(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 December 2003 (11.12.2003)

PCT

(10) International Publication Number
WO 03/102159 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US03/17573

(22) International Filing Date: 4 June 2003 (04.06.2003)

(25) Filing Language: English

(26) Publication Language: English

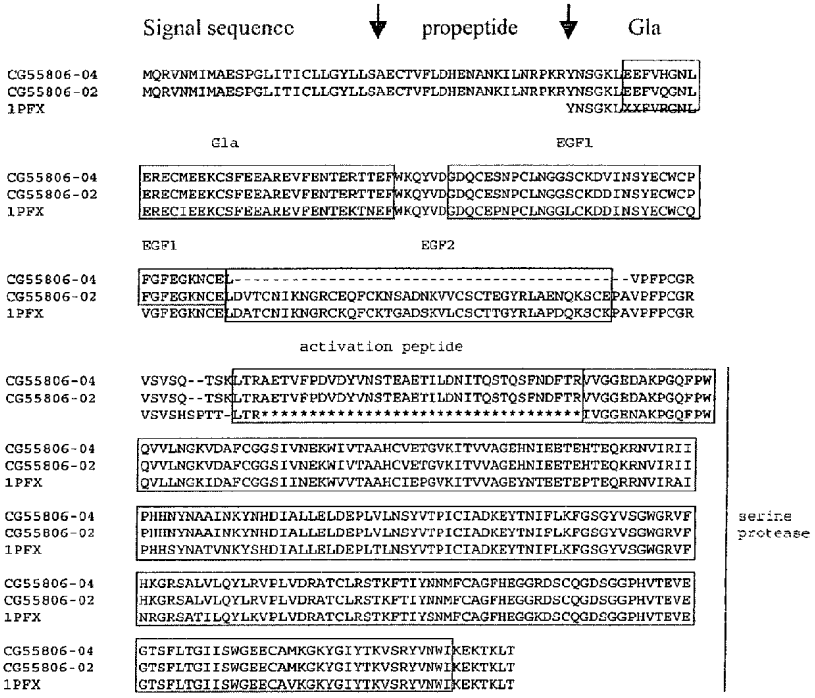
(30) Priority Data:

60/385,615	4 June 2002 (04.06.2002)	US
60/385,490	4 June 2002 (04.06.2002)	US
60/385,755	4 June 2002 (04.06.2002)	US
60/386,041	5 June 2002 (05.06.2002)	US
60/386,357	6 June 2002 (06.06.2002)	US
60/386,465	6 June 2002 (06.06.2002)	US
60/386,355	6 June 2002 (06.06.2002)	US
60/386,447	6 June 2002 (06.06.2002)	US
60/386,459	6 June 2002 (06.06.2002)	US
60/386,864	6 June 2002 (06.06.2002)	US
60/386,796	7 June 2002 (07.06.2002)	US
60/386,701	7 June 2002 (07.06.2002)	US
60/387,081	7 June 2002 (07.06.2002)	US
60/387,083	7 June 2002 (07.06.2002)	US
60/387,078	7 June 2002 (07.06.2002)	US
60/386,931	7 June 2002 (07.06.2002)	US
60/387,866	10 June 2002 (10.06.2002)	US
60/387,429	10 June 2002 (10.06.2002)	US
60/387,540	10 June 2002 (10.06.2002)	US
60/387,606	11 June 2002 (11.06.2002)	US
60/387,859	11 June 2002 (11.06.2002)	US
60/387,610	11 June 2002 (11.06.2002)	US
60/387,659	11 June 2002 (11.06.2002)	US
60/387,668	11 June 2002 (11.06.2002)	US
60/387,696	11 June 2002 (11.06.2002)	US
60/387,934	12 June 2002 (12.06.2002)	US
60/388,022	12 June 2002 (12.06.2002)	US
60/388,096	12 June 2002 (12.06.2002)	US
60/388,432	12 June 2002 (12.06.2002)	US

[Continued on next page]

(54) Title: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

Alignment of CG55806-04 (SEQ ID NO:748), CG55806-02 (SEQ ID NO:752), and 1PFX (SEQ ID NO: 1476).



(57) Abstract: The present invention provides novel isolated polynucleotides and small molecule target polypeptides encoded by the polynucleotides. Antibodies that immunospecifically bind to a novel small molecule target polypeptide or any derivative, variant, mutant or fragment of that polypeptide, polynucleotide or antibody are disclosed, as are methods in which the small molecule target polypeptide, polynucleotide and antibody are utilized in the detection and treatment of a broad range of pathological states. More specifically, the present invention discloses methods of using recombinantly expressed and/or endogenously expressed proteins in various screening procedures for the purpose of identifying therapeutic antibodies and therapeutic small molecules associated with diseases. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

WO 03/102159 A2

60/387,960	12 June 2002 (12.06.2002)	US
60/388,479	12 June 2002 (12.06.2002)	US
60/389,123	13 June 2002 (13.06.2002)	US
60/389,120	14 June 2002 (14.06.2002)	US
60/389,146	14 June 2002 (14.06.2002)	US
60/389,742	17 June 2002 (17.06.2002)	US
60/389,604	18 June 2002 (18.06.2002)	US
60/389,884	18 June 2002 (18.06.2002)	US
60/390,006	19 June 2002 (19.06.2002)	US
60/390,144	19 June 2002 (19.06.2002)	US
60/390,209	19 June 2002 (19.06.2002)	US
60/391,726	25 June 2002 (25.06.2002)	US
60/401,628	6 August 2002 (06.08.2002)	US
60/402,268	9 August 2002 (09.08.2002)	US
60/402,822	12 August 2002 (12.08.2002)	US
60/403,458	13 August 2002 (13.08.2002)	US
60/403,617	15 August 2002 (15.08.2002)	US
60/403,732	15 August 2002 (15.08.2002)	US
60/406,182	26 August 2002 (26.08.2002)	US
60/410,085	12 September 2002 (12.09.2002)	US
60/410,505	13 September 2002 (13.09.2002)	US
60/412,955	23 September 2002 (23.09.2002)	US
60/415,195	30 September 2002 (30.09.2002)	US
60/420,718	23 October 2002 (23.10.2002)	US
60/420,627	23 October 2002 (23.10.2002)	US
60/420,852	24 October 2002 (24.10.2002)	US
60/422,750	31 October 2002 (31.10.2002)	US
60/423,095	1 November 2002 (01.11.2002)	US
60/423,748	5 November 2002 (05.11.2002)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/385,615 (CIP)
Filed on	4 June 2002 (04.06.2002)
US	60/402,268 (CIP)
Filed on	9 August 2002 (09.08.2002)
US	60/387,606 (CIP)
Filed on	11 June 2002 (11.06.2002)
US	60/386,357 (CIP)
Filed on	6 June 2002 (06.06.2002)
US	60/385,755 (CIP)
Filed on	4 June 2002 (04.06.2002)
US	60/386,355 (CIP)
Filed on	6 June 2002 (06.06.2002)
US	60/385,490 (CIP)
Filed on	4 June 2002 (04.06.2002)
US	60/420,718 (CIP)
Filed on	23 October 2002 (23.10.2002)
US	60/386,447 (CIP)
Filed on	6 June 2002 (06.06.2002)
US	60/386,465 (CIP)
Filed on	6 June 2002 (06.06.2002)
US	60/420,627 (CIP)
Filed on	23 October 2002 (23.10.2002)
US	60/386,459 (CIP)
Filed on	6 June 2002 (06.06.2002)
US	60/410,505 (CIP)
Filed on	13 September 2002 (13.09.2002)
US	60/420,852 (CIP)
Filed on	24 October 2002 (24.10.2002)
US	60/386,769 (CIP)
Filed on	7 June 2002 (07.06.2002)
US	60/387,078 (CIP)
Filed on	7 June 2002 (07.06.2002)
US	60/387,083 (CIP)
Filed on	7 June 2002 (07.06.2002)
US	60/387,081 (CIP)
Filed on	7 June 2002 (07.06.2002)
US	60/386,041 (CIP)
Filed on	5 June 2002 (05.06.2002)
US	60/386,701 (CIP)
Filed on	7 June 2002 (07.06.2002)
US	60/387,610 (CIP)
Filed on	11 June 2002 (11.06.2002)
US	60/387,540 (CIP)
Filed on	10 June 2002 (10.06.2002)
US	60/387,429 (CIP)
Filed on	10 June 2002 (10.06.2002)
US	60/410,085 (CIP)
Filed on	12 September 2002 (12.09.2002)
US	60/389,120 (CIP)
Filed on	14 June 2002 (14.06.2002)
US	60/386,931 (CIP)
Filed on	7 June 2002 (07.06.2002)
US	60/387,866 (CIP)
Filed on	10 June 2002 (10.06.2002)
US	60/387,859 (CIP)
Filed on	11 June 2002 (11.06.2002)
US	60/387,659 (CIP)
Filed on	11 June 2002 (11.06.2002)
US	60/387,934 (CIP)
Filed on	12 June 2002 (12.06.2002)
US	60/387,696 (CIP)
Filed on	11 June 2002 (11.06.2002)
US	60/390,006 (CIP)
Filed on	19 June 2002 (19.06.2002)
US	60/389,604 (CIP)
Filed on	18 June 2002 (18.06.2002)
US	60/387,668 (CIP)
Filed on	11 June 2002 (11.06.2002)
US	60/386,864 (CIP)
Filed on	6 June 2002 (06.06.2002)
US	60/401,628 (CIP)
Filed on	6 August 2002 (06.08.2002)
US	60/406,182 (CIP)
Filed on	26 August 2002 (26.08.2002)
US	60/412,955 (CIP)
Filed on	23 September 2002 (23.09.2002)
US	60/415,195 (CIP)
Filed on	30 September 2002 (30.09.2002)
US	60/422,750 (CIP)
Filed on	31 October 2002 (31.10.2002)
US	60/390,144 (CIP)
Filed on	19 June 2002 (19.06.2002)
US	60/388,022 (CIP)
Filed on	12 June 2002 (12.06.2002)
US	60/402,822 (CIP)
Filed on	12 August 2002 (12.08.2002)
US	60/388,096 (CIP)
Filed on	12 June 2002 (12.06.2002)
US	60/389,123 (CIP)
Filed on	13 June 2002 (13.06.2002)
US	60/390,209 (CIP)
Filed on	19 June 2002 (19.06.2002)
US	60/388,479 (CIP)
Filed on	12 June 2002 (12.06.2002)

[Continued on next page]

US 60/403,458 (CIP) Filed on 13 August 2002 (13.08.2002)
US 60/389,884 (CIP) Filed on 18 June 2002 (18.06.2002)
US 60/389,146 (CIP) Filed on 14 June 2002 (14.06.2002)
US 60/387,960 (CIP) Filed on 12 June 2002 (12.06.2002)
US 60/388,432 (CIP) Filed on 12 June 2002 (12.06.2002)
US 60/403,617 (CIP) Filed on 15 August 2002 (15.08.2002)
US 60/423,095 (CIP) Filed on 1 November 2002 (01.11.2002)
US 60/423,748 (CIP) Filed on 5 November 2002 (05.11.2002)
US 60/391,726 (CIP) Filed on 25 June 2002 (25.06.2002)
US 60/403,732 (CIP) Filed on 15 August 2002 (15.08.2002)
US 60/389,742 (CIP) Filed on 17 June 2002 (17.06.2002)

(71) **Applicant** *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **ALSOBROOK, John, P., II** [US/US]; 60 Lake Drive, Madison, CT 06443 (US). **ANDERSON, David, W.** [US/US]; 85 Montoya Drive, Branford, CT 06405 (US). **BAUMGARTNER, Jason, C.** [US/US]; 1697 Quinnipiac Avenue, New Haven, CT 06513 (US). **BERGHS, Constance** [NL/US]; 459 Orange Street, New Haven, CT 06511 (US). **BOLDOG, Ferenc, L.** [HU/US]; 1687 Hartford Turnpike, North Haven, CT 06473 (US). **BURGESS, Catherine, E.** [US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109 (US). **CASMAN, Stacie, J.** [US/US]; 17 Peck Street, North Haven, CT 06473 (US). **CATTERTON, Elina** [FI/US]; 584, Boston Post Road, Madison, CT 06443 (US). **DHANABAL, Mohanraj** [US/US]; 9 C Stonegate Circle, Branford, CT 06405 (US). **EDINGER, Shlomit, R.** [US/US]; 766 Edgewood Avenue, New Haven, CT 06515 (US). **ELLERMAN, Karen** [US/US]; 87 Montoya Drive, Branford, CT 06405 (US). **ETTENBERG, Seth** [US/US]; 583 Orange Street, New Haven, CT 06511 (US). **GANGOLLI, Esha, A.** [IN/US]; 31 Strawberry Hill Road, Madison, CT 06443 (US). **GERLACH, Valerie, L.** [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). **GORMAN, Linda** [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). **GROSSE, William, M.** [US/US]; 15 Rice Terrace Road, Apartment C, Branford, CT 06405 (US). **GUNTHER, Erik** [US/US]; 34 Bryan Road, Branford, CT 06405 (US). **GUO, Xiaojia** [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). **GUSEV, Vladimir, Y.** [UA/US]; 1209 Durham Road, Madison, CT 06443 (US). **HERRMANN, John, L.** [US/US]; 78 Barnshed Lane, Guilford, CT 06437 (US). **JI, Weizhen** [CN/US]; 375 Old Rock Road, Branford, CT 06405 (US). **KEKUDA, Ramesh** [IN/US]; 71 Aiken Street, Unit R3, Norwalk, CT 06851 (US). **KHRAMTSOV, Nikolai, V.** [RU/US]; 38 Montoya Circle, Branford, CT 06405 (US). **LAROCHELLE, William, J.** [US/US]; 15 Devonshire Lane, Madison, CT 06443 (US). **LI, Li** [CN/US]; 56 Jerimoth Drive, Branford, CT 06405 (US). **LIANG, Hongping** [CN/US]; 2 Skiff Street, Apartment 109, Hamden, CT 06514 (US). **LOW, Kenneth** [US/US]; 188 Livimgston Street, New Haven, CT 06511 (US). **MACDOUGALL, John, R.** [CA/US]; 117 Russell Street, Hamden, CT 06517 (US). **MACLACHLAN, Timothy** [US/US]; 8 Jefferson Street, Unionville, CT 06085 (US). **MALYANKAR, Uriel, M.** [IN/US]; 229 Branford Road, Number 330, North Branford, CT 06471 (US). **MCQUEENEY, Kelly** [US/US]; 7 Pin Oak Lane, Ansonia, CT 06401 (US). **MEZICK, Amanda, J.** [US/US]; 240 West Woods Road, Hamden, CT 06518 (US). **MILLER, Charles, E.** [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US). **MILLET, Isabelle** [FR/US]; 74 Carrington Avenue, Milford, CT 06460 (US). **PADIGARU, Muralidhara** [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). **PEYMAN, John, A.** [US/US]; 336 West Rock Avenue, New Haven, CT 06515 (US). **QIAN, Xiaozhong** [US/US]; 79 Florence Road, Branford, CT 06405 (US). **RASTELLI, Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). **RIEGER, Daniel, K.** [DE/US]; 10A McKinnel Court, Branford, CT 06405 (US). **ROTHENBERG, Mark, E.** [US/US]; 2 Allen Road, Clinton, CT 06413 (US). **SHENOY, Suresh, G.** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **SHIMKETS, Richard, A.** [US/US]; 5 Indian Meadows Drive, Guilford, CT 06437 (US). **SMITHSON, Glennda** [US/US]; 125 Michael Drive, Guilford, CT 06435 (US). **SPYTEK, Kimberly, A.** [US/US]; 7 Hotel Road, Ellington, CT 06029 (US). **STONE, David, J.** [US/US]; 223 Whitehorn Drive, Guilford, CT 06437 (US). **SUKUMARAN, Sujatha** [US/US]; 8B Stonegate Circle, Branford, CT 06405 (US). **SZEKERES, Edward, S., Jr.** [US/US]; 110 Constitution Street, Wallington, CT 06492 (US). **VERNET, Corine, A., M.** [FR/US]; 1739 Foxon Road, Apartment L6, North Branford, CT 06471 (US). **VOSS, Edward, Z.** [US/US]; 123 Knollwood Drive, Wallingford, CT 06492 (US). **WOLENC, Adam, R.** [US/US]; 183 Fairmont Avenue, New Haven, CT 06513 (US). **ZHONG, Mei** [CA/US]; 152 Brushy Plain Road, Branford, CT 06405 (US). **ZHONG, Haihong** [CN/US]; 2269 Long Hill Road, Guilford, CT 06437 (US).

(74) **Agent: ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., ., One Financial Center, Boston, MA 02111 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,

[Continued on next page]

ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

FIELD OF THE INVENTION

The present invention relates to novel polypeptides that are targets of small molecule drugs and that have properties related to stimulation of biochemical or physiological responses in a cell, a tissue, an organ or an organism. More particularly, the novel polypeptides are gene products of novel genes, or are specified biologically active fragments or derivatives thereof. Methods of use encompass diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions.

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

BACKGROUND

Eukaryotic cells are characterized by biochemical and physiological processes which under normal conditions are exquisitely balanced to achieve the preservation and propagation of the cells. When such cells are components of multicellular organisms such as vertebrates, or more particularly organisms such as mammals, the regulation of the biochemical and physiological processes involves intricate signaling pathways. Frequently, such signaling pathways involve extracellular signaling proteins, cellular receptors that bind the signaling proteins and signal transducing components located within the cells.

Signaling proteins may be classified as endocrine effectors, paracrine effectors or autocrine effectors. Endocrine effectors are signaling molecules secreted by a given organ into the circulatory system, which are then transported to a distant target organ or tissue. The target cells include the receptors for the endocrine effector, and when the endocrine effector binds, a signaling cascade is induced. Paracrine effectors involve secreting cells and receptor cells in close proximity to each other, for example two different classes of cells in the same tissue or organ. One class of cells secretes the paracrine effector, which then reaches the second class of cells, for example by diffusion through the extracellular fluid. The second class of cells contains the receptors for the paracrine effector; binding of the effector results in induction of the signaling cascade that elicits the corresponding biochemical or physiological effect. Autocrine effectors are highly analogous to paracrine effectors, except that the same cell type that secretes the autocrine effector also contains the receptor. Thus the autocrine effector binds to receptors on the same cell, or on identical neighboring cells. The binding process then elicits the characteristic biochemical or physiological effect.

Signaling processes may elicit a variety of effects on cells and tissues including by way of nonlimiting example induction of cell or tissue proliferation, suppression of growth or proliferation, induction of differentiation or maturation of a cell or tissue, and suppression of differentiation or maturation of a cell or tissue.

Many pathological conditions involve dysregulation of expression of important effector proteins. In certain classes of pathologies the dysregulation is manifested as diminished or suppressed level of synthesis and secretion of protein effectors. In other classes of pathologies the dysregulation is manifested as increased or up-regulated level

of synthesis and secretion of protein effectors. In a clinical setting a subject may be suspected of suffering from a condition brought on by altered or mis-regulated levels of a protein effector of interest. Therefore there is a need to assay for the level of the protein effector of interest in a biological sample from such a subject, and to compare the level with that characteristic of a nonpathological condition. There also is a need to provide the protein effector as a product of manufacture. Administration of the effector to a subject in need thereof is useful in treatment of the pathological condition. Accordingly, there is a need for a method of treatment of a pathological condition brought on by a diminished or suppressed levels of the protein effector of interest. In addition, there is a need for a method of treatment of a pathological condition brought on by a increased or up-regulated levels of the protein effector of interest.

Small molecule targets have been implicated in various disease states or pathologies. These targets may be proteins, and particularly enzymatic proteins, which are acted upon by small molecule drugs for the purpose of altering target function and achieving a desired result. Cellular, animal and clinical studies can be performed to elucidate the genetic contribution to the etiology and pathogenesis of conditions in which small molecule targets are implicated in a variety of physiologic, pharmacologic or native states. These studies utilize the core technologies at CuraGen Corporation to look at differential gene expression, protein-protein interactions, large-scale sequencing of expressed genes and the association of genetic variations such as, but not limited to, single nucleotide polymorphisms (SNPs) or splice variants in and between biological samples from experimental and control groups. The goal of such studies is to identify potential avenues for therapeutic intervention in order to prevent, treat the consequences or cure the conditions.

In order to treat diseases, pathologies and other abnormal states or conditions in which a mammalian organism has been diagnosed as being, or as being at risk for becoming, other than in a normal state or condition, it is important to identify new therapeutic agents. Such a procedure includes at least the steps of identifying a target component within an affected tissue or organ, and identifying a candidate therapeutic agent that modulates the functional attributes of the target. The target component may be any biological macromolecule implicated in the disease or pathology. Commonly the target is a polypeptide or protein with specific functional attributes. Other classes of macromolecule may be a nucleic acid, a polysaccharide, a lipid such as a complex lipid

or a glycolipid; in addition a target may be a sub-cellular structure or extra-cellular structure that is comprised of more than one of these classes of macromolecule. Once such a target has been identified, it may be employed in a screening assay in order to identify favorable candidate therapeutic agents from among a large population of substances or compounds.

In many cases the objective of such screening assays is to identify small molecule candidates; this is commonly approached by the use of combinatorial methodologies to develop the population of substances to be tested. The implementation of high throughput screening methodologies is advantageous when working with large, combinatorial libraries of compounds.

SUMMARY OF THE INVENTION

The invention includes nucleic acid sequences and the novel polypeptides they encode. The novel nucleic acids and polypeptides are referred to herein as NOVX, or NOV1, NOV2, NOV3, *etc.*, nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid, which represents the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 566, or polypeptide sequences, which represents the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 566.

In one aspect, the invention provides an isolated polypeptide comprising a mature form of a NOVX amino acid. One example is a variant of a mature form of a NOVX amino acid sequence, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. The amino acid can be, for example, a NOVX amino acid sequence or a variant of a NOVX amino acid sequence, wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed. The invention also includes fragments of any of these. In another aspect, the invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof.

Also included in the invention is a NOVX polypeptide that is a naturally occurring allelic variant of a NOVX sequence. In one embodiment, the allelic variant

includes an amino acid sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a NOVX nucleic acid sequence. In another embodiment, the NOVX polypeptide is a variant polypeptide described therein, wherein any amino acid specified in the chosen sequence is changed to provide a conservative substitution. In one embodiment, the invention discloses a method for determining the presence or amount of the NOVX polypeptide in a sample. The method involves the steps of: providing a sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the NOVX polypeptide, thereby determining the presence or amount of the NOVX polypeptide in the sample. In another embodiment, the invention provides a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide in a mammalian subject. This method involves the steps of: measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in the sample of the first step to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease, wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

In a further embodiment, the invention includes a method of identifying an agent that binds to a NOVX polypeptide. This method involves the steps of: introducing the polypeptide to the agent; and determining whether the agent binds to the polypeptide. In various embodiments, the agent is a cellular receptor or a downstream effector.

In another aspect, the invention provides a method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of a NOVX polypeptide. The method involves the steps of: providing a cell expressing the NOVX polypeptide and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent. In another aspect, the invention describes a method for screening for a modulator of activity or of latency or predisposition to a pathology associated with the

NOVX polypeptide. This method involves the following steps: administering a test compound to a test animal at increased risk for a pathology associated with the NOVX polypeptide, wherein the test animal recombinantly expresses the NOVX polypeptide. This method involves the steps of measuring the activity of the NOVX polypeptide in the test animal after administering the compound of step; and comparing the activity of the protein in the test animal with the activity of the NOVX polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the NOVX polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of, or predisposition to, a pathology associated with the NOVX polypeptide. In one embodiment, the test animal is a recombinant test animal that expresses a test protein transgene or expresses the transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein the promoter is not the native gene promoter of the transgene. In another aspect, the invention includes a method for modulating the activity of the NOVX polypeptide, the method comprising introducing a cell sample expressing the NOVX polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. In a preferred embodiment, the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant. In another embodiment, the nucleic acid encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant. In another embodiment, the nucleic acid molecule differs by a single nucleotide from a NOVX nucleic acid sequence. In one embodiment, the NOVX nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 566, or a complement of the nucleotide sequence. In another aspect, the invention provides a vector or a cell expressing a NOVX nucleotide sequence.

In one embodiment, the invention discloses a method for modulating the activity of a NOVX polypeptide. The method includes the steps of: introducing a cell sample expressing the NOVX polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide. In another embodiment,

the invention includes an isolated NOVX nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising a NOVX amino acid sequence or a variant of a mature form of the NOVX amino acid sequence, wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. In another embodiment, the invention includes an amino acid sequence that is a variant of the NOVX amino acid sequence, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed.

In one embodiment, the invention discloses a NOVX nucleic acid fragment encoding at least a portion of a NOVX polypeptide or any variant of the polypeptide, wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed. In another embodiment, the invention includes the complement of any of the NOVX nucleic acid molecules or a naturally occurring allelic nucleic acid variant. In another embodiment, the invention discloses a NOVX nucleic acid molecule that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant. In another embodiment, the invention discloses a NOVX nucleic acid, wherein the nucleic acid molecule differs by a single nucleotide from a NOVX nucleic acid sequence.

In another aspect, the invention includes a NOVX nucleic acid, wherein one or more nucleotides in the NOVX nucleotide sequence is changed to a different nucleotide provided that no more than 15% of the nucleotides are so changed. In one embodiment, the invention discloses a nucleic acid fragment of the NOVX nucleotide sequence and a nucleic acid fragment wherein one or more nucleotides in the NOVX nucleotide sequence is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed. In another embodiment, the invention includes a nucleic acid molecule wherein the nucleic acid molecule hybridizes under stringent conditions to a NOVX nucleotide sequence or a complement of the NOVX nucleotide sequence. In one embodiment, the invention includes a nucleic acid molecule, wherein the sequence is changed such that no more than 15% of the nucleotides in the coding sequence differ from the NOVX nucleotide sequence or a fragment thereof.

In a further aspect, the invention includes a method for determining the presence or amount of the NOVX nucleic acid in a sample. The method involves the steps of: providing the sample; introducing the sample to a probe that binds to the nucleic acid molecule; and determining the presence or amount of the probe bound to the NOVX nucleic acid molecule, thereby determining the presence or amount of the NOVX nucleic acid molecule in the sample. In one embodiment, the presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

In another aspect, the invention discloses a method for determining the presence of or predisposition to a disease associated with altered levels of the NOVX nucleic acid molecule of in a first mammalian subject. The method involves the steps of: measuring the amount of NOVX nucleic acid in a sample from the first mammalian subject; and comparing the amount of the nucleic acid in the sample of step (a) to the amount of NOVX nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease; wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

BRIEF DESCRIPTION OF THE FIGURES

Fig. D1: Alignment of CG55806-04 (SEQ ID NO:748), CG55806-02 (SEQ ID NO:752), and 1PFX (SEQ ID NO: 1476).

Fig. D2: Structure of porcine factor IXa (1PFX).

Fig. E1: Data showing effect on cell growth by knockdown of CG59693-01.

Fig. E2: Data showing effect on cell growth by knockdown of CG59693-01 with subsequent treatment with Paclitaxel (48 hr).

Fig. E3: Data showing effect on cell viability by knockdown of CG59693-01 with subsequent treatment with Paclitaxel (48 hr).

Fig. E4: Data showing effect on cell growth by knockdown of CG59693-01 with subsequent treatment with Paclitaxel (72 hr).

Fig. E5: Data showing effect on cell viability by knockdown of CG59693-01 with subsequent treatment with Paclitaxel (72 hr).

Fig. E6: Data showing effect on cell growth by knockdown of CG59693-01 by AS4 antisense oligonucleotide followed by subsequent treatment with Gemcitabine.

Fig. E7: Data showing effect on cell growth by knockdown of CG59693-01 by AS4 antisense oligonucleotide followed by subsequent treatment with Daunorubicin.

Fig. E8: Data showing effect on cell growth by knockdown of CG59693-01 by AS4 antisense oligonucleotide followed by subsequent treatment with Etoposide.

Fig. E9: Data showing effect on cell growth by knockdown of CG59693-01 by AS4 antisense oligonucleotide followed by subsequent treatment with Cisplatin.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences, their encoded polypeptides, antibodies, and other related compounds. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the

corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (amino acid)	Homology
NOV1a	CG101340-01	1	2	Putative G protein-coupled receptor 92 (Putative G-protein coupled receptor) - Homo sapiens
NOV1b	SNP13382465	3	4	Putative G protein-coupled receptor 92 (Putative G-protein coupled receptor) - Homo sapiens
NOV2a	CG101396-01	5	6	Glutamate receptor delta-1 subunit precursor - Rattus norvegicus
NOV2b	267253224	7	8	Glutamate receptor delta-1 subunit precursor - Rattus norvegicus
NOV2c	315490179	9	10	Glutamate receptor delta-1 subunit precursor - Rattus norvegicus
NOV2d	CG101396-02	11	12	Glutamate receptor delta-1 subunit precursor - Rattus norvegicus
NOV2e	SNP13379211	13	14	Glutamate receptor delta-1 subunit precursor - Rattus norvegicus
NOV3a	CG102348-01	15	16	Complement C1r-like proteinase - Homo sapiens
NOV3b	199842645	17	18	Complement C1r-like proteinase - Homo sapiens
NOV3c	198306343	19	20	Complement C1r-like proteinase - Homo sapiens
NOV3d	199842665	21	22	Complement C1r-like proteinase - Homo sapiens
NOV3e	199842661	23	24	Complement C1r-like proteinase - Homo sapiens
NOV3f	199597024	25	26	Complement C1r-like proteinase - Homo sapiens
NOV3g	199842653	27	28	Complement C1r-like proteinase - Homo sapiens
NOV3h	199652830	29	30	Complement C1r-like proteinase - Homo sapiens

NOV3i	199652835	31	32	Complement C1r-like proteinase - Homo sapiens
NOV3j	198306308	33	34	Complement C1r-like proteinase - Homo sapiens
NOV3k	CG102348-02	35	36	Complement C1r-like proteinase - Homo sapiens
NOV3l	CG102348-03	37	38	Complement C1r-like proteinase - Homo sapiens
NOV3m	CG102348-04	39	40	Complement C1r-like proteinase - Homo sapiens
NOV3n	CG102348-05	41	42	Complement C1r-like proteinase - Homo sapiens
NOV3o	CG102348-06	43	44	Complement C1r-like proteinase - Homo sapiens
NOV3p	SNP13376570	45	46	Complement C1r-like proteinase - Homo sapiens
NOV3q	SNP13376568	47	48	Complement C1r-like proteinase - Homo sapiens
NOV3r	SNP13382463	49	50	Complement C1r-like proteinase - Homo sapiens
NOV3s	SNP13374245	51	52	Complement C1r-like proteinase - Homo sapiens
NOV4a	CG125860-02	53	54	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens
NOV4b	CG125860-01	55	56	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens
NOV4c	SNP13376012	57	58	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens
NOV4d	SNP13376014	59	60	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens
NOV4e	SNP13376015	61	62	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens
NOV4f	SNP13376016	63	64	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens

NOV4g	SNP13376017	65	66	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens
NOV4h	SNP13376011	67	68	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens
NOV4i	SNP13376018	69	70	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens
NOV4j	SNP13382467	71	72	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens
NOV4k	SNP13382466	73	74	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens
NOV5a	CG155759-02	75	76	Seven transmembrane helix receptor - Homo sapiens
NOV5b	CG155759-01	77	78	Seven transmembrane helix receptor - Homo sapiens
NOV6a	CG187667-01	79	80	LOC133308 protein - Homo sapiens
NOV6b	CG187667-02	81	82	LOC133308 protein - Homo sapiens
NOV7a	CG187676-01	83	84	Sentrin-specific protease 5 (EC 3.4.22.-) (Sentrin/SUMO-specific protease SENP5) (Protease FKSG45) - Homo sapiens
NOV8a	CG50235-04	85	86	Tolloid-like 2 protein - Homo sapiens
NOV8b	CG50235-01	87	88	Tolloid-like 2 protein - Homo sapiens
NOV8c	CG50235-02	89	90	Tolloid-like 2 protein - Homo sapiens
NOV8d	CG50235-03	91	92	Tolloid-like 2 protein - Homo sapiens
NOV8e	SNP13377383	93	94	Tolloid-like 2 protein - Homo sapiens
NOV8f	SNP13377384	95	96	Tolloid-like 2 protein - Homo sapiens
NOV8g	SNP13377385	97	98	Tolloid-like 2 protein - Homo sapiens
NOV8h	SNP13377386	99	100	Tolloid-like 2 protein - Homo sapiens
NOV8i	SNP13377387	101	102	Tolloid-like 2 protein - Homo sapiens
NOV9a	CG50249-01	103	104	Voltage gated potassium channel Kv3.2b (Potassium voltage-gated potassium channel subfamily C member 2) - Homo sapiens
NOV9b	207885588	105	106	Voltage gated potassium channel Kv3.2b (Potassium voltage-gated potassium channel subfamily C member 2) - Homo sapiens

NOV9c	CG50249-02	107	108	Voltage gated potassium channel Kv3.2b (Potassium voltage-gated potassium channel subfamily C member 2) - Homo sapiens
NOV9d	CG50249-03	109	110	Voltage gated potassium channel Kv3.2b (Potassium voltage-gated potassium channel subfamily C member 2) - Homo sapiens
NOV9e	CG50249-04	111	112	Voltage gated potassium channel Kv3.2b (Potassium voltage-gated potassium channel subfamily C member 2) - Homo sapiens
NOV10a	CG50307-03	113	114	Steroid dehydrogenase-like protein
NOV10b	275624102	115	116	Steroid dehydrogenase-like protein
NOV10c	CG50307-01	117	118	Steroid dehydrogenase-like protein
NOV10d	CG50307-02	119	120	Steroid dehydrogenase-like protein
NOV10e	SNP13375811	121	122	Steroid dehydrogenase-like protein
NOV11a	CG50315-01	123	124	Seven transmembrane helix receptor - Homo sapiens
NOV11b	207580272	125	126	Seven transmembrane helix receptor - Homo sapiens
NOV11c	314411778	127	128	Seven transmembrane helix receptor - Homo sapiens
NOV11d	CG50315-02	129	130	Seven transmembrane helix receptor - Homo sapiens
NOV11e	CG50315-03	131	132	Seven transmembrane helix receptor - Homo sapiens
NOV12a	CG50341-01	133	134	Seven transmembrane helix receptor - Homo sapiens
NOV12b	169475616	135	136	Seven transmembrane helix receptor - Homo sapiens
NOV12c	CG50341-02	137	138	Seven transmembrane helix receptor - Homo sapiens
NOV12d	CG50341-03	139	140	Seven transmembrane helix receptor - Homo sapiens
NOV12e	CG50341-04	141	142	Seven transmembrane helix receptor - Homo sapiens
NOV13a	CG50365-01	143	144	Carbonic anhydrase XIII (EC 4.2.1.1) (Carbonate dehydratase XIII) (CA-XIII) - Homo sapiens
NOV13b	278019595	145	146	Carbonic anhydrase XIII (EC 4.2.1.1) (Carbonate dehydratase XIII) (CA-XIII) - Homo sapiens

NOV13c	CG50365-02	147	148	Carbonic anhydrase XIII (EC 4.2.1.1) (Carbonate dehydratase XIII) (CA-XIII) - Homo sapiens
NOV14a	CG50367-01	149	150	ADAM 33 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 33) - Homo sapiens
NOV14b	CG50367-05	151	152	ADAM 33 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 33) - Homo sapiens
NOV14c	CG50367-06	153	154	ADAM 33 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 33) - Homo sapiens
NOV14d	CG50367-07	155	156	ADAM 33 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 33) - Homo sapiens
NOV14e	249356906	157	158	ADAM 33 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 33) - Homo sapiens
NOV14f	CG50367-02	159	160	ADAM 33 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 33) - Homo sapiens
NOV14g	CG50367-03	161	162	ADAM 33 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 33) - Homo sapiens
NOV14h	CG50367-04	163	164	ADAM 33 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 33) - Homo sapiens
NOV15a	CG50718-02	165	166	Glomerular mesangial cell receptor protein-tyrosine phosphatase precursor - Rattus norvegicus
NOV15b	CG50718-06	167	168	Glomerular mesangial cell receptor protein-tyrosine phosphatase precursor - Rattus norvegicus
NOV15c	258979883	169	170	Glomerular mesangial cell receptor protein-tyrosine phosphatase precursor - Rattus norvegicus
NOV15d	CG50718-01	171	172	Glomerular mesangial cell receptor protein-tyrosine phosphatase precursor - Rattus norvegicus
NOV15e	CG50718-03	173	174	Glomerular mesangial cell receptor protein-tyrosine phosphatase precursor - Rattus norvegicus
NOV15f	CG50718-04	175	176	Glomerular mesangial cell receptor protein-tyrosine phosphatase precursor - Rattus norvegicus

NOV15g	CG50718-05	177	178	Glomerular mesangial cell receptor protein-tyrosine phosphatase precursor - Rattus norvegicus
NOV16a	CG50934-03	179	180	Mast cell protease-11 - Mus musculus
NOV16b	CG50934-01	181	182	Mast cell protease-11 - Mus musculus
NOV16c	CG50934-02	183	184	Mast cell protease-11 - Mus musculus
NOV16d	SNP13381559	185	186	Mast cell protease-11 - Mus musculus
NOV16e	SNP13381558	187	188	Mast cell protease-11 - Mus musculus
NOV16f	SNP13376399	189	190	Mast cell protease-11 - Mus musculus
NOV16g	SNP13378301	191	192	Mast cell protease-11 - Mus musculus
NOV16h	SNP13378300	193	194	Mast cell protease-11 - Mus musculus
NOV17a	CG51213-04	195	196	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Homo sapiens (Human), 1077 aa
NOV17b	CG51213-01	197	198	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Homo sapiens (Human), 1077 aa
NOV17c	306345264	199	200	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Homo sapiens (Human), 1077 aa
NOV17d	CG51213-02	201	202	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Homo sapiens (Human), 1077 aa
NOV17e	CG51213-03	203	204	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Homo sapiens (Human), 1077 aa

NOV17f	CG51213-05	205	206	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Homo sapiens (Human), 1077 aa
NOV17g	CG51213-06	207	208	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Homo sapiens (Human), 1077 aa
NOV17h	CG51213-07	209	210	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Homo sapiens (Human), 1077 aa
NOV18a	CG51448-05	211	212	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18b	CG51448-01	213	214	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18c	274051198	215	216	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18d	274051170	217	218	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18e	CG51448-02	219	220	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18f	CG51448-03	221	222	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18g	CG51448-04	223	224	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18h	SNP13375535	225	226	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18i	SNP13375536	227	228	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18j	SNP13375537	229	230	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18k	SNP13375538	231	232	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18l	SNP13375539	233	234	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18m	SNP13375540	235	236	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18n	SNP13375541	237	238	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D

NOV18o	SNP13375542	239	240	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV18p	SNP13375543	241	242	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D
NOV19a	CG51752-02	243	244	Human FCTR6b polypeptide sequence
NOV19b	CG51752-03	245	246	Human FCTR6b polypeptide sequence
NOV19c	175069825	247	248	Human FCTR6b polypeptide sequence
NOV19d	175069842	249	250	Human FCTR6b polypeptide sequence
NOV19e	258076315	251	252	Human FCTR6b polypeptide sequence
NOV19f	258076366	253	254	Human FCTR6b polypeptide sequence
NOV19g	CG51752-04	255	256	Human FCTR6b polypeptide sequence
NOV19h	191887409	257	258	Human FCTR6b polypeptide sequence
NOV19i	CG51752-01	259	260	Human FCTR6b polypeptide sequence
NOV19j	CG51752-05	261	262	Human FCTR6b polypeptide sequence
NOV19k	CG51752-06	263	264	Human FCTR6b polypeptide sequence
NOV19l	CG51752-07	265	266	Human FCTR6b polypeptide sequence
NOV19m	SNP13374584	267	268	Human FCTR6b polypeptide sequence
NOV19n	SNP13374585	269	270	Human FCTR6b polypeptide sequence
NOV20a	CG51914-02	271	272	Human kinase polypeptide (PKIN-16)
NOV20b	CG51914-01	273	274	Human kinase polypeptide (PKIN-16)
NOV21a	CG51965-01	275	276	Cadherin EGF LAG seven-pass G-type receptor 1 precursor (Flamingo homolog 2) (hFmi2) - Homo sapiens
NOV21b	258076370	277	278	Cadherin EGF LAG seven-pass G-type receptor 1 precursor (Flamingo homolog 2) (hFmi2) - Homo sapiens
NOV21c	317619862	279	280	Cadherin EGF LAG seven-pass G-type receptor 1 precursor (Flamingo homolog 2) (hFmi2) - Homo sapiens

NOV21d	317460050	281	282	Cadherin EGF LAG seven-pass G-type receptor 1 precursor (Flamingo homolog 2) (hFmi2) - Homo sapiens
NOV21e	SNP13382483	283	284	Cadherin EGF LAG seven-pass G-type receptor 1 precursor (Flamingo homolog 2) (hFmi2) - Homo sapiens
NOV21f	SNP13382484	285	286	Cadherin EGF LAG seven-pass G-type receptor 1 precursor (Flamingo homolog 2) (hFmi2) - Homo sapiens
NOV21g	SNP13382485	287	288	Cadherin EGF LAG seven-pass G-type receptor 1 precursor (Flamingo homolog 2) (hFmi2) - Homo sapiens
NOV22a	CG51983-05	289	290	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens
NOV22b	CG51983-01	291	292	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens
NOV22c	CG51983-02	293	294	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens
NOV22d	CG51983-03	295	296	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens
NOV22e	CG51983-04	297	298	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens
NOV22f	CG51983-06	299	300	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens
NOV22g	SNP13376585	301	302	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens
NOV22h	SNP13376586	303	304	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens
NOV23a	CG53390-02	305	306	Seven transmembrane helix receptor - Homo sapiens

NOV23b	CG53390-01	307	308	Seven transmembrane helix receptor - Homo sapiens
NOV23c	CG53390-03	309	310	Seven transmembrane helix receptor - Homo sapiens
NOV24a	CG53482-01	311	312	Human odorant receptor (OR)
NOV24b	CG53482-02	313	314	Human odorant receptor (OR)
NOV24c	CG53482-03	315	316	Human odorant receptor (OR)
NOV24d	SNP13373787	317	318	Human odorant receptor (OR)
NOV24e	SNP13373786	319	320	Human odorant receptor (OR)
NOV24f	SNP13373785	321	322	Human odorant receptor (OR)
NOV25a	CG53530-03	323	324	Seven transmembrane helix receptor - Homo sapiens
NOV25b	CG53530-01	325	326	Seven transmembrane helix receptor - Homo sapiens
NOV25c	CG53530-02	327	328	Seven transmembrane helix receptor - Homo sapiens
NOV26a	CG53563-03	329	330	Seven transmembrane helix receptor - Homo sapiens
NOV26b	CG53563-01	331	332	Seven transmembrane helix receptor - Homo sapiens
NOV26c	CG53563-02	333	334	Seven transmembrane helix receptor - Homo sapiens
NOV27a	CG53719-02	335	336	Olfactory receptor MOR106-5 - Mus musculus
NOV27b	CG53719-01	337	338	Olfactory receptor MOR106-5 - Mus musculus
NOV28a	CG53746-04	339	340	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens
NOV28b	CG53746-01	341	342	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens
NOV28c	CG53746-02	343	344	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens
NOV28d	CG53746-03	345	346	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens
NOV28e	SNP13373967	347	348	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens
NOV28f	SNP13373968	349	350	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens

NOV28g	SNP13382432	351	352	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens
NOV28h	SNP13373969	353	354	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens
NOV28i	SNP13373970	355	356	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens
NOV28j	SNP13373971	357	358	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens
NOV28k	SNP13373972	359	360	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens
NOV28l	SNP13382433	361	362	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens
NOV29a	CG53767-02	363	364	Seven transmembrane helix receptor - Homo sapiens
NOV29b	CG53767-01	365	366	Seven transmembrane helix receptor - Homo sapiens
NOV29c	SNP13382437	367	368	Seven transmembrane helix receptor - Homo sapiens
NOV29d	SNP13382436	369	370	Seven transmembrane helix receptor - Homo sapiens
NOV30a	CG53776-02	371	372	Olfactory receptor 5P2 (Olfactory receptor-like protein JCG3) - Homo sapiens
NOV30b	CG53776-03	373	374	Olfactory receptor 5P2 (Olfactory receptor-like protein JCG3) - Homo sapiens
NOV30c	CG53776-01	375	376	Olfactory receptor 5P2 (Olfactory receptor-like protein JCG3) - Homo sapiens
NOV31a	CG53803-02	377	378	Sequence 3 from Patent WO0159113 - Homo sapiens
NOV31b	CG53803-01	379	380	Sequence 3 from Patent WO0159113 - Homo sapiens
NOV32a	CG53989-03	381	382	Mast cell protease-11 - Mus musculus
NOV32b	CG53989-04	383	384	Mast cell protease-11 - Mus musculus
NOV32c	306076095	385	386	Mast cell protease-11 - Mus musculus
NOV32d	CG53989-01	387	388	Mast cell protease-11 - Mus musculus
NOV32e	CG53989-02	389	390	Mast cell protease-11 - Mus musculus

NOV32f	CG53989-05	391	392	Mast cell protease-11 - Mus musculus
NOV32g	CG53989-06	393	394	Mast cell protease-11 - Mus musculus
NOV32h	CG53989-07	395	396	Mast cell protease-11 - Mus musculus
NOV32i	CG53989-08	397	398	Mast cell protease-11 - Mus musculus
NOV32j	CG53989-09	399	400	Mast cell protease-11 - Mus musculus
NOV32k	CG53989-10	401	402	Mast cell protease-11 - Mus musculus
NOV32l	CG53989-11	403	404	Mast cell protease-11 - Mus musculus
NOV32m	CG53989-12	405	406	Mast cell protease-11 - Mus musculus
NOV32n	CG53989-13	407	408	Mast cell protease-11 - Mus musculus
NOV33a	CG54203-02	409	410	Olfactory receptor 2S2 - Homo sapiens
NOV33b	CG54203-01	411	412	Olfactory receptor 2S2 - Homo sapiens
NOV33c	SNP13382442	413	414	Olfactory receptor 2S2 - Homo sapiens
NOV33d	SNP13373880	415	416	Olfactory receptor 2S2 - Homo sapiens
NOV33e	SNP13373881	417	418	Olfactory receptor 2S2 - Homo sapiens
NOV33f	SNP13382443	419	420	Olfactory receptor 2S2 - Homo sapiens
NOV33g	SNP13373844	421	422	Olfactory receptor 2S2 - Homo sapiens
NOV33h	SNP13373883	423	424	Olfactory receptor 2S2 - Homo sapiens
NOV34a	CG54212-02	425	426	Seven transmembrane helix receptor - Homo sapiens
NOV34b	CG54212-01	427	428	Seven transmembrane helix receptor - Homo sapiens
NOV34c	CG54212-03	429	430	Seven transmembrane helix receptor - Homo sapiens
NOV34d	CG54212-04	431	432	Seven transmembrane helix receptor - Homo sapiens
NOV34e	SNP13373981	433	434	Seven transmembrane helix receptor - Homo sapiens
NOV34f	SNP13373982	435	436	Seven transmembrane helix receptor - Homo sapiens
NOV34g	SNP13373984	437	438	Seven transmembrane helix receptor - Homo sapiens

NOV34h	SNP13373985	439	440	Seven transmembrane helix receptor - Homo sapiens
NOV35a	CG54236-01	441	442	Cysteinyl leukotriene receptor 2 (CysLTR2) (PSEC0146) (HG57) (HPN321) (hGPCR21) - Homo sapiens
NOV35b	CG54236-02	443	444	Cysteinyl leukotriene receptor 2 (CysLTR2) (PSEC0146) (HG57) (HPN321) (hGPCR21) - Homo sapiens
NOV36a	CG54479-06	445	446	Macrophage stimulating 1 (Hepatocyte growth factor-like) - Homo sapiens
NOV36b	CG54479-05	447	448	Macrophage stimulating 1 (Hepatocyte growth factor-like) - Homo sapiens
NOV36c	CG54479-01	449	450	Macrophage stimulating 1 (Hepatocyte growth factor-like) - Homo sapiens
NOV36d	CG54479-02	451	452	Macrophage stimulating 1 (Hepatocyte growth factor-like) - Homo sapiens
NOV36e	CG54479-03	453	454	Macrophage stimulating 1 (Hepatocyte growth factor-like) - Homo sapiens
NOV36f	CG54479-04	455	456	Macrophage stimulating 1 (Hepatocyte growth factor-like) - Homo sapiens
NOV37a	CG54539-02	457	458	Zinc transporter 1 (ZnT-1) - Homo sapiens
NOV37b	CG54539-01	459	460	Zinc transporter 1 (ZnT-1) - Homo sapiens
NOV37c	SNP13382438	461	462	Zinc transporter 1 (ZnT-1) - Homo sapiens
NOV38a	CG54683-05	463	464	Gamma-aminobutyric-acid receptor rho-3 subunit precursor (GABA(A) receptor) - Rattus norvegicus
NOV38b	CG54683-01	465	466	Gamma-aminobutyric-acid receptor rho-3 subunit precursor (GABA(A) receptor) - Rattus norvegicus
NOV38c	CG54683-02	467	468	Gamma-aminobutyric-acid receptor rho-3 subunit precursor (GABA(A) receptor) - Rattus norvegicus

NOV38d	CG54683-03	469	470	Gamma-aminobutyric-acid receptor rho-3 subunit precursor (GABA(A) receptor) - Rattus norvegicus
NOV38e	CG54683-04	471	472	Gamma-aminobutyric-acid receptor rho-3 subunit precursor (GABA(A) receptor) - Rattus norvegicus
NOV39a	CG54692-06	473	474	Human G-protein coupled receptor (GCREC-15)
NOV39b	CG54692-01	475	476	Human G-protein coupled receptor (GCREC-15)
NOV39c	CG54692-02	477	478	Human G-protein coupled receptor (GCREC-15)
NOV39d	CG54692-03	479	480	Human G-protein coupled receptor (GCREC-15)
NOV39e	CG54692-04	481	482	Human G-protein coupled receptor (GCREC-15)
NOV39f	CG54692-05	483	484	Human G-protein coupled receptor (GCREC-15)
NOV40a	CG55069-01	485	486	Ten-m3 - Mus musculus
NOV40b	CG55069-04	487	488	Ten-m3 - Mus musculus
NOV40c	248993047	489	490	Ten-m3 - Mus musculus
NOV40d	262802488	491	492	Ten-m3 - Mus musculus
NOV40e	248993606	493	494	Ten-m3 - Mus musculus
NOV40f	314411758	495	496	Ten-m3 - Mus musculus
NOV40g	319067006	497	498	Ten-m3 - Mus musculus
NOV40h	319506086	499	500	Ten-m3 - Mus musculus
NOV40i	CG55069-03	501	502	Ten-m3 - Mus musculus
NOV40j	219937039	503	504	Ten-m3 - Mus musculus
NOV40k	219937046	505	506	Ten-m3 - Mus musculus
NOV40l	219937583	507	508	Ten-m3 - Mus musculus
NOV40m	219937000	509	510	Ten-m3 - Mus musculus
NOV40n	219937005	511	512	Ten-m3 - Mus musculus
NOV40o	219937013	513	514	Ten-m3 - Mus musculus
NOV40p	219937063	515	516	Ten-m3 - Mus musculus
NOV40q	CG55069-09	517	518	Ten-m3 - Mus musculus
NOV40r	309327410	519	520	Ten-m3 - Mus musculus
NOV40s	CG55069-02	521	522	Ten-m3 - Mus musculus
NOV40t	CG55069-05	523	524	Ten-m3 - Mus musculus
NOV40u	CG55069-06	525	526	Ten-m3 - Mus musculus
NOV40v	CG55069-07	527	528	Ten-m3 - Mus musculus
NOV40w	CG55069-08	529	530	Ten-m3 - Mus musculus
NOV40x	CG55069-10	531	532	Ten-m3 - Mus musculus

NOV40y	CG55069-11	533	534	Ten-m3 - Mus musculus
NOV40z	CG55069-12	535	536	Ten-m3 - Mus musculus
NOV40a	CG55069-13	537	538	Ten-m3 - Mus musculus
NOV40b	CG55069-14	539	540	Ten-m3 - Mus musculus
NOV40c	CG55069-15	541	542	Ten-m3 - Mus musculus
NOV40d	SNP13374479	543	544	Ten-m3 - Mus musculus
NOV40e	SNP13382453	545	546	Ten-m3 - Mus musculus
NOV40f	SNP13382454	547	548	Ten-m3 - Mus musculus
NOV40g	SNP13382455	549	550	Ten-m3 - Mus musculus
NOV40h	SNP13378354	551	552	Ten-m3 - Mus musculus
NOV41a	CG55343-03	553	554	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV41b	260568382	555	556	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV41c	314361391	557	558	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV41d	317286137	559	560	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV41e	CG55343-01	561	562	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV41f	CG55343-02	563	564	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV41g	CG55343-04	565	566	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV41h	CG55343-05	567	568	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV41i	CG55343-06	569	570	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV41j	SNP13373740	571	572	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV41k	SNP13376425	573	574	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens

NOV41l	SNP13376424	575	576	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV41m	SNP13376423	577	578	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens
NOV42a	CG55358-04	579	580	G-protein coupled receptor 4a (GPCR4a)
NOV42b	CG55358-03	581	582	G-protein coupled receptor 4a (GPCR4a)
NOV42c	317863291	583	584	G-protein coupled receptor 4a (GPCR4a)
NOV42d	317863328	585	586	G-protein coupled receptor 4a (GPCR4a)
NOV42e	317863350	587	588	G-protein coupled receptor 4a (GPCR4a)
NOV42f	271624076	589	590	G-protein coupled receptor 4a (GPCR4a)
NOV42g	CG55358-01	591	592	G-protein coupled receptor 4a (GPCR4a)
NOV42h	CG55358-02	593	594	G-protein coupled receptor 4a (GPCR4a)
NOV42i	CG55358-05	595	596	G-protein coupled receptor 4a (GPCR4a)
NOV42j	CG55358-06	597	598	G-protein coupled receptor 4a (GPCR4a)
NOV42k	CG55358-07	599	600	G-protein coupled receptor 4a (GPCR4a)
NOV42l	SNP13375122	601	602	G-protein coupled receptor 4a (GPCR4a)
NOV42m	SNP13375123	603	604	G-protein coupled receptor 4a (GPCR4a)
NOV42n	SNP13382494	605	606	G-protein coupled receptor 4a (GPCR4a)
NOV42o	SNP13375124	607	608	G-protein coupled receptor 4a (GPCR4a)
NOV42p	SNP13375125	609	610	G-protein coupled receptor 4a (GPCR4a)
NOV42q	SNP13376426	611	612	G-protein coupled receptor 4a (GPCR4a)
NOV42r	SNP13376427	613	614	G-protein coupled receptor 4a (GPCR4a)
NOV42s	SNP13375126	615	616	G-protein coupled receptor 4a (GPCR4a)

NOV43a	CG55604-04	617	618	Seven transmembrane helix receptor
NOV43b	CG55604-01	619	620	Seven transmembrane helix receptor
NOV43c	CG55604-02	621	622	Seven transmembrane helix receptor
NOV43d	CG55604-03	623	624	Seven transmembrane helix receptor
NOV43e	CG55604-05	625	626	Seven transmembrane helix receptor
NOV43f	CG55604-06	627	628	Seven transmembrane helix receptor
NOV43g	CG55604-07	629	630	Seven transmembrane helix receptor
NOV43h	SNP13019742	631	632	Seven transmembrane helix receptor
NOV43i	SNP13373776	633	634	Seven transmembrane helix receptor
NOV43j	SNP13373777	635	636	Seven transmembrane helix receptor
NOV43k	SNP13373778	637	638	Seven transmembrane helix receptor
NOV43l	SNP13373838	639	640	Seven transmembrane helix receptor
NOV43m	SNP13373837	641	642	Seven transmembrane helix receptor
NOV43n	SNP13373836	643	644	Seven transmembrane helix receptor
NOV43o	SNP13373780	645	646	Seven transmembrane helix receptor
NOV43p	SNP13373781	647	648	Seven transmembrane helix receptor
NOV43q	SNP13373782	649	650	Seven transmembrane helix receptor
NOV43r	SNP13373783	651	652	Seven transmembrane helix receptor
NOV43s	SNP13373833	653	654	Seven transmembrane helix receptor
NOV43t	SNP13373784	655	656	Seven transmembrane helix receptor
NOV43u	SNP13373832	657	658	Seven transmembrane helix receptor
NOV44a	CG55752-07	659	660	Glucosidase - Homo sapiens (Human), 769 aa
NOV44b	CG55752-06	661	662	Glucosidase - Homo sapiens (Human), 769 aa
NOV44c	CG55752-01	663	664	Glucosidase - Homo sapiens (Human), 769 aa
NOV44d	CG55752-02	665	666	Glucosidase - Homo sapiens (Human), 769 aa
NOV44e	CG55752-03	667	668	Glucosidase - Homo sapiens (Human), 769 aa
NOV44f	CG55752-04	669	670	Glucosidase - Homo sapiens (Human), 769 aa
NOV44g	CG55752-05	671	672	Glucosidase - Homo sapiens (Human), 769 aa
NOV44h	SNP13379656	673	674	Glucosidase - Homo sapiens (Human), 769 aa
NOV44i	SNP13379655	675	676	Glucosidase - Homo sapiens (Human), 769 aa
NOV44j	SNP13379654	677	678	Glucosidase - Homo sapiens (Human), 769 aa

NOV45a	CG55778-03	679	680	Aldo-keto reductase related protein 3 - Homo sapiens
NOV45b	CG55778-06	681	682	Aldo-keto reductase related protein 3 - Homo sapiens
NOV45c	275480984	683	684	Aldo-keto reductase related protein 3 - Homo sapiens
NOV45d	CG55778-01	685	686	Aldo-keto reductase related protein 3 - Homo sapiens
NOV45e	CG55778-02	687	688	Aldo-keto reductase related protein 3 - Homo sapiens
NOV45f	CG55778-04	689	690	Aldo-keto reductase related protein 3 - Homo sapiens
NOV45g	CG55778-05	691	692	Aldo-keto reductase related protein 3 - Homo sapiens
NOV45h	CG55778-07	693	694	Aldo-keto reductase related protein 3 - Homo sapiens
NOV45i	CG55778-08	695	696	Aldo-keto reductase related protein 3 - Homo sapiens
NOV45j	SNP13375813	697	698	Aldo-keto reductase related protein 3 - Homo sapiens
NOV46a	CG55794-03	699	700	Zn-carboxypeptidase - Homo sapiens
NOV46b	210223559	701	702	Zn-carboxypeptidase - Homo sapiens
NOV46c	210223626	703	704	Zn-carboxypeptidase - Homo sapiens
NOV46d	171095097	705	706	Zn-carboxypeptidase - Homo sapiens
NOV46e	183852229	707	708	Zn-carboxypeptidase - Homo sapiens
NOV46f	183852264	709	710	Zn-carboxypeptidase - Homo sapiens
NOV46g	183852410	711	712	Zn-carboxypeptidase - Homo sapiens
NOV46h	183523337	713	714	Zn-carboxypeptidase - Homo sapiens
NOV46i	CG55794-01	715	716	Zn-carboxypeptidase - Homo sapiens
NOV46j	CG55794-02	717	718	Zn-carboxypeptidase - Homo sapiens
NOV46k	CG55794-04	719	720	Zn-carboxypeptidase - Homo sapiens
NOV46l	CG55794-05	721	722	Zn-carboxypeptidase - Homo sapiens
NOV46m	CG55794-06	723	724	Zn-carboxypeptidase - Homo sapiens
NOV46n	CG55794-07	725	726	Zn-carboxypeptidase - Homo sapiens
NOV46o	CG55794-08	727	728	Zn-carboxypeptidase - Homo sapiens
NOV46p	CG55794-09	729	730	Zn-carboxypeptidase - Homo sapiens
NOV46q	CG55794-10	731	732	Zn-carboxypeptidase - Homo sapiens
NOV46r	CG55794-11	733	734	Zn-carboxypeptidase - Homo sapiens
NOV46s	CG55794-12	735	736	Zn-carboxypeptidase - Homo sapiens
NOV46t	CG55794-13	737	738	Zn-carboxypeptidase - Homo sapiens
NOV46u	SNP13375362	739	740	Zn-carboxypeptidase - Homo sapiens
NOV46v	SNP13379598	741	742	Zn-carboxypeptidase - Homo sapiens

NOV46w	SNP13375066	743	744	Zn-carboxypeptidase - Homo sapiens
NOV46x	SNP13375067	745	746	Zn-carboxypeptidase - Homo sapiens
NOV47a	CG55806-04	747	748	FACTOR IX - Homo sapiens
NOV47b	CG55806-01	749	750	FACTOR IX - Homo sapiens
NOV47c	CG55806-02	751	752	FACTOR IX - Homo sapiens
NOV47d	CG55806-03	753	754	FACTOR IX - Homo sapiens
NOV47e	SNP13382503	755	756	FACTOR IX - Homo sapiens
NOV47f	SNP13382492	757	758	FACTOR IX - Homo sapiens
NOV48a	CG55828-02	759	760	Serine/threonine-protein kinase PAK 7 (EC 2.7.1.-) (p21-activated kinase 7) (PAK-7) (PAK-5) - Homo sapiens
NOV48b	CG55828-01	761	762	Serine/threonine-protein kinase PAK 7 (EC 2.7.1.-) (p21-activated kinase 7) (PAK-7) (PAK-5) - Homo sapiens
NOV48c	SNP13379517	763	764	Serine/threonine-protein kinase PAK 7 (EC 2.7.1.-) (p21-activated kinase 7) (PAK-7) (PAK-5) - Homo sapiens
NOV48d	SNP13376535	765	766	Serine/threonine-protein kinase PAK 7 (EC 2.7.1.-) (p21-activated kinase 7) (PAK-7) (PAK-5) - Homo sapiens
NOV48e	SNP13382500	767	768	Serine/threonine-protein kinase PAK 7 (EC 2.7.1.-) (p21-activated kinase 7) (PAK-7) (PAK-5) - Homo sapiens
NOV48f	SNP13375705	769	770	Serine/threonine-protein kinase PAK 7 (EC 2.7.1.-) (p21-activated kinase 7) (PAK-7) (PAK-5) - Homo sapiens
NOV49a	CG55980-02	771	772	Seven transmembrane helix receptor – Homo sapiens
NOV49b	CG55980-01	773	774	Seven transmembrane helix receptor – Homo sapiens
NOV50a	CG55988-03	775	776	Organic cation transporter OKB1 – Homo sapiens
NOV50b	CG55988-04	777	778	Organic cation transporter OKB1 – Homo sapiens
NOV50c	CG55988-01	779	780	Organic cation transporter OKB1 – Homo sapiens
NOV50d	CG55988-02	781	782	Organic cation transporter OKB1 – Homo sapiens
NOV51a	CG56071-01	783	784	Mixed lineage kinase MLK1 - Homo sapiens (Human), 1066 aa
NOV51b	274082270	785	786	Mixed lineage kinase MLK1 - Homo sapiens (Human), 1066 aa
NOV51c	SNP13376041	787	788	Mixed lineage kinase MLK1 - Homo sapiens (Human), 1066 aa

NOV52a	CG56142-01	789	790	Similar to protease, serine, 8 (Prostasin) - Homo sapiens (Human), 327 aa
NOV52b	CG56142-04	791	792	Similar to protease, serine, 8 (Prostasin) - Homo sapiens (Human), 327 aa
NOV52c	276873337	793	794	Similar to protease, serine, 8 (Prostasin) - Homo sapiens (Human), 327 aa
NOV52d	276863970	795	796	Similar to protease, serine, 8 (Prostasin) - Homo sapiens (Human), 327 aa
NOV52e	276863992	797	798	Similar to protease, serine, 8 (Prostasin) - Homo sapiens (Human), 327 aa
NOV52f	276873330	799	800	Similar to protease, serine, 8 (Prostasin) - Homo sapiens (Human), 327 aa
NOV52g	CG56142-02	801	802	Similar to protease, serine, 8 (Prostasin) - Homo sapiens (Human), 327 aa
NOV52h	CG56142-03	803	804	Similar to protease, serine, 8 (Prostasin) - Homo sapiens (Human), 327 aa
NOV52i	CG56142-05	805	806	Similar to protease, serine, 8 (Prostasin) - Homo sapiens (Human), 327 aa
NOV52j	CG56142-06	807	808	Similar to protease, serine, 8 (Prostasin) - Homo sapiens (Human), 327 aa
NOV53a	CG56144-01	809	810	Seven transmembrane helix receptor – Homo sapiens
NOV53b	170645965	811	812	Seven transmembrane helix receptor – Homo sapiens
NOV53c	168869277	813	814	Seven transmembrane helix receptor – Homo sapiens
NOV53d	170645981	815	816	Seven transmembrane helix receptor – Homo sapiens
NOV53e	168869262	817	818	Seven transmembrane helix receptor – Homo sapiens
NOV53f	168869254	819	820	Seven transmembrane helix receptor – Homo sapiens
NOV53g	CG56144-02	821	822	Seven transmembrane helix receptor – Homo sapiens
NOV53h	CG56144-03	823	824	Seven transmembrane helix receptor –

				Homo sapiens
NOV53i	CG56144-04	825	826	Seven transmembrane helix receptor – Homo sapiens
NOV53j	CG56144-05	827	828	Seven transmembrane helix receptor – Homo sapiens
NOV53k	CG56144-06	829	830	Seven transmembrane helix receptor – Homo sapiens
NOV54a	CG56146-02	831	832	Seven transmembrane helix receptor – Homo sapiens
NOV54b	CG56146-01	833	834	Seven transmembrane helix receptor – Homo sapiens
NOV54c	170646057	835	836	Seven transmembrane helix receptor – Homo sapiens
NOV54d	170646049	837	838	Seven transmembrane helix receptor – Homo sapiens
NOV54e	170646053	839	840	Seven transmembrane helix receptor – Homo sapiens
NOV54f	174307717	841	842	Seven transmembrane helix receptor – Homo sapiens
NOV54g	168869383	843	844	Seven transmembrane helix receptor – Homo sapiens
NOV54h	CG56146-03	845	846	Seven transmembrane helix receptor – Homo sapiens
NOV54i	262939640	847	848	Seven transmembrane helix receptor – Homo sapiens
NOV54j	CG56146-04	849	850	Seven transmembrane helix receptor – Homo sapiens
NOV54k	CG56146-05	851	852	Seven transmembrane helix receptor – Homo sapiens
NOV54l	CG56146-06	853	854	Seven transmembrane helix receptor – Homo sapiens
NOV55a	CG56258-04	855	856	Na exchanger isoform 3 splice variant 2 - Homo sapiens
NOV55b	CG56258-02	857	858	Na exchanger isoform 3 splice variant 2 - Homo sapiens
NOV55c	258076220	859	860	Na exchanger isoform 3 splice variant 2 - Homo sapiens
NOV55d	248057963	861	862	Na exchanger isoform 3 splice variant 2 - Homo sapiens
NOV55e	CG56258-01	863	864	Na exchanger isoform 3 splice variant 2 - Homo sapiens
NOV55f	CG56258-03	865	866	Na exchanger isoform 3 splice variant 2 - Homo sapiens

NOV55g	CG56258-05	867	868	Na exchanger isoform 3 splice variant 2 - Homo sapiens
NOV55h	CG56258-06	869	870	Na exchanger isoform 3 splice variant 2 - Homo sapiens
NOV56a	CG56262-01	871	872	Putative calcium binding transporter - Homo sapiens
NOV56b	266120550	873	874	Putative calcium binding transporter - Homo sapiens
NOV57a	CG56315-01	875	876	Connexin25 - Homo sapiens
NOV57b	247679382	877	878	Connexin25 - Homo sapiens
NOV57c	247678321	879	880	Connexin25 - Homo sapiens
NOV57d	247679418	881	882	Connexin25 - Homo sapiens
NOV57e	247679395	883	884	Connexin25 - Homo sapiens
NOV57f	247679328	885	886	Connexin25 - Homo sapiens
NOV57g	CG56315-02	887	888	Connexin25 - Homo sapiens
NOV57h	CG56315-03	889	890	Connexin25 - Homo sapiens
NOV57i	CG56315-04	891	892	Connexin25 - Homo sapiens
NOV57j	CG56315-05	893	894	Connexin25 - Homo sapiens
NOV57k	CG56315-06	895	896	Connexin25 - Homo sapiens
NOV57l	CG56315-07	897	898	Connexin25 - Homo sapiens
NOV57m	CG56315-08	899	900	Connexin25 - Homo sapiens
NOV57n	SNP13381650	901	902	Connexin25 - Homo sapiens
NOV57o	SNP13381651	903	904	Connexin25 - Homo sapiens
NOV57p	SNP13381652	905	906	Connexin25 - Homo sapiens
NOV57q	SNP13381653	907	908	Connexin25 - Homo sapiens
NOV57r	SNP13381654	909	910	Connexin25 - Homo sapiens
NOV58a	CG56398-01	911	912	Sodium/glucose cotransporter KST1 - Homo sapiens
NOV58b	265726152	913	914	Sodium/glucose cotransporter KST1 - Homo sapiens
NOV58c	267254040	915	916	Sodium/glucose cotransporter KST1 - Homo sapiens
NOV58d	SNP13379242	917	918	Sodium/glucose cotransporter KST1 - Homo sapiens
NOV59a	CG56605-03	919	920	Protease PRTS-5
NOV59b	CG56605-01	921	922	Protease PRTS-5
NOV59c	CG56605-02	923	924	Protease PRTS-5
NOV60a	CG56645-03	925	926	sodium/glucose cotransporter Fbh68723pat
NOV60b	CG56645-04	927	928	sodium/glucose cotransporter Fbh68723pat

NOV60c	CG56645-01	929	930	sodium/glucose cotransporter Fbh68723pat
NOV60d	CG56645-02	931	932	sodium/glucose cotransporter Fbh68723pat
NOV61a	CG56667-01	933	934	Seven transmembrane helix receptor - Homo sapiens
NOV61b	268952113	935	936	Seven transmembrane helix receptor - Homo sapiens
NOV62a	CG56868-01	937	938	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens
NOV62b	276580332	939	940	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens
NOV63a	CG56870-01	941	942	NDRG3 protein - Homo sapiens
NOV63b	CG56870-06	943	944	NDRG3 protein - Homo sapiens
NOV63c	276585681	945	946	NDRG3 protein - Homo sapiens
NOV63d	CG56870-02	947	948	NDRG3 protein - Homo sapiens
NOV63e	CG56870-03	949	950	NDRG3 protein - Homo sapiens
NOV63f	CG56870-04	951	952	NDRG3 protein - Homo sapiens
NOV63g	CG56870-05	953	954	NDRG3 protein - Homo sapiens
NOV64a	CG57109-01	955	956	Human kinase (PKIN)-4
NOV64b	260457400	957	958	Human kinase (PKIN)-4
NOV64c	260457409	959	960	Human kinase (PKIN)-4
NOV64d	CG57109-05	961	962	Human kinase (PKIN)-4
NOV64e	267253965	963	964	Human kinase (PKIN)-4
NOV64f	267254000	965	966	Human kinase (PKIN)-4
NOV64g	267253987	967	968	Human kinase (PKIN)-4
NOV64h	CG57109-02	969	970	Human kinase (PKIN)-4
NOV64i	CG57109-03	971	972	Human kinase (PKIN)-4
NOV64j	CG57109-04	973	974	Human kinase (PKIN)-4
NOV64k	CG57109-06	975	976	Human kinase (PKIN)-4
NOV65a	CG57399-04	977	978	Human phospholipase-like enzyme
NOV65b	CG57399-01	979	980	Human phospholipase-like enzyme
NOV65c	CG57399-02	981	982	Human phospholipase-like enzyme
NOV65d	CG57399-03	983	984	Human phospholipase-like enzyme
NOV66a	CG57562-02	985	986	cation-transporting ATPase 2 (EC 3.6.3.-) (CGI-152) - Homo sapiens
NOV66b	CG57562-01	987	988	cation-transporting ATPase 2 (EC 3.6.3.-) (CGI-152) - Homo sapiens

NOV66c	SNP13380762	989	990	cation-transporting ATPase 2 (EC 3.6.3.-) (CGI-152) - Homo sapiens
NOV66d	SNP13380787	991	992	cation-transporting ATPase 2 (EC 3.6.3.-) (CGI-152) - Homo sapiens
NOV67a	CG57758-03	993	994	Na+coupled citrate transporter protein - Homo sapiens
NOV67b	308537854	995	996	Na+coupled citrate transporter protein - Homo sapiens
NOV67c	CG57758-01	997	998	Na+coupled citrate transporter protein - Homo sapiens
NOV67d	CG57758-02	999	1000	Na+coupled citrate transporter protein - Homo sapiens
NOV67e	CG57758-04	1001	1002	Na+coupled citrate transporter protein - Homo sapiens
NOV67f	CG57758-05	1003	1004	Na+coupled citrate transporter protein - Homo sapiens
NOV68a	CG58504-01	1005	1006	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV68b	169648376	1007	1008	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV68c	169648388	1009	1010	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV68d	169648365	1011	1012	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV68e	284068250	1013	1014	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV68f	305867866	1015	1016	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens

NOV68g	318176397	1017	1018	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV68h	CG58504-02	1019	1020	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV68i	CG58504-03	1021	1022	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV68j	CG58504-04	1023	1024	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV68k	CG58504-05	1025	1026	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV68l	CG58504-06	1027	1028	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV68m	CG58504-07	1029	1030	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens
NOV69a	CG58510-01	1031	1032	Cation transport protein YRDO
NOV70a	CG59309-01	1033	1034	Human Acyl-CoA thioesterase
NOV70b	278901386	1035	1036	Human Acyl-CoA thioesterase
NOV71a	CG59490-01	1037	1038	Mast cell protease-11 - Mus musculus
NOV71b	207639512	1039	1040	Mast cell protease-11 - Mus musculus
NOV71c	207639476	1041	1042	Mast cell protease-11 - Mus musculus
NOV71d	207639523	1043	1044	Mast cell protease-11 - Mus musculus
NOV71e	CG59490-02	1045	1046	Mast cell protease-11 - Mus musculus

NOV72a	CG59693-01	1047	1048	Aldo-keto reductase family 1 member C1 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (High-affinity hepatic bile acid-binding protein) (HBAB) (Chlordecone reductase homolog HAKRC) (Dihydrodiol dehydrogenase 2) (DD2) (20 alpha- hydroxysteroid dehydrogenase) - Homo sapiens
NOV72b	CG59693-03	1049	1050	Aldo-keto reductase family 1 member C1 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (High-affinity hepatic bile acid-binding protein) (HBAB) (Chlordecone reductase homolog HAKRC) (Dihydrodiol dehydrogenase 2) (DD2) (20 alpha- hydroxysteroid dehydrogenase) - Homo sapiens
NOV72c	277637252	1051	1052	Aldo-keto reductase family 1 member C1 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (High-affinity hepatic bile acid-binding protein) (HBAB) (Chlordecone reductase homolog HAKRC) (Dihydrodiol dehydrogenase 2) (DD2) (20 alpha- hydroxysteroid dehydrogenase) - Homo sapiens
NOV72d	CG59693-02	1053	1054	Aldo-keto reductase family 1 member C1 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (High-affinity hepatic bile acid-binding protein) (HBAB) (Chlordecone reductase homolog HAKRC) (Dihydrodiol dehydrogenase 2) (DD2) (20 alpha- hydroxysteroid dehydrogenase) - Homo sapiens

NOV72e	CG59693-04	1055	1056	Aldo-keto reductase family 1 member C1 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (High-affinity hepatic bile acid-binding protein) (HBAB) (Chlordecone reductase homolog HAKRC) (Dihydrodiol dehydrogenase 2) (DD2) (20 alpha- hydroxysteroid dehydrogenase) - Homo sapiens
NOV72f	CG59693-05	1057	1058	Aldo-keto reductase family 1 member C1 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (High-affinity hepatic bile acid-binding protein) (HBAB) (Chlordecone reductase homolog HAKRC) (Dihydrodiol dehydrogenase 2) (DD2) (20 alpha- hydroxysteroid dehydrogenase) - Homo sapiens
NOV72g	CG59693-06	1059	1060	Aldo-keto reductase family 1 member C1 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (High-affinity hepatic bile acid-binding protein) (HBAB) (Chlordecone reductase homolog HAKRC) (Dihydrodiol dehydrogenase 2) (DD2) (20 alpha- hydroxysteroid dehydrogenase) - Homo sapiens
NOV72h	CG59693-07	1061	1062	Aldo-keto reductase family 1 member C1 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (High-affinity hepatic bile acid-binding protein) (HBAB) (Chlordecone reductase homolog HAKRC) (Dihydrodiol dehydrogenase 2) (DD2) (20 alpha- hydroxysteroid dehydrogenase) - Homo sapiens

NOV72i	CG59693-08	1063	1064	Aldo-keto reductase family 1 member C1 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (High-affinity hepatic bile acid-binding protein) (HBAB) (Chlordecone reductase homolog HAKRC) (Dihydrodiol dehydrogenase 2) (DD2) (20 alpha- hydroxysteroid dehydrogenase) - Homo sapiens
NOV72j	CG59693-09	1065	1066	Aldo-keto reductase family 1 member C1 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (High-affinity hepatic bile acid-binding protein) (HBAB) (Chlordecone reductase homolog HAKRC) (Dihydrodiol dehydrogenase 2) (DD2) (20 alpha- hydroxysteroid dehydrogenase) - Homo sapiens
NOV73a	CG59839-02	1067	1068	cation-transporting ATPase 3 (EC 3.6.3.-) (ATPase family homolog up-regulated in senescence cells 1) - Homo sapiens
NOV73b	CG59839-01	1069	1070	cation-transporting ATPase 3 (EC 3.6.3.-) (ATPase family homolog up-regulated in senescence cells 1) - Homo sapiens
NOV74a	CG90866-04	1071	1072	Human protein kinase
NOV74b	CG90866-03	1073	1074	Human protein kinase
NOV74c	CG90866-01	1075	1076	Human protein kinase
NOV74d	CG90866-02	1077	1078	Human protein kinase
NOV74e	CG90866-05	1079	1080	Human protein kinase
NOV75a	CG91708-02	1081	1082	Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3) (MMP-3) (Transin-1) (SL-1) - Homo sapiens
NOV75b	262751856	1083	1084	Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3) (MMP-3) (Transin-1) (SL-1) - Homo sapiens
NOV75c	CG91708-01	1085	1086	Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3) (MMP-3) (Transin-1) (SL-1) - Homo sapiens

NOV75d	CG91708-03	1087	1088	Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3) (MMP-3) (Transin-1) (SL-1) - Homo sapiens
NOV75e	CG91708-04	1089	1090	Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3) (MMP-3) (Transin-1) (SL-1) - Homo sapiens
NOV75f	SNP13380740	1091	1092	Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3) (MMP-3) (Transin-1) (SL-1) - Homo sapiens
NOV76a	CG92078-02	1093	1094	Yolk SAC permease-like YSPL-1 form 1 (Yolk SAC permease-like YSPL-1 form 4) (Yolk SAC permease-like YSPL-1 form 3) (Yolk SAC permease-like YSPL-1 form 2) - Mus musculus
NOV76b	CG92078-01	1095	1096	Yolk SAC permease-like YSPL-1 form 1 (Yolk SAC permease-like YSPL-1 form 4) (Yolk SAC permease-like YSPL-1 form 3) (Yolk SAC permease-like YSPL-1 form 2) - Mus musculus
NOV77a	CG93669-04	1097	1098	NIMA-related protein kinase 3 - Homo sapiens
NOV77b	CG93669-01	1099	1100	NIMA-related protein kinase 3 - Homo sapiens
NOV77c	CG93669-02	1101	1102	NIMA-related protein kinase 3 - Homo sapiens
NOV77d	CG93669-03	1103	1104	NIMA-related protein kinase 3 - Homo sapiens
NOV77e	SNP13376464	1105	1106	NIMA-related protein kinase 3 - Homo sapiens
NOV77f	SNP13376462	1107	1108	NIMA-related protein kinase 3 - Homo sapiens
NOV77g	SNP13382521	1109	1110	NIMA-related protein kinase 3 - Homo sapiens
NOV78a	CG94235-01	1111	1112	Thymidylate kinase homologue
NOV78b	254647864	1113	1114	Thymidylate kinase homologue
NOV78c	254347797	1115	1116	Thymidylate kinase homologue
NOV78d	CG94235-02	1117	1118	Thymidylate kinase homologue
NOV79a	CG95175-01	1119	1120	Human EphA full length kinase
NOV79b	275697118	1121	1122	Human EphA full length kinase

NOV79c	275697150	1123	1124	Human EphA full length kinase
NOV80a	CG99638-01	1125	1126	Concentrative Na+-nucleoside cotransporter hCNT3
NOV81a	CG99650-01	1127	1128	Concentrative Na+-nucleoside cotransporter hCNT3
NOV81b	SNP13382525	1129	1130	Concentrative Na+-nucleoside cotransporter hCNT3
NOV81c	SNP13382526	1131	1132	Concentrative Na+-nucleoside cotransporter hCNT3

Table A indicates the homology of NOVX polypeptides to known protein families. Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table A will be useful in therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table A.

Pathologies, diseases, disorders and condition and the like that are associated with NOVX sequences include, but are not limited to, *e.g.*, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, metabolic disturbances associated with obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary Ostoeodystrophy, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers, as well as conditions such as transplantation and fertility.

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the

presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

Consistent with other known members of the family of proteins, identified in column 5 of Table A, the NOVX polypeptides of the present invention show homology to, and contain domains that are characteristic of, other members of such protein families. Details of the sequence relatedness and domain analysis for each NOVX are presented in Example A.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit diseases associated with the protein families listed in Table A.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are presented in Example C. Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the detection of a variety of diseases with differential expression in normal vs. diseased tissues, *e.g.*, detection of a variety of cancers. SNP analysis for each NOVX, if applicable, is presented in Example D.

Additional utilities for NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOVX clones

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

The NOVX genes and their corresponding encoded proteins are useful for preventing, treating or ameliorating medical conditions, *e.g.*, by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein

in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

The NOVX nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) a biological defense weapon.

In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 566; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 566, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 566; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 566 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and (e) a fragment of any of (a) through (d).

In another specific embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 566; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 566 wherein any

amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 566; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 566, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; (e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 566 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and (f) the complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 566; (b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 566 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 566; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 566 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the

invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

A NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product "mature" form arises, by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell (*e.g.*, host cell) in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probe", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), about 100 nt, or as many as

approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- stranded or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as used herein, is a nucleic acid that is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, *etc.*). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium, or of chemical precursors or other chemicals.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, or a complement of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template with appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis.

Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of a NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, is one that is sufficiently complementary to the nucleotide sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, that it can hydrogen bond with few or no mismatches to the nucleotide sequence shown in SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

A "fragment" provided herein is defined as a sequence of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, and is at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice.

A full-length NOVX clone is identified as containing an ATG translation start codon and an in-frame stop codon. Any disclosed NOVX nucleotide sequence lacking an ATG start codon therefore encodes a truncated C-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 5' direction of the disclosed sequence. Any disclosed NOVX nucleotide sequence lacking an in-frame stop codon similarly encodes a truncated N-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 3' direction of the disclosed sequence.

A "derivative" is a nucleic acid sequence or amino acid sequence formed from the native compounds either directly, by modification or partial substitution. An "analog" is a nucleic acid sequence or amino acid sequence that has a structure similar to, but not identical to, the native compound, *e.g.*, they differs from it in respect to certain components or side chains. Analogs may be synthetic or derived from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. A "homolog" is a nucleic acid sequence or amino acid sequence of a particular gene that is derived from different species.

Derivatives and analogs may be full length or other than full length. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.*, Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences include those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

A NOVX polypeptide is encoded by the open reading frame ("ORF") of a NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.*, from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25,

50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566; or an anti-sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566; or of a naturally occurring mutant of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe has a detectable label attached, *e.g.*, the label can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of a NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, that encodes a polypeptide having a NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566.

In addition to the human NOVX nucleotide sequences of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding a NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from a human SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 65% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low,

moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60 °C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of SEQ ID NO:$2n$-1, wherein n is an integer between 1 and 566, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Reinhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55 °C, followed by one or more washes in 1X SSC, 0.1% SDS at 37 °C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.*, Ausubel, et *al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Krieger, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* 78: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, thereby leading to changes in the amino acid sequences of the encoded NOVX protein, without altering the functional ability of that NOVX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:2*n*,

wherein n is an integer between 1 and 566. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are not particularly amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 40% homologous to the amino acid sequences of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NO:2*n*, wherein n is an integer between 1 and 566; more preferably at least about 70% homologous to SEQ ID NO:2*n*, wherein n is an integer between 1 and 566; still more preferably at least about 80% homologous to SEQ ID NO:2*n*, wherein n is an integer between 1 and 566; even more preferably at least about 90% homologous to SEQ ID NO:2*n*, wherein n is an integer between 1 and 566; and most preferably at least about 95% homologous to SEQ ID NO:2*n*, wherein n is an integer between 1 and 566.

An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced any one of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side

chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of a nucleic acid of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (*i*) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (*ii*) complex formation between a mutant NOVX protein and a NOVX ligand; or (*iii*) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (*e.g.*, avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (*e.g.*, regulation of insulin release).

Interfering RNA

In one aspect of the invention, NOVX gene expression can be attenuated by RNA interference. One approach well-known in the art is short interfering RNA (siRNA) mediated gene silencing where expression products of a NOVX gene are targeted by specific double stranded NOVX derived siRNA nucleotide sequences that are complementary to at least a 19-25 nt long segment of the NOVX gene transcript, including the 5' untranslated (UT) region, the ORF, or the 3' UT region. *See, e.g.*, PCT applications WO00/44895, WO99/32619, WO01/75164, WO01/92513, WO 01/29058, WO01/89304, WO02/16620, and WO02/29858, each incorporated by reference herein in their entirety. Targeted genes can be a NOVX gene, or an upstream or downstream modulator of the NOVX gene. Nonlimiting examples of upstream or downstream modulators of a NOVX gene include, *e.g.*, a transcription factor that binds the NOVX gene promoter, a kinase or phosphatase that interacts with a NOVX polypeptide, and polypeptides involved in a NOVX regulatory pathway.

According to the methods of the present invention, NOVX gene expression is silenced using short interfering RNA. A NOVX polynucleotide according to the invention includes a siRNA polynucleotide. Such a NOVX siRNA can be obtained using a NOVX polynucleotide sequence, for example, by processing the NOVX ribopolynucleotide sequence in a cell-free system, such as but not limited to a Drosophila extract, or by transcription of recombinant double stranded NOVX RNA or by chemical synthesis of nucleotide sequences homologous to a NOVX sequence. *See, e.g.*, Tuschl, Zamore, Lehmann, Bartel and Sharp (1999), Genes & Dev. 13: 3191-3197, incorporated herein by reference in its entirety. When synthesized, a typical 0.2 micromolar-scale RNA synthesis provides about 1 milligram of siRNA, which is sufficient for 1000 transfection experiments using a 24-well tissue culture plate format.

The most efficient silencing is generally observed with siRNA duplexes composed of a 21-nt sense strand and a 21-nt antisense strand, paired in a manner to have a 2-nt 3' overhang. The sequence of the 2-nt 3' overhang makes an additional small contribution to the specificity of siRNA target recognition. The contribution to specificity is localized to the unpaired nucleotide adjacent to the first paired bases. In one embodiment, the nucleotides in the 3' overhang are ribonucleotides. In an alternative embodiment, the nucleotides in the 3' overhang are deoxyribonucleotides.

Using 2'-deoxyribonucleotides in the 3' overhangs is as efficient as using ribonucleotides, but deoxyribonucleotides are often cheaper to synthesize and are most likely more nuclease resistant.

A contemplated recombinant expression vector of the invention comprises a NOVX DNA molecule cloned into an expression vector comprising operatively-linked regulatory sequences flanking the NOVX sequence in a manner that allows for expression (by transcription of the DNA molecule) of both strands. An RNA molecule that is antisense to NOVX mRNA is transcribed by a first promoter (*e.g.*, a promoter sequence 3' of the cloned DNA) and an RNA molecule that is the sense strand for the NOVX mRNA is transcribed by a second promoter (*e.g.*, a promoter sequence 5' of the cloned DNA). The sense and antisense strands may hybridize in vivo to generate siRNA constructs for silencing of the NOVX gene. Alternatively, two constructs can be utilized to create the sense and anti-sense strands of a siRNA construct. Finally, cloned DNA can encode a construct having secondary structure, wherein a single transcript has both the sense and complementary antisense sequences from the target gene or genes. In an example of this embodiment, a hairpin RNAi product is homologous to all or a portion of the target gene. In another example, a hairpin RNAi product is a siRNA. The regulatory sequences flanking the NOVX sequence may be identical or may be different, such that their expression may be modulated independently, or in a temporal or spatial manner.

In a specific embodiment, siRNAs are transcribed intracellularly by cloning the NOVX gene templates into a vector containing, *e.g.*, a RNA pol III transcription unit from the smaller nuclear RNA (snRNA) U6 or the human RNase P RNA H1. One example of a vector system is the GeneSuppressor™ RNA Interference kit (commercially available from Imgenex). The U6 and H1 promoters are members of the type III class of Pol III promoters. The +1 nucleotide of the U6-like promoters is always guanosine, whereas the +1 for H1 promoters is adenosine. The termination signal for these promoters is defined by five consecutive thymidines. The transcript is typically cleaved after the second uridine. Cleavage at this position generates a 3' UU overhang in the expressed siRNA, which is similar to the 3' overhangs of synthetic siRNAs. Any sequence less than 400 nucleotides in length can be transcribed by these promoter, therefore they are ideally suited for the expression of around 21-nucleotide siRNAs in, *e.g.*, an approximately 50-nucleotide RNA stem-loop transcript.

A siRNA vector appears to have an advantage over synthetic siRNAs where long term knock-down of expression is desired. Cells transfected with a siRNA expression vector would experience steady, long-term mRNA inhibition. In contrast, cells transfected with exogenous synthetic siRNAs typically recover from mRNA suppression within seven days or ten rounds of cell division. The long-term gene silencing ability of siRNA expression vectors may provide for applications in gene therapy.

In general, siRNAs are chopped from longer dsRNA by an ATP-dependent ribonuclease called DICER. DICER is a member of the RNase III family of double-stranded RNA-specific endonucleases. The siRNAs assemble with cellular proteins into an endonuclease complex. *In vitro* studies in Drosophila suggest that the siRNAs/protein complex (siRNP) is then transferred to a second enzyme complex, called an RNA-induced silencing complex (RISC), which contains an endoribonuclease that is distinct from DICER. RISC uses the sequence encoded by the antisense siRNA strand to find and destroy mRNAs of complementary sequence. The siRNA thus acts as a guide, restricting the ribonuclease to cleave only mRNAs complementary to one of the two siRNA strands.

A NOVX mRNA region to be targeted by siRNA is generally selected from a desired NOVX sequence beginning 50 to100 nt downstream of the start codon. Alternatively, 5' or 3' UTRs and regions nearby the start codon can be used but are generally avoided, as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. An initial BLAST homology search for the selected siRNA sequence is done against an available nucleotide sequence library to ensure that only one gene is targeted. Specificity of target recognition by siRNA duplexes indicate that a single point mutation located in the paired region of an siRNA duplex is sufficient to abolish target mRNA degradation. See, Elbashir *et al.* 2001 EMBO J. 20(23):6877-88. Hence, consideration should be taken to accommodate SNPs, polymorphisms, allelic variants or species-specific variations when targeting a desired gene.

In one embodiment, a complete NOVX siRNA experiment includes the proper negative control. A negative control siRNA generally has the same nucleotide composition as the NOVX siRNA but lack significant sequence homology to the

genome. Typically, one would scramble the nucleotide sequence of the NOVX siRNA and do a homology search to make sure it lacks homology to any other gene.

Two independent NOVX siRNA duplexes can be used to knock-down a target NOVX gene. This helps to control for specificity of the silencing effect. In addition, expression of two independent genes can be simultaneously knocked down by using equal concentrations of different NOVX siRNA duplexes, *e.g.*, a NOVX siRNA and an siRNA for a regulator of a NOVX gene or polypeptide. Availability of siRNA-associating proteins is believed to be more limiting than target mRNA accessibility.

A targeted NOVX region is typically a sequence of two adenines (AA) and two thymidines (TT) divided by a spacer region of nineteen (N19) residues (*e.g.*, AA(N19)TT). A desirable spacer region has a G/C-content of approximately 30% to 70%, and more preferably of about 50%. If the sequence AA(N19)TT is not present in the target sequence, an alternative target region would be AA(N21). The sequence of the NOVX sense siRNA corresponds to (N19)TT or N21, respectively. In the latter case, conversion of the 3' end of the sense siRNA to TT can be performed if such a sequence does not naturally occur in the NOVX polynucleotide. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3' overhangs. Symmetric 3' overhangs may help to ensure that the siRNPs are formed with approximately equal ratios of sense and antisense target RNA-cleaving siRNPs. *See, e.g.*, Elbashir, Lendeckel and Tuschl (2001). Genes & Dev. 15: 188-200, incorporated by reference herein in its entirely. The modification of the overhang of the sense sequence of the siRNA duplex is not expected to affect targeted mRNA recognition, as the antisense siRNA strand guides target recognition.

Alternatively, if the NOVX target mRNA does not contain a suitable AA(N21) sequence, one may search for the sequence NA(N21). Further, the sequence of the sense strand and antisense strand may still be synthesized as 5' (N19)TT, as it is believed that the sequence of the 3'-most nucleotide of the antisense siRNA does not contribute to specificity. Unlike antisense or ribozyme technology, the secondary structure of the target mRNA does not appear to have a strong effect on silencing. *See*, Harborth, *et al.* (2001) J. Cell Science 114: 4557-4565, incorporated by reference in its entirety.

Transfection of NOVX siRNA duplexes can be achieved using standard nucleic acid transfection methods, for example, OLIGOFECTAMINE Reagent (commercially available from Invitrogen). An assay for NOVX gene silencing is generally performed approximately 2 days after transfection. No NOVX gene silencing has been observed in the absence of transfection reagent, allowing for a comparative analysis of the wild-type and silenced NOVX phenotypes. In a specific embodiment, for one well of a 24-well plate, approximately 0.84 μg of the siRNA duplex is generally sufficient. Cells are typically seeded the previous day, and are transfected at about 50% confluence. The choice of cell culture media and conditions are routine to those of skill in the art, and will vary with the choice of cell type. The efficiency of transfection may depend on the cell type, but also on the passage number and the confluency of the cells. The time and the manner of formation of siRNA-liposome complexes (*e.g.*, inversion versus vortexing) are also critical. Low transfection efficiencies are the most frequent cause of unsuccessful NOVX silencing. The efficiency of transfection needs to be carefully examined for each new cell line to be used. Preferred cell are derived from a mammal, more preferably from a rodent such as a rat or mouse, and most preferably from a human. Where used for therapeutic treatment, the cells are preferentially autologous, although non-autologous cell sources are also contemplated as within the scope of the present invention.

For a control experiment, transfection of 0.84 μg single-stranded sense NOVX siRNA will have no effect on NOVX silencing, and 0.84 μg antisense siRNA has a weak silencing effect when compared to 0.84 μg of duplex siRNAs. Control experiments again allow for a comparative analysis of the wild-type and silenced NOVX phenotypes. To control for transfection efficiency, targeting of common proteins is typically performed, for example targeting of lamin A/C or transfection of a CMV-driven EGFP-expression plasmid (*e.g.*, commercially available from Clontech). In the above example, a determination of the fraction of lamin A/C knockdown in cells is determined the next day by such techniques as immunofluorescence, Western blot, Northern blot or other similar assays for protein expression or gene expression. Lamin A/C monoclonal antibodies may be obtained from Santa Cruz Biotechnology.

Depending on the abundance and the half life (or turnover) of the targeted NOVX polynucleotide in a cell, a knock-down phenotype may become apparent after 1 to 3 days, or even later. In cases where no NOVX knock-down phenotype is observed,

depletion of the NOVX polynucleotide may be observed by immunofluorescence or Western blotting. If the NOVX polynucleotide is still abundant after 3 days, cells need to be split and transferred to a fresh 24-well plate for re-transfection. If no knock-down of the targeted protein is observed, it may be desirable to analyze whether the target mRNA (NOVX or a NOVX upstream or downstream gene) was effectively destroyed by the transfected siRNA duplex. Two days after transfection, total RNA is prepared, reverse transcribed using a target-specific primer, and PCR-amplified with a primer pair covering at least one exon-exon junction in order to control for amplification of pre-mRNAs. RT/PCR of a non-targeted mRNA is also needed as control. Effective depletion of the mRNA yet undetectable reduction of target protein may indicate that a large reservoir of stable NOVX protein may exist in the cell. Multiple transfection in sufficiently long intervals may be necessary until the target protein is finally depleted to a point where a phenotype may become apparent. If multiple transfection steps are required, cells are split 2 to 3 days after transfection. The cells may be transfected immediately after splitting.

An inventive therapeutic method of the invention contemplates administering a NOVX siRNA construct as therapy to compensate for increased or aberrant NOVX expression or activity. The NOVX ribopolynucleotide is obtained and processed into siRNA fragments, or a NOVX siRNA is synthesized, as described above. The NOVX siRNA is administered to cells or tissues using known nucleic acid transfection techniques, as described above. A NOVX siRNA specific for a NOVX gene will decrease or knockdown NOVX transcription products, which will lead to reduced NOVX polypeptide production, resulting in reduced NOVX polypeptide activity in the cells or tissues.

The present invention also encompasses a method of treating a disease or condition associated with the presence of a NOVX protein in an individual comprising administering to the individual an RNAi construct that targets the mRNA of the protein (the mRNA that encodes the protein) for degradation. A specific RNAi construct includes a siRNA or a double stranded gene transcript that is processed into siRNAs. Upon treatment, the target protein is not produced or is not produced to the extent it would be in the absence of the treatment.

Where the NOVX gene function is not correlated with a known phenotype, a control sample of cells or tissues from healthy individuals provides a reference standard

for determining NOVX expression levels. Expression levels are detected using the assays described, *e.g.*, RT-PCR, Northern blotting, Western blotting, ELISA, and the like. A subject sample of cells or tissues is taken from a mammal, preferably a human subject, suffering from a disease state. The NOVX ribopolynucleotide is used to produce siRNA constructs, that are specific for the NOVX gene product. These cells or tissues are treated by administering NOVX siRNA's to the cells or tissues by methods described for the transfection of nucleic acids into a cell or tissue, and a change in NOVX polypeptide or polynucleotide expression is observed in the subject sample relative to the control sample, using the assays described. This NOVX gene knockdown approach provides a rapid method for determination of a NOVX minus ($NOVX^-$) phenotype in the treated subject sample. The $NOVX^-$ phenotype observed in the treated subject sample thus serves as a marker for monitoring the course of a disease state during treatment.

In specific embodiments, a NOVX siRNA is used in therapy. Methods for the generation and use of a NOVX siRNA are known to those skilled in the art. Example techniques are provided below.

Production of RNAs

Sense RNA (ssRNA) and antisense RNA (asRNA) of NOVX are produced using known methods such as transcription in RNA expression vectors. In the initial experiments, the sense and antisense RNA are about 500 bases in length each. The produced ssRNA and asRNA (0.5 μM) in 10 mM Tris-HCl (pH 7.5) with 20 mM NaCl were heated to 95° C for 1 min then cooled and annealed at room temperature for 12 to 16 h. The RNAs are precipitated and resuspended in lysis buffer (below). To monitor annealing, RNAs are electrophoresed in a 2% agarose gel in TBE buffer and stained with ethidium bromide. See, *e.g.*, Sambrook et al., Molecular Cloning. Cold Spring Harbor Laboratory Press, Plainview, N.Y. (1989).

Lysate Preparation

Untreated rabbit reticulocyte lysate (Ambion) are assembled according to the manufacturer's directions. dsRNA is incubated in the lysate at 30° C for 10 min prior to the addition of mRNAs. Then NOVX mRNAs are added and the incubation continued for an additional 60 min. The molar ratio of double stranded RNA and mRNA is about

200:1. The NOVX mRNA is radiolabeled (using known techniques) and its stability is monitored by gel electrophoresis.

In a parallel experiment made with the same conditions, the double stranded RNA is internally radiolabeled with a ^{32}P-ATP. Reactions are stopped by the addition of 2X-proteinase-K buffer and deproteinized as described previously (Tuschl *et al.*, Genes Dev., 13:3191-3197 (1999)). Products are analyzed by electrophoresis in 15% or 18% polyacrylamide sequencing gels using appropriate RNA standards. By monitoring the gels for radioactivity, the natural production of 10 to 25 nt RNAs from the double stranded RNA can be determined.

The band of double stranded RNA, about 21-23 bps, is eluded. The efficacy of these 21-23 mers for suppressing NOVX transcription is assayed in vitro using the same rabbit reticulocyte assay described above using 50 nanomolar of double stranded 21-23 mer for each assay. The sequence of these 21-23 mers is then determined using standard nucleic acid sequencing techniques.

RNA Preparation

21 nt RNAs, based on the sequence determined above, are chemically synthesized using Expedite RNA phosphoramidites and thymidine phosphoramidite (Proligo, Germany). Synthetic oligonucleotides are deprotected and gel-purified (Elbashir, Lendeckel, & Tuschl, Genes & Dev. 15, 188-200 (2001)), followed by Sep-Pak C18 cartridge (Waters, Milford, Mass., USA) purification (Tuschl, et al., Biochemistry, 32:11658-11668 (1993)).

These RNAs (20 μM) single strands are incubated in annealing buffer (100 mM potassium acetate, 30 mM HEPES-KOH at pH 7.4, 2 mM magnesium acetate) for 1 min at 90° C followed by 1 h at 37° C.

Cell Culture

A cell culture known in the art to regularly express NOVX is propagated using standard conditions. 24 hours before transfection, at approx. 80% confluency, the cells are trypsinized and diluted 1:5 with fresh medium without antibiotics (1-3 X 105 cells/ml) and transferred to 24-well plates (500 ml/well). Transfection is performed using a commercially available lipofection kit and NOVX expression is monitored using standard techniques with positive and negative control. A positive control is cells that naturally express NOVX while a negative control is cells that do not express NOVX.

Base-paired 21 and 22 nt siRNAs with overhanging 3' ends mediate efficient sequence-specific mRNA degradation in lysates and in cell culture. Different concentrations of siRNAs are used. An efficient concentration for suppression in vitro in mammalian culture is between 25 nM to 100 nM final concentration. This indicates that siRNAs are effective at concentrations that are several orders of magnitude below the concentrations applied in conventional antisense or ribozyme gene targeting experiments.

The above method provides a way both for the deduction of NOVX siRNA sequence and the use of such siRNA for in vitro suppression. In vivo suppression may be performed using the same siRNA using well known in-vivo transfection or gene therapy transfection techniques.

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566, or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers

to 5' and 3' sequences that flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-carboxymethylaminomethyl-2-thiouridine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 5-methoxyuracil, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, 2-thiouracil, 4-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense

orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other. *See, e.g.,* Gaultier, *et al.*, 1987. *Nucl. Acids Res.* 15: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g.,* Inoue, *et al.* 1987. *Nucl. Acids Res.* 15: 6131-6148) or a chimeric RNA-DNA analogue (*See, e.g.,* Inoue, *et al.*, 1987. *FEBS Lett.* 215: 327-330.

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of

the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX cDNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleotide bases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomer can be performed using standard solid

phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S_1 nucleases (*See,* Hyrup, *et al.*, 1996.*supra*); or as probes or primers for DNA sequence and hybridization (*See,* Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.*, 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleotide bases, and orientation (*see,* Hyrup, et al., 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)-amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad.*

Sci. U.S.A. 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in any one of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in any one of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, a NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of a NOVX protein can be

a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO:2*n*, wherein n is an integer between 1 and 566, and retains the functional activity of the protein of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566, and retains the functional activity of the NOVX proteins of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See,* Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity

preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a NOVX protein of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically-active portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at least two biologically-active portions of a NOVX protein. In yet another embodiment, a NOVX fusion protein comprises at least three biologically-active portions of a NOVX protein. Within the fusion protein, the term

"operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is a NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers.

Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.,* Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.,* mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.,* mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a

variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.,* Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in

combination with the screening assays to identify NOVX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

Anti-NOVX Antibodies

Included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated protein of the invention intended to serve as an antigen, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence of SEQ ID NO:2*n*, wherein n is an integer between 1 and 566, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, *e.g.*,

a hydrophilic region. A hydrophobicity analysis of the human NOVX protein sequence will indicate which regions of a NOVX polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. A NOVX polypeptide or a fragment thereof comprises at least one antigenic epitope. An anti-NOVX antibody of the present invention is said to specifically bind to antigen NOVX when the equilibrium binding constant (K_D) is ≤ 1 μM, preferably $\leq$ 100 nM, more preferably $\leq$ 10 nM, and most preferably $\leq$ 100 pM to about 1 pM, as measured by assays such as radioligand binding assays or similar assays known to those skilled in the art.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, *etc.*), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen that is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining

regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods (Goding,1986). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (*e.g.*, by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide.

Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

Human Antibodies

Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human

antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals. For example, mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al,(Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and

is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see *e.g.*, U.S.

Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see *e.g.*, Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain,

are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers that are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (*e.g.*, tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (*e.g.*, alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (*e.g.*, $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2

receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker that is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (*e.g.*, CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, *e.g.*, the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (*e.g.*, an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (*e.g.*, avidin) that is in turn conjugated to a cytotoxic agent.

Immunoliposomes

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al.,

Proc. Natl Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin et al., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon *et al.*, J. National Cancer Inst., 81(19): 1484 (1989).

Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme linked immunosorbent assay (ELISA) and other immunologically mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Antibodies directed against a NOVX protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of a NOVX protein (*e.g.*, for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies specific to a NOVX protein, or derivative, fragment, analog or homolog thereof, that contain the antibody derived antigen binding domain, are utilized as pharmacologically active compounds (referred to hereinafter as "Therapeutics").

An antibody specific for a NOVX protein of the invention (*e.g.*, a monoclonal antibody or a polyclonal antibody) can be used to isolate a NOVX polypeptide by standard techniques, such as immunoaffinity, chromatography or immunoprecipitation.

An antibody to a NOVX polypeptide can facilitate the purification of a natural NOVX antigen from cells, or of a recombinantly produced NOVX antigen expressed in host cells. Moreover, such an anti-NOVX antibody can be used to detect the antigenic NOVX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic NOVX protein. Antibodies directed against a NOVX protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibody Therapeutics

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand that may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York.

If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain

of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, *e.g.*, Marasco et al., Proc. Natl. Acad. Sci. USA, 90: 7889-7893 (1993). The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for in vivo administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, *e.g.*, films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and γ ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT ™ (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

ELISA Assay

An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, F_{ab} or $F_{(ab)2}$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of an analyte mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of an analyte genomic DNA include Southern hybridizations. Procedures for conducting immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulus, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Theory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, *in vivo* techniques for detection of an analyte protein include introducing into a subject a labeled anti-an analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, useful expression vectors in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory

sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, *etc.* The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, *etc.*).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse

glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.,* Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.,* Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*,

MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, *etc.* A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a

pseudopregnant female foster animal. The human NOVX cDNA sequences, *i.e.*, any one of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of any one of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby

alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.,* Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *see, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae. See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a

transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation),

transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent that delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a NOVX protein or anti-NOVX antibody) in the required amount in an

appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser that contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.,* U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease (possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds that bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in

combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a NOVX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound

with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention. In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.*, a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The

target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.*, intracellular Ca^{2+}, diacylglycerol, IP_3, *etc.*), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to a NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the

activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate a NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione

derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of

NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue typing); and (*iii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX, sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific

human chromosomes. *See, e.g.,* D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see,* Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.,* in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can

then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each

500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If coding sequences, such as those of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in a NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1, wherein n is an integer between 1 and 566, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody

using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration

affecting the integrity of a gene encoding a NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from a NOVX gene; (*ii*) an addition of one or more nucleotides to a NOVX gene; (*iii*) a substitution of one or more nucleotides of a NOVX gene, (*iv*) a chromosomal rearrangement of a NOVX gene; (*v*) an alteration in the level of a messenger RNA transcript of a NOVX gene, (*vi*) aberrant modification of a NOVX gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a NOVX gene, (*viii*) a non-wild-type level of a NOVX protein, (*ix*) allelic loss of a NOVX gene, and (*x*) inappropriate post-translational modification of a NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional

amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al,* 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.,* U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.,* Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al., supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560

or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.,* Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.,* Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on a NOVX sequence, *e.g.*, a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch

repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.,* 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.,* 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.,* 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.,* Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.,* 1986. *Nature* 324: 163; Saiki, *et al.,* 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of

different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,* Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.*, 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.*, NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or

conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome pregnancy zone protein precursor enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the

population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX

and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of a NOVX protein, mRNA, or genomic DNA in the preadministration sample; (*iii*) obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

These methods of treatment will be discussed more fully, below.

Diseases and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, *etc.*) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, a NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced

into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situ*ations in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders. The

disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from diseases, disorders, conditions and the like, including but not limited to those listed herein.

Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example A: Polynucleotide and Polypeptide Sequences, and Homology Data

Example 1.

The NOV1 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 1A.

Table 1A. NOV1 Sequence Analysis

NOV1a, CG101340-01 DNA Sequence	SEQ ID NO: 1	1327 bp
	ORF Start: ATG at 76	ORF Stop: TGA at 1192

GAGGTGAATGCCATGCCATGATTCTGGTGTGCTCCATGGCATCCCCAGCCTAGCTCCCAATCCCACTT
TGGCACGATGTTAGCCAACAGCTCCTCAACCAACAGTTCTGTTCTCCCGTGTCCTGACTACCGACCTA
CCCACCGCCTGCACTTGGTGGTCTACAGCTTGGTGCTGGCTGCCGGGCTCCCCCTCAACGCGCTAGCC
CTCTGGGTCTTCCTGCGCGCGCTGCGCGTGCACTCGGTGGTGAGCGTGTACATGTGTAACCTGGCGGC
CAGCGACCTGCTCTTCACCCTCTCGCTGCCCGTTCGTCTCTCCTACTACGCACTGCACCACTGGCCCT
TCCCCGACCTCCTGTGCCAGACGACGGGCGCCATCTTCCAGATGAACATGTACGGCAGCTGCATCTTC
CTGATGCTCATCAACGTGGACCGCTACGCCGCCATCGTGCACCCGCTGCGACTGCGCCACCTGTGGCG
GCCCCGCGTGGCGCGGCTGCTCTACCTGGGCGTGTGGGCGCTCATCCTGGTGTTTGCCGTGCCCGCCG
CCCGCGTGCACAGGCCCTCGCGTTGCCGCTACCGGGACCTCGAGGTGCGCCTATGCTTCGAGAGCTTC
TGCGACGAGCTGTGGAAAGGCAGGCTGCTGCCCCTCGTGCTGCTGGCCGAGGCGCTGGGCTTCCTGCT
GCCCCTGGCGGCGGTGGTCTACTCGTCGGGCCGAGTCTTCTGGACGCTGGCGCGCCCCGACGCCACGC
AGAGCCAGCGGCGGTGGAAGACCGTGCGCCTCCTGCTGGCTAACCTCGTCATCTTCCTGCTGTGCTTC
GTGCCCTACAACAGCACGCTGGCGGTCTACGGGCTGCTGCGGAGCAAGCTGGTGGCGGCCAGCGTGCC
TGCCCGCGATCGCGTGCGCGGGGTGCTGATGGTGATGGTGCTGCTGGCCGGCGCCAACTGCGTGCTGG
ACCCGCTGGTGTACTACTTTAGCGCCGAGGGCTTCCGCAACACCCTGCGCGGCCTGGGCACTCCGCAC
CGGGCCAGGACCTCGGCCACCAACGGGACGCGGGCGGCGCTCGCGCAATCCGAAAGGTCCGCCGTCAC
CACCGACGCCACCAGGCCGGATGCCGCCAGTCAGGGGCTGCTCCGACCCTCCGACTCCCACTCTCTGT
CTTCCTTCACACAGTGTCCCCAGGATTCCGCCCTC**TGA**ACACACATGCCATTGCGCTGTCCGTGCCCG
ACTCCCAACGCCTCTCGTTCTGGGAGGCTTACAGGGCGTACACACAAGAAGGTGGGCTGGGCACTTGG
ACCTTTGGGTGGCAATTCCAGCTTAGCAACGCAGA

NOV1a, CG101340-01 Protein Sequence	SEQ ID NO: 2	372 aa	MW at 41482.1kD

MLANSSSTNSSVLPCPDYRPTHRLHLVVYSLVLAAGLPLNALALWVFLRALRVHSVVSVYMCNLAASD
LLFTLSLPVRLSYYALHHWPFPDLLCQTTGAIFQMNMYGSCIFLMLINVDRYAAIVHPLRLRHLWRPR
VARLLYLGVWALILVFAVPAARVHRPSRCRYRDLEVRLCFESFCDELWKGRLLPLVLLAEALGFLLPL
AAVVYSSGRVFWTLARPDATQSQRRWKTVRLLLANLVIFLLCFVPYNSTLAVYGLLRSKLVAASVPAR
DRVRGVLMVMVLLAGANCVLDPLVYYFSAEGFRNTLRGLGTPHRARTSATNGTRAALAQSERSAVTTD
ATRPDAASQGLLRPSDSHSLSSFTQCPQDSAL

NOV1b, SNP13382465 of CG101340-01, DNA Sequence	SEQ ID NO: 3	1327 bp
	ORF Start: ATG at 76	ORF Stop: TGA at 1192
	SNP Pos: 472	SNP Change: T to C

GAGGTGAATGCCATGCCATGATTCTGGTGTGCTCCATGGCATCCCCAGCCTAGCTCCCAATCCCACTT
TGGCACGATGTTAGCCAACAGCTCCTCAACCAACAGTTCTGTTCTCCCGTGTCCTGACTACCGACCTA
CCCACCGCCTGCACTTGGTGGTCTACAGCTTGGTGCTGGCTGCCGGGCTCCCCCTCAACGCGCTAGCC
CTCTGGGTCTTCCTGCGCGCGCTGCGCGTGCACTCGGTGGTGAGCGTGTACATGTGTAACCTGGCGGC
CAGCGACCTGCTCTTCACCCTCTCGCTGCCCGTTCGTCTCTCCTACTACGCACTGCACCACTGGCCCT
TCCCCGACCTCCTGTGCCAGACGACGGGCGCCATCTTCCAGATGAACATGTACGGCAGCTGCATCTTC
CTGATGCTCATCAACGTGGACCGCTACGCCGCCATCGTGCACCCGCTGCGACTGCGCCACCTGCGGCG
GCCCCGCGTGGCGCGGCTGCTCTACCTGGGCGTGTGGGCGCTCATCCTGGTGTTTGCCGTGCCCGCCG
CCCGCGTGCACAGGCCCTCGCGTTGCCGCTACCGGGACCTCGAGGTGCGCCTATGCTTCGAGAGCTTC
TGCGACGAGCTGTGGAAAGGCAGGCTGCTGCCCCTCGTGCTGCTGGCCGAGGCGCTGGGCTTCCTGCT
GCCCCTGGCGGCGGTGGTCTACTCGTCGGGCCGAGTCTTCTGGACGCTGGCGCGCCCCGACGCCACGC
AGAGCCAGCGGCGGTGGAAGACCGTGCGCCTCCTGCTGGCTAACCTCGTCATCTTCCTGCTGTGCTTC
GTGCCCTACAACAGCACGCTGGCGGTCTACGGGCTGCTGCGGAGCAAGCTGGTGGCGGCCAGCGTGCC
TGCCCGCGATCGCGTGCGCGGGGTGCTGATGGTGATGGTGCTGCTGGCCGGCGCCAACTGCGTGCTGG
ACCCGCTGGTGTACTACTTTAGCGCCGAGGGCTTCCGCAACACCCTGCGCGGCCTGGGCACTCCGCAC
CGGGCCAGGACCTCGGCCACCAACGGGACGCGGGCGGCGCTCGCGCAATCCGAAAGGTCCGCCGTCAC
CACCGACGCCACCAGGCCGGATGCCGCCAGTCAGGGGCTGCTCCGACCCTCCGACTCCCACTCTCTGT
CTTCCTTCACACAGTGTCCCCAGGATTCCGCCCTCTGAACACACATGCCATTGCGCTGTCCGTGCCCG
ACTCCCAACGCCTCTCGTTCTGGGAGGCTTACAGGGCGTACACACAAGAAGGTGGGCTGGGCACTTGG
ACCTTTGGGTGGCAATTCCAGCTTAGCAACGCAGA

NOV1b, SNP13382465 of CG101340-01, Protein Sequence	SEQ ID NO: 4	372 aa	MW at 41452.1kD
	SNP Pos: 133		SNP Change: Trp to Arg

MLANSSSTNSSVLPCPDYRPTHRLHLVVYSLVLAAGLPLNALALWVFLRALRVHSVVSVYMCNLAASD
LLFTLSLPVRLSYYALHHWPFPDLLCQTTGAIFQMNMYGSCIFLMLINVDRYAAIVHPLRLRHLRRPR
VARLLYLGVWALILVFAVPAARVHRPSRCRYRDLEVRLCFESFCDELWKGRLLPLVLLAEALGFLLPL
AAVVYSSGRVFWTLARPDATQSQRRWKTVRLLLANLVIFLLCFVPYNSTLAVYGLLRSKLVAASVPAR
DRVRGVLMVMVLLAGANCVLDPLVYYFSAEGFRNTLRGLGTPHRARTSATNGTRAALAQSERSAVTTD
ATRPDAASQGLLRPSDSHSLSSFTQCPQDSAL

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 1B.

Table 1B. Comparison of the NOV1 protein sequences.

```
NOV18a  MCSLPMARYYRIKYADQKALYTRDGQLLVGDPVADNCCAEKICILPNRGLARTKVPIFLG
NOV18b  MCSLPMARYYIIKYADQKALYTRDGQLLVGDPVADNCCAEKICILPNRGLARTKVPIFLG
NOV18c  -----LSYCFRIKYADQKALYTRDGQLLVGDPVADNCCAEKICILPNRGLARTKVPIFLG

NOV18a  IQGGSRCLACVETEEGPSLQLEPSTLPPQDVNIEELYKGGEEATRFTFFQSSSGSAFRLE
NOV18b  IQGGSRCLACVETEEGPSLQLE-------DVNIEELYKGGEEATRFTFFQSSSGSAFRLE
```

```
NOV18c  IQGGSRCLACVETEEGPSLQLEPSTLPPQDVNIEELYKGGEEATRFTFFQSSSGSAFRLE

NOV18a  AAAWPGWFLCGPAEPQQPVQLTKESEPSARTKFYFEQSW
NOV18b  AAAWPGWFLCGPAEPQQPVQLTKESEPSARTKFYFEQSW
NOV18c  AAAWPGWFLCGPAEPQQPVQLTKESEPSARTKFYFEQSW

NOV1a  (SEQ ID NO:  2)
```

Further analysis of the NOV1a protein yielded the following properties shown in Table 1C.

Table 1C. Protein Sequence Properties NOV1a

SignalP analysis:	Cleavage site between residues 59 and 60

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 16;  peak value   3.12
      PSG score:  -1.28

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   0.37
      possible cleavage site: between 43 and 44

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood = -4.19   Transmembrane   31 -  47
      INTEGRAL    Likelihood = -0.16   Transmembrane   99 - 115
      INTEGRAL    Likelihood = -7.01   Transmembrane  140 - 156
      INTEGRAL    Likelihood = -6.90   Transmembrane  192 - 208
      INTEGRAL    Likelihood = -7.54   Transmembrane  232 - 248
      INTEGRAL    Likelihood = -3.82   Transmembrane  277 - 293
      PERIPHERAL  Likelihood =  1.43 (at 56)
      ALOM score:  -7.54  (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 38
      Charge difference:  0.5   C( 2.5) - N( 2.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  4.95
      Hyd Moment(95):  3.49    G content:       0
      D/E content:     1       S/T content:     6
      Score: -3.65

Gavel: prediction of cleavage sites for mitochondrial preseq
```

```
      R-2 motif at 176  CRY|RD

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  10.2%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       11.1 %: Golgi
```

```
        11.1 %: vacuolar
        11.1 %: vesicles of secretory system
        11.1 %: mitochondrial

>> prediction for CG101340-01 is end (k=9)
```

A search of the NOV1a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 1D.

Table 1D. Geneseq Results for NOV1a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV1a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABP81962	Human G protein-coupled receptor GPR92/3 protein SEQ ID NO:410 - Homo sapiens, 372 aa. [WO200261087-A2, 08-AUG-2002]	1..372 1..372	368/372 (98%) 368/372 (98%)	0.0
AAM52650	Human G protein-coupled receptor TGR4 - Homo sapiens, 372 aa. [WO200177326-A1, 18-OCT-2001]	1..372 1..372	368/372 (98%) 368/372 (98%)	0.0
AAU11897	Human novel G protein-coupled receptor, GPCR9 - Homo sapiens, 372 aa. [WO200190187-A2, 29-NOV-2001]	1..372 1..372	368/372 (98%) 368/372 (98%)	0.0
AAE16170	Human G-protein coupled receptor 1 (GCREC-1) protein - Homo sapiens, 372 aa. [WO200187937-A2, 22-NOV-2001]	1..372 1..372	368/372 (98%) 368/372 (98%)	0.0
AAU97915	Human HIPHUM0000001 purinergic receptor protein - Homo sapiens, 372 aa. [GB2360523-A, 26-SEP-2001]	1..372 1..372	368/372 (98%) 368/372 (98%)	0.0

In a BLAST search of public sequence databases, the NOV1a protein was found to have homology to the proteins shown in the BLASTP data in Table 1E.

Table 1E. Public BLASTP Results for NOV1a				
Protein Accession Number	**Protein/Organism/Length**	**NOV1a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9H1C0	Putative G protein-coupled receptor 92 (Putative G-protein coupled receptor) - Homo sapiens (Human), 372 aa.	1..372 1..372	368/372 (98%) 368/372 (98%)	0.0
CAD38444	Sequence 10 from Patent WO0238607 - Mus musculus (Mouse), 488 aa.	3..372 123..488	294/370 (79%) 316/370 (84%)	e-164
P32250	P2Y purinoceptor 5 (P2Y5) (Purinergic receptor 5) (6H1) - Gallus gallus (Chicken), 308 aa.	22..302 14..293	111/284 (39%) 169/284 (59%)	1e-51
Q8BMC0	P2Y purinoceptor 5 - Mus musculus (Mouse), 344 aa.	7..309 3..303	113/306 (36%) 178/306 (57%)	1e-49
P43657	P2Y purinoceptor 5 (P2Y5) (Purinergic receptor 5) (RB intron encoded G-protein coupled receptor) - Homo sapiens (Human), 344 aa.	22..309 17..303	106/291 (36%) 171/291 (58%)	1e-49

PFam analysis indicates that the NOV1a protein contains the domains shown in the Table 1F.

Table 1F. Domain Analysis of NOV1a			
Pfam Domain	**NOV1a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	39..297	87/276 (32%) 196/276 (71%)	7.2e-53

Example 2.

The NOV2 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 2A.

5

Table 2A. NOV2 Sequence Analysis

NOV2a, CG101396-01 DNA Sequence	SEQ ID NO: 5	3189 bp
	ORF Start: ATG at 86	ORF Stop: TGA at 3113

CTGCACCGGGACCAGCGCCTCCCCGCTTCGCGCTGCCCTCGGCCTCGCCCCGGGCCCGGGTGGATGAG
CCGCGCGCCCGGGGGACATGGAAGCGCTGACGCTGTGGCTTCTCCCCTGGATATGCCAGTGCGTGTCG
GTGCGGGCCGACTCCATCATCCACATCGGTGCCATCTTCGAGGAGAACGCGGCCAAGGACGACAGGGT
GTTCCAGTTGGCGGTATCCGACCTGAGCCTCAACGATGACATCCTGCAGAGCGAGAAGATCACCTACT
CCATCAAGGTCATCGAGGCCAACAACCCATTCCAGGCTGTGCAGGAAGCCTGTGACCTCATGACCCAG
GGGATTTTGGCCTTGGTCACGTCCACTGGCTGTGCATCTGCCAATGCCCTGCAGTCCCTCACGGATGC
CATGCACATCCCACACCTCTTTGTCCAGCGCAACCCGGGAGGGTCGCCACGCACCGCATGCCACCTGA
ACCCCAGCCCCGATGGTGAGGCCTACACACTGGCTTCGAGACCACCCGTCCGCCTCAATGATGTCATG
CTCAGGCTGGTGACGGAGCTGCGCTGGCAGAAGTTCGTCATGTTCTACGACAGCGAGTATGATATCCG
TGGGCTTCAAAGCTTTCTGGACCAGGCCTCGCGGCTGGGCCTTGACGTCTCTTTACAAAAGGTGGACA
AGAACATTAGCCACGTATTCACCAGCCTCTTCACCACGATGAAGACAGAGGAGCTGAATCGCTACCGG
GACACGCTTCGCCGCGCCATCCTGCTGCTCAGCCCACAGGGAGCCCACTCCTTCATCAACGAGGCCGT
GGAGACCAACCTGGCTTCCAAGGACAGCCACTGGGTCTTTGTGAATGAGGAAATCAGTGACCCGGAGA
TCCTGGATCTGGTCCATAGTGCCCTTGGAAGGATGACCGTGGTCCGGCAAATCTTTCCGTCTGCAAAG
GACAATCAGAAATGCACGAGGAACAACCACCGCATCTCCTCCCTGCTCTGCGACCCCCAGGAAGGCTA
CCTCCAGATGCTGCAGATCTCCAACCTCTATCTGTATGACAGTGTTCTGATGCTGGCCAACGCCTTTC
ACAGGAAGCTGGAGGACCGGAAGTGGCATAGCATGGCGAGCCTCAACTGCATACGGAAATCCACTAAG
CCATGGAATGGTGGGAGGTCCATGCTGGATACCATCAAAAAGGGCCACATCACTGGCCTCACTGGGGT
GATGGAGTTTCGGGAGGACAGTTCGAATCCCTATGTCCAGTTTGAAATCCTTGGCACTACCTATAGTG
AGACTTTTGGCAAAGACATGCGCAAGTTGGCGACATGGGACTCAGAGAAGGGCTTGAATGGCAGCTTG
CAAGAGAGGCCCATGGGCAGCCGCCTCCAAGGATTGACTCTTAAAGTGGTGACTGTCTTGGAAGAGCC
TTTCGTGATGGTGGCTGAGAACATCCTAGGACAGCCCAAGCGCTACAAAGGGTTCTCCATAGATGTCC
TGGATGCACTGGCCAAGGCTCTGGGCTTTAAATATGAGATTTACCAAGCCCCTGATGGCAGGTACGGT
CACCAGCTCCATAACACCTCCTGGAACGGGATGATCGGGGAGCTCATCAGCAAGAGAGCAGACTTGGC
CATCTCTGCCATCACCATCACCCCAGAGAGGGAGAGCGTTGTGGACTTCAGCAAGCGGTACATGGACT
ATTCAGTGGGGATTCTAATTAAGAAGCCCGAGGAGAAAATCAGCATCTTCTCCCTCTTTGCTCCATTT
GATTTCGCTGTGTGGGCCTGCATTGCAGCAGCCATCCCTGTGGTTGGTGTGCTGATATTTGTGTTGAA
CAGGATACAGGCTGTGAGGGCTCAGAGTGCTGCCCAGCCCAGGCCGTCAGCTTCTGCCACTCTGCACA
GCGCCATCTGGATTGTCTATGGAGCCTTCGTACAGCAAGGTGGCGAATCTTCCGTGAACTCCATGGCC
ATGCGCATCGTGATGGGCAGCTGGTGGCTCTTCACGCTCATTGTGTGCTCCTCCTACACAGCCAACCT
TGCTGCCTTCCTCACAGTGTCCAGGATGGACAACCCCATAAGGACTTTCCAGGACCTGTCCAAACAAG
TGGAAATGTCTTATGGCACTGTCCGGGATTCTGCTGTATATGAGTACTTCCGAGCCAAGGGCACCAAC
CCCCTGGAGCAGGACAGCACGTTTGCTGAACTCTGGCGGACCATCAGCAAGAACGGAGGGGCTGACAA
CTGCGTGTCCAGTCCTTCAGAAGGCATCAGGAAGGCAAAGAAGGGGAACTACGCCTTCCTGTGGGATG
TGGCCGTGGTGGAATACGCAGCCCTGACGGATGACGACTGCTCGGTGACTGTCATCGGCAACAGCATC
AGCAGCAAGGGTTACGGGATTGCCCTGCAGCATGGCAGCCCCTACAGGGACCTCTTCTCCCAGAGGAT
CCTGGAGCTGCAGGACACAGGGGACCTGGATGTGCTGAAGCAGAAGTGGTGGCCGCACATGGGCCGCT
GTGACCTCACCAGCCATGCCAGCGCCCAGGCCGACGGCAAATCCCTCAAGCTGCACAGCTTCGCCGGG
GTCTTCTGCATCCTGGCCATTGGCCTGCTCCTGGCCTGCCTGGTGGCTGCCCTGGAGTTGTGGTGGAA
CAGCAACCGGTGCCACCAGGAGACCCCCAAGGAGGACAAAGAAGTGAACTTGGAGCAGGTCCACCGGC
GCATGAACAGCCTCATGGATGAAGACATTGCTCACAAGCAGATTTCCCCAGCGTCGATTGAGCTCTCG
GCCCTGGAGATGGGGGGCCTGGCTCCCACCCAGACCTTGGAGCCGACACGGGAGTACCAGAACACCCA
GCTCTCGGTCAGCACCTTTCTGCCAGAGCAGAGCAGCCATGGCACCAGCCGGACACTCTCATCAGGGC
CCAGCAGCAACCTGCCGCTGCCGCTGAGCAGCTCGGCGACCATGCCCTCCATGCAGTGCAAACACAGG
TCACCCAACGGGGGGCTGTTCCGGCAGAGCCCGGTGAAGACCCCCATCCCCATGTCCTTCCAGCCCGT
GCCTGGAGGCGTCCTTCCAGAGGCTCTGGACACCTCCCACGGGACCTCCATCTGACTGCGCCGCCTGC
CCTCCTGCCCACCCTCCCACCCACCCGACCAGCAGAGCTTTTTAATACAAGAAAACAACAA

NOV2a, CG101396-01 Protein Sequence	SEQ ID NO: 6	1009 aa	MW at 112129.7kD

MEALTLWLLPWICQCVSVRADSIIHIGAIFEENAAKDDRVFQLAVSDLSLNDDILQSEKITYSIKVIE
ANNPFQAVQEACDLMTQGILALVTSTGCASANALQSLTDAMHIPHLFVQRNPGGSPRTACHLNPSPDG
EAYTLASRPPVRLNDVMLRLVTELRWQKFVMFYDSEYDIRGLQSFLDQASRLGLDVSLQKVDKNISHV
FTSLFTTMKTEELNRYRDTLRRAILLLSPQGAHSFINEAVETNLASKDSHWVFVNEEISDPEILDLVH
SALGRMTVVRQIFPSAKDNQKCTRNNHRISSLLCDPQEGYLQMLQISNLYLYDSVLMLANAFHRKLED
RKWHSMASLNCIRKSTKPWNGGRSMLDTIKKGHITGLTGVMEFREDSSNPYVQFEILGTTYSETFGKD
MRKLATWDSEKGLNGSLQERPMGSRLQGLTLKVVTVLEEPFVMVAENILGQPKRYKGFSIDVLDALAK
ALGFKYEIYQAPDGRYGHQLHNTSWNGMIGELISKRADLAISAITITPERESVVDFSKRYMDYSVGIL
IKKPEEKISIFSLFAPFDFAVWACIAAAIPVVGVLIFVLNRIQAVRAQSAAQPRPSASATLHSAIWIV
YGAFVQQGGESSVNSMAMRIVMGSWWLFTLIVCSSYTANLAAFLTVSRMDNPIRTFQDLSKQVEMSYG
TVRDSAVYEYFRAKGTNPLEQDSTFAELWRTISKNGGADNCVSSPSEGIRKAKKGNYAFLWDVAVVEY
AALTDDDCSVTVIGNSISSKGYGIALQHGSPYRDLFSQRILELQDTGDLDVLKQKWWPHMGRCDLTSH
ASAQADGKSLKLHSFAGVFCILAIGLLLACLVAALELWWNSNRCHQETPKEDKEVNLEQVHRRMNSLM
DEDIAHKQISPASIELSALEMGGLAPTQTLEPTREYQNTQLSVSTFLPEQSSHGTSRTLSSGPSSNLP
LPLSSSATMPSMQCKHRSPNGGLFRQSPVKTPIPMSFQPVPGGVLPEALDTSHGTSI

NOV2b, 267253224 DNA Sequence	SEQ ID NO: 7	2986 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGTACCGACTCCATCATCCACATCGGTGCCATCTTCGAGGAGAACGCGGCCAAGGACGACAGGG
TGTTCCAGTTGGCGGTATCCGACCTGAGCCTCAACGATGACATCCTGCAGAGCGAGAAGATCACCTAC
TCCATCAAGGTCATCGAGGCCAACAACCCATTCCAGGCTGTGCAGGAAGCCTGTGACCTCATGACCCA
GGGGATTTTGGCCTTGGTCACGTCCACTGGCTGTGCATCTGCCAATGCCCTGCAGTCCCTCACGGATG
CCATGCACATCCCACACCTCTTTGTCCAGCGCAACCCGGGAGGGTCGCCACGCACCGCATGCCACCTG
AACCCCAGCCCCGATGGTGAGGCCTACACACTGGCTTCGAGACCACCCGTCCGCCTCAATGATGTCAT
GCTCAGGCTGGTGACGGAGCTGCGCTGGCAGAAGTTCGTCATGTTCTACGACAGCGAGTATGATATCC
GTGGGCTTCAAAGCTTTCTGGACCAGGCCTCGCGGCTGGGCCTTGACGTCTCTTTACAAAAGGTGGAC
AAGAACATTAGCCACGTATTCACCAGCCTCTTCACCACGATGAAGACAGAGGAGCTGAATCGCTACCG
GGACACGCTTCGCCGCGCCATCCTGCTGCTCAGCCCACAGGGAGCCCACTCCTTCATCAACGAGGCCG
TGGAGACCAACCTGGCTTCCAAGGACAGCCACTGGGTCTTTGTGAATGAGGAAATCAGTGACCCGGAG
ATCCTGGATCTGGTCCATAGTGCCCTTGGAAGGATGACCGTGGTCCGGCAAATCTTTCCGTCTGCAAA
GGACAATCAGAAATGCACGAGGAACAACCACCGCATCTCCTCCCTGCTCTGCGACCCCCAGGAAGGCT
ACCTCCAGATGCTGCAGATCTCCAACCTCTATCTGTATGACAGTGTTCTGATGCTGGCCAACGCCTTT
CACAGGAAGCTGGAGGACCGGAAGTGGCATAGCATGGCGAGCCTCAACTGCATACGGAAATCCACTAA
GCCATGGAATGGTGGGAGGTCCATGCTGGATACCATCAAAAAGGGCCACATCACTGGCCTCACTGGGG
TGATGGAGTTTCGGGAGGACAGTTCGAATCCCTATGTCCAGTTTGAAATCCTTGGCACTACCTATAGT
GAGACTTTTGGCAAAGACATGCGCAAGTTGGCGACATGGGACTCAGAGAAGGGCTTGAATGGCAGCTT
GCAAGAGAGGCCCATGGGCAGCCGCCTCCAAGGATTGACTCTTAAAGTGGTGACTGTCTTGGAAGAGC
CTTTCGTGATGGTGGCTGAGAACATCCTAGGACAGCCCAAGCGCTACAAAGGGTTCTCCATAGATGTC
CTGGATGCACTGGCCAAGGCTCTGGGCTTTAAATATGAGATTTACCAAGCCCCTGATGGCAGGTACGG
TCACCAGCTCCATAACACCTCCTGGAACGGGATGATCGGGGAGCTCATCAGCAAGAGAGCAGACTTGG
CCATCTCTGCCATCACCATCACCCCAGAGAGGGAGAGCGTTGTGGACTTCAGCAAGCGGTACATGGAC
TATTCAGTGGGGATTCTAATTAAGAAGCCCGAGGAGAAAATCAGCATCTTCTCCCTCTTTGCTCCATT
TGATTTCGCTGTGTGGGCCTGCATTGCAGCAGCCATCCCTGTGGTTGGTGTGCTGATATTTGTGTTGA
ACAGGATACAGGCTGTGAGGGCTCAGAGTGCTGCCCAGCCCAGGCCGTCAGCTTCTGCCACTCTGCAC
AGCGCCATCTGGATTGTCTATGGAGCCTTCGTACAGCAAGGTGGCGAATCTTCCGTGAACTCCATGGC
CATGCGCATCGTGATGGGCAGCTGGTGGCTCTTCACGCTCATTGTGTGCTCCTCCTACACAGCCAACC
TTGCTGCCTTCCTCACAGTGTCCAGGATGGACAACCCCATAAGGACTTTCCAGGACCTGTCCAAACAA
GTGGAAATGTCTTATGGCACTGTCCGGGATTCTGCTGTATATGAGTACTTCCGAGCCAAGGGCACCAA
CCCCCTGGAGCAGGACAGCACGTTTGCTGAACTCTGGCGGACCATCAGCAAGAACGGAGGGGCTGACA
ACTGCGTGTCCAGTCCTTCAGAAGGCATCAGGAAGGCAAAGAAGGGGAACTACGCCTTCCTGTGGGAT
GTGGCCGTGGTGGAATACGCAGCCCTGACGGATGACGACTGCTCGGTGACTGTCATCGGCAACAGCAT
CAGCAGCAAGGGTTACGGGATTGCCCTGCAGCATGGCAGCCCCTACAGGGACCTCTTCTCCCAGAGGA
TCCTGGAGCTGCAGGACACAGGGGACCTGGATGTGCTGAAGCAGAAGTGGTGGCCGCACATGGGCCGC
TGTGACCTCACCAGCCATGCCAGCGCCCAGGCCGACGGCAAATCCCTCAAGCTGCACAGCTTCGCCGG
GGTCTTCTGCATCCTGGCCATTGGCCTGCTCCTGGCCTGCCTGGTGGCTGCCCTGGAGTTGTGGTGGA
ACAGCAACCGGTGCCACCAGGAGACCCCCAAGGAGGACAAAGAAGTGAACTTGGAGCAGGTCCACCGG

CGCATGAACAGCCTCATGGATGAAGACATTGCTCACAAGCAGATTTCCCCAGCGTCGATTGAGCTCTC
GGCCCTGGAGATGGGGAGCCTGGCTCCCACCCAGACCTTGGAGCCGACACGGGAGTACCAGAACACCC
AGCTCTCGGTCAGCACCTTTCTGCCAGAGCAGAGCAGCCATGGCACCAGCCGGACACTCTCATCAGGG
CCCAGCAGCAACCTGCCGCTGCCGCTGAGCAGCTCGGCGACCATGCCCTCCATGCAGTGCAAACACAG
GTCACCCAACGGGGGGCTGTTCCGGCAGAGCCCGGTGAAGACCCCCATCCCCATGTCCTTCCAGCCCG
TGCCTGGAGGCGTCCTTCCAGAGGCTCTGGACACCTCCCACGGGACCTCCATCCTCGAGGGC

NOV2b, 267253224 Protein Sequence	SEQ ID NO: 8	995 aa	MW at 110403.5kD

TGTDSIIHIGAIFEENAAKDDRVFQLAVSDLSLNDDILQSEKITYSIKVIEANNPFQAVQEACDLMTQ
GILALVTSTGCASANALQSLTDAMHIPHLFVQRNPGGSPRTACHLNPSPDGEAYTLASRPPVRLNDVM
LRLVTELRWQKFVMFYDSEYDIRGLQSFLDQASRLGLDVSLQKVDKNISHVFTSLFTTMKTEELNRYR
DTLRRAILLLSPQGAHSFINEAVETNLASKDSHWVFVNEEISDPEILDLVHSALGRMTVVRQIFPSAK
DNQKCTRNNHRISSLLCDPQEGYLQMLQISNLYLYDSVLMLANAFHRKLEDRKWHSMASLNCIRKSTK
PWNGGRSMLDTIKKGHITGLTGVMEFREDSSNPYVQFEILGTTYSETFGKDMRKLATWDSEKGLNGSL
QERPMGSRLQGLTLKVVTVLEEPFVMVAENILGQPKRYKGFSIDVLDALAKALGFKYEIYQAPDGRYG
HQLHNTSWNGMIGELISKRADLAISAITITPERESVVDFSKRYMDYSVGILIKKPEEKISIFSLFAPF
DFAVWACIAAAIPVVGVLIFVLNRIQAVRAQSAAQPRPSASATLHSAIWIVYGAFVQQGGESSVNSMA
MRIVMGSWWLFTLIVCSSYTANLAAFLTVSRMDNPIRTFQDLSKQVEMSYGTVRDSAVYEYFRAKGTN
PLEQDSTFAELWRTISKNGGADNCVSSPSEGIRKAKKGNYAFLWDVAVVEYAALTDDDCSVTVIGNSI
SSKGYGIALQHGSPYRDLFSQRILELQDTGDLDVLKQKWWPHMGRCDLTSHASAQADGKSLKLHSFAG
VFCILAIGLLLACLVAALELWWNSNRCHQETPKEDKEVNLEQVHRRMNSLMDEDIAHKQISPASIELS
ALEMGSLAPTQTLEPTREYQNTQLSVSTFLPEQSSHGTSRTLSSGPSSNLPLPLSSSATMPSMQCKHR
SPNGGLFRQSPVKTPIPMSFQPVPGGVLPEALDTSHGTSILEG

NOV2c, 315490179 DNA Sequence	SEQ ID NO: 9	3049 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCACCATGGAAGCGCTGACGCTGTGGCTTCTCCCCTGGATATGCCAGTGCGTGTCGGTGC
GGGCCGACTCCATCATCCACATTGGTGCCATCTTCGAGGAGAACGCGGCCAAGGACGACAGGGTGTTC
CAGTTGGCGGTATCCGACCTGAGCCTCAACGATGACATCCTGCAGAGCGAGAAGATCACCTACTCCAT
CAAGGTCATCGAGGCCAACAACCCATTCCAGGCTGTGCAGGAAGCCTGTGACCTCATGACCCAGGGGA
TTTTGGCCTTGGTCACGTCCACTGGCTGTGCATCTGCCAATGCCCTGCAGTCCCTCACGGATGCCATG
CACATCCCACACCTCTTTGTCCAGCGCAACCCGGGAGGGTCGCCACGCACCGCATGCCACCTGAACCC
CAGCCCCGATGGTGAGGCCTACACACTGGCTTCGAGACCACCCGTCCGCCTCAATGATGTCATGCTCA
GGCTGGTGACGGAGCTGCGCTGGCAGAAGTTCGTCATGTTCTACGACAGCGAGTATGATATCCGTGGG
CTTCAAAGCTTTCTGGACCAGGCCTCGCGGCTGGGCCTTGACGTCTCTTTACAAAAGGTGGACAAGAA
CATTAGCCACGTATTCACCAGCCTCTTCACCACGATGAAGACAGAGGAGCTGAATCGCTACCGGGACA
CGCTTCGCCGCGCCATCCTGCTGCTCAGCCCACAGGGAGCCCACTCCTTCATCAACGAGGCCGTGGAG
ACCAACCTGGCTTCCAAGGACAGCCACTGGGTCTTTGTGAATGAGGAAATCAGTGACCCGGAGATCCT
GGATCTGGTCCATAGTGCCCTTGGAAGGATGACCGTGGTCCGGCAAATCTTTCCGTCTGCAAAGGACA
ATCAGAAATGCACGAGGAACAACCACCGCATCTCCTCCCTGCTCTGCGACCCCCAGGAAGGCTACCTC
CAGATGCTGCAGATCTCCAACCTCTATCTGTATGACAGTGTTCTGATGCTGGCCAACGCCTTTCACAG
GAAGCTGGAGGACCGGAAGTGGCATAGCATGGCGAGCCTCAACTGCATACGGAAATCCACTAAGCCAT
GGAATGGTGGGAGGTCCATGCTGGATACCATCAAAAAGGGCCACATCACTGGCCTCACTGGGGTGATG
GAGTTTCGGGAGGACAGTTCGAATCCCTATGTCCAGTTTGAAATCCTTGGCACTACCTATAGTGAGAC
TTTTGGCAAAGACATGCGCAAGTTGGCGACATGGGACTCAGAGAAGGGCTTGAATGGCAGCTTGCAAG
AGAGGCCCATGGGCAGCCGCCTCCAAGGATTGACTCTTAAAGTGGTGACTGTCTTGGAAGAGCCTTTC
GTGATGGTGGCTGAGAACATCCTAGGACAGCCCAAGCGCTACAAAGGGTTCTCCATAGATGTCCTGGA
TGCACTGGCCAAGGCTCTGGGCTTTAAATATGAGATTTACCAAGCCCCTGATGGCAGGTACGGTCACC
AGCTCCATAACACCTCCTGGAACGGGATGATCGGGGAGCTCATCAGCAAGAGAGCAGACTTGGCCATC
TCTGCCATCACCATCACCCCAGAGAGGGAGAGCGTTGTGGACTTCAGCAAGCGGTACATGGACTATTC
AGTGGGGATTCTAATTAAGAAGCCCGAGGAGAAAATCAGCATCTTCTCCCTCTTTGCTCCATTTGATT
TCGCTGTGTGGGCCTGCATTGCAGCAGCCATCCCTGTGGTTGGTGTGCTGATATTTGTGTTGAACAGG
ATACAGGCTGTGAGGGCTCAGAGTGCTGCCCAGCCCAGGCCGTCAGCTTCTGCCACTCTGCACAGCGC
CATCTGGATTGTCTATGGAGCCTTCGTACAGCAAGGTGGCGAATCTTCCGTGAACTCCATGGCCATGC
GCATCGTGATGGGCAGCTGGTGGCTCTTCACGCTCATTGTGTGCTCCTCCTACACAGCCAACCTTGCT
GCCTTCCTCACAGTGTCCAGGATGGACAACCCCATAAGGACTTTCCAGGACCTGTCCAAACAAGTGGA
AATGTCTTATGGCACTGTCCGGGATTCTGCTGTATATGAGTACTTCCGAGCCAAGGGCACCAACCCCC
TGGAGCAGGACAGCACGTTTGCTGAACTCTGGCGGACCATCAGCAAGAACGGAGGGGCTGACAACTGC

GTGTCCAGTCCTTCAGAAGGCATCAGGAAGGCAAAGAAGGGGAACTACGCCTTCCTGTGGGATGTGGC
CGTGGTGGAATACGCAGCCCTGACGGATGACGACTGCTCGGTGACTGTCATCGGCAACAGCATCAGCA
GCAAGGGTTACGGGATTGCCCTGCAGCATGGCAGCCCCTACAGGGACCTCTTCTCCCAGAGGATCCTG
GAGCTGCAGGACACAGGGGACCTGGATGTGCTGAAGCAGAAGTGGTGGCCGCACATGGGCCGCTGTGA
CCTCACCAGCCATGCCAGCGCCCAGGCCGACGGCAAATCCCTCAAGCTGCACAGCTTCGCCGGGGTCT
TCTGCATCCTGGCCATTGGCCTGCTCCTGGCCTGCCTGGTGGCTGCCCTGGAGTTGTGGTGGAACAGC
AACCGGTGCCACCAGGAGACCCCCAAGGAGGACAAAGAAGTGAACTTGGAGCAGGTCCACCGGCGCAT
GAACAGCCTCATGGATGAAGACATTGCTCACAAGCAGATTTCCCCAGCGTCGATTGAGCTCTCGGCCC
TGGAGATGGGGGGCCTGGCTCCCACCCAGACCTTGGAGCCGACACGGGAGTACCAGAACACCCAGCTC
TCGGTCAGCACCTTTCTGCCAGAGCAGAGCAGCCATGGCACCAGCCGGACACTCTCATCAGGGCCCAG
CAGCAACCTGCCGCTGCCGCTGAGCAGCTCGGCGACCATGCCCTCCATGCAGTGCAAACACAGGTCAC
CCAACGGGGGGCTGTTCCGGCAGAGCCCGGTGAAGACCCCCATCCCCATGTCCTTCCAGCCCGTGCCT
GGAGGCGTCCTTCCAGAGGCTCTGGACACCTCCCACGGGACCTCCATCCTCGAGGGC

NOV2c, 315490179 Protein Sequence	SEQ ID NO: 10	1016 aa	MW at 112775.3kD

TGSTMEALTLWLLPWICQCVSVRADSIIHIGAIFEENAAKDDRVFQLAVSDLSLNDDILQSEKITYSI
KVIEANNPFQAVQEACDLMTQGILALVTSTGCASANALQSLTDAMHIPHLFVQRNPGGSPRTACHLNP
SPDGEAYTLASRPPVRLNDVMLRLVTELRWQKFVMFYDSEYDIRGLQSFLDQASRLGLDVSLQKVDKN
ISHVFTSLFTTMKTEELNRYRDTLRRAILLLSPQGAHSFINEAVETNLASKDSHWVFVNEEISDPEIL
DLVHSALGRMTVVRQIFPSAKDNQKCTRNNHRISSLLCDPQEGYLQMLQISNLYLYDSVLMLANAFHR
KLEDRKWHSMASLNCIRKSTKPWNGGRSMLDTIKKGHITGLTGVMEFREDSSNPYVQFEILGTTYSET
FGKDMRKLATWDSEKGLNGSLQERPMGSRLQGLTLKVVTVLEEPFVMVAENILGQPKRYKGFSIDVLD
ALAKALGFKYEIYQAPDGRYGHQLHNTSWNGMIGELISKRADLAISAITITPERESVVDFSKRYMDYS
VGILIKKPEEKISIFSLFAPFDFAVWACIAAAIPVVGVLIFVLNRIQAVRAQSAAQPRPSASATLHSA
IWIVYGAFVQQGGESSVNSMAMRIVMGSWWLFTLIVCSSYTANLAAFLTVSRMDNPIRTFQDLSKQVE
MSYGTVRDSAVYEYFRAKGTNPLEQDSTFAELWRTISKNGGADNCVSSPSEGIRKAKKGNYAFLWDVA
VVEYAALTDDDCSVTVIGNSISSKGYGIALQHGSPYRDLFSQRILELQDTGDLDVLKQKWWPHMGRCD
LTSHASAQADGKSLKLHSFAGVFCILAIGLLLACLVAALELWWNSNRCHQETPKEDKEVNLEQVHRRM
NSLMDEDIAHKQISPASIELSALEMGGLAPTQTLEPTREYQNTQLSVSTFLPEQSSHGTSRTLSSGPS
SNLPLPLSSSATMPSMQCKHRSPNGGLFRQSPVKTPIPMSFQPVPGGVLPEALDTSHGTSILEG

NOV2d, CG101396-02 DNA Sequence	SEQ ID NO: 11	3094 bp
	ORF Start: ATG at 56	ORF Stop: TAG at 3083

TGTCGACGGCGCCAGTGTGATGATATTGCAGATTCGCCTTCACCGCGGCCGCACC**ATG**GAAGCGCTGA
CGCTGTGGCTTCTCCCCTGGATATGCCAGTGCGTGTCGGTGCGGGCCGACTCCATCATCCACATCGGT
GCCATCTTCGAGGAGAACGCGGCCAAGGACGACAGGGTGTTCCAGTTGGCGGTATCCGACCTGAGCCT
CAGCGATGACATCCTGCAGAGCGAGAAGATCACCTACTCCATCAAGGTCATCGAGGCCAACAACCCAT
TCCAGGCTGTGCAGGAAGCCTGTGACCTCATGACCCAGGGGATTTTGGCCTTGGTCACGTCCACTGGC
TGTGCATCTGCCAATGCCCTGCAGTCCCTCACGGATGCCATGCACATCCCACACCTCTTTGTCCAGCG
CAACCCGGGAGGGTCGCCACGCACCGCATGCCACCTGAACCCCAGCCCCGATGGTGAGGCCTACACAC
TGGCTTCGAGACCACCCGTCCGCCTCAATGATGTCATGCTCAGGCTGGTGACGGAGCTGCGCTGGCAG
AAGTTCGTCATGTTCTACGACAGCGAGTATGATATCCGTGGGCTTCAAAGCTTTCTGGACCAGGCCTC
GCGGCTGGGCCTTGACGTCTCTTTACAAAAGGTGGACAAGAACATTAGCCACGTATTCACCAGCCTCT
TCACCACGATGAAGACAGAGGAGCTGAATCGCTACCGGGACACGCTTCGCCGCGCCATCCTGCTGCTC
AGCCCACAGGGAGCCCACTCCTTCATCAACGAGGCCGTGGAGACCAACCTGGCTTCCAAGGACAGCCA
CTGGGTCTTTGTGAATGAGGAAATCAGTGACCCGGAGATCCTGGATCTGGTCCATAGTGCCCTTGGAA
GGATGACCGTGGTCCGGCAAATCTTTCCGTCTGCAAAGGACAATCAGAAATGCACGAGGAACAACCAC
CGCATCTCCTCCCTGCTCTGCGACCCCCAGGAAGGCTACCTCCAGATGCTGCAGATCTCCAACCTCTA
TCTGTATGACAGTGTTCTGATGCTGGCCAACGCCTTTCACAGGAAGCTGGAGGACCGGAAGTGGCATA
GCATGGCGAGCCTCAACTGCATACGGAAATCCACTAAGCCATGGAATGGTGGGAGGTCCATGCTGGAT
ACCATCAAAAAGGGCCACATCACTGGCCTCACTGGGGTGATGGAGTTTCGGGAGGACAGTTCGAATCC
CTATGTCCAGTTTGAAATCCTTGGCACTACCTATAGTGAGACTTTTGGCAAAGACATGCGCAAGTTGG
CGACATGGGACTCAGAGAAGGGCTTGAATGGCAGCTTGCAAGAGAGGCCCATGGGCAGCCGCCTCCAA
GGATTGACTCTTAAAGTGGTGACTGTCTTGGAAGAGCCTTTCGTGATGGTGGCTGAGAACATCCTAGG
ACAGCCCAAGCGCTACAAAGGGTTCTCCATAGATGTCCTGGATGCACTGGCCAAGGCTCTGGGCTTTA
AATATGAGATTTACCAAGCCCCTGATGGCAGGTACGGTCACCAGCTCCATAACACCTCCTGGAACGGG
ATGATCGGGGAGCTCATCAGCAAGAGAGCAGACTTGGCCATCTCTGCCATCACCATCACCCCAGAGAG
GGAGAGCGTTGTGGACTTCAGCAAGCGGTACATGGACTATTCAGTGGGGATTCTAATTAAGAAGCCCG

AGGAGAAAATCAGCATCTTCTCCCTCTTTGCTCCATTTGATTTCGCTGTGTGGGCCTGCATTGCAGCA
GCCATCCCTGTGGTTGGTGTGCTGATATTTGTGTTGAACAGGATACAGGCTGTGAGGGCTCAGAGTGC
TGCCCAGCCCAGGCCGTCAGCTTCTGCCACTCTGCACAGCGCCATCTGGATTGTCTATGGAGCCTTCG
TACAGCAAGGTGGCGAATCTTCCGTGAACTCCATGGCCATGCGCATCGTGATGGGCAGCTGGTGGCTC
TTCACGCTCATTGTGTGCTCCTCCTACACAGCCAACCTTGCTGCCTTCCTCACAGTGTCCAGGATGGA
CAACCCCATAAGGACTTTCCAGGACCTGTCCAAACAAGTGGAAATGTCTTATGGCACTGTCCGGGATT
CTGCTGTATATGAGTACTTCCGAGCCAAGGGCACCAACCCCCTGGAGCAGGACAGCACGTTTGCTGAA
CTCTGGCGGACCATCAGCAAGAACGGAGGGGCTGACAACTGCGTGTCCAGTCCTTCAGAAGGCATCAG
GAAGGCAAAGAAGGGGAACTACGCCTTCCTGTGGGATGTGGCCGTGGTGGAATACGCAGCCCTGACGG
ATGACGACTGCTCGGTGACTGTCATCGGCAACAGCATCAGCAGCAAGGGTTACGGGATTGCCCTGCAG
CATGGCAGCCCCTACAGGGACCTCTTCTCCCAGAGGATCCTGGAGCTGCAGGACACAGGGGACCTGGA
TGTGCTGAAGCAGAAGTGGTGGCCGCACATGGGCCGCTGTGACCTCACCAGCCATGCCAGCGCCCAGG
CCGACGGCAAATCCCTCAAGCTGCACAGCTTCGCCGGGGTCTTCTGCATCCTGGCCATTGGCCTGCTC
CTGGCCTGCCTGGTGGCTGCCCTGGAGTTGTGGTGGAACAGCAACCGGTGCCACCAGGAGACCCCCAA
GGAGGACAAAGAAGTGAACTTGGAGCAGGTCCACCGGCGCATGAACAGCCTCATGGATGAAGACATTG
CTCACAAGCAGATTTCCCCAGCGTCGATTGAGCTCTCGGCCCTGGAGATGGGGGGCCTGGCTCCCACC
CAGACCTTGGAGCCGACACGGGAGTACCAGAACACCCAGCTCTCGGTCAGCACCTTTCTGCCAGAGCA
GAGCAGCCATGGCACCAGCCGGACACTCTCATCAGGGCCCAGCAGCAACCTGCCGCTGCCGCTGAGCA
GCTCGGCGACCATGCCCTCCATGCAGTGCAAACACAGGTCACCCAACGGGGGGCTGTTCCGGCAGAGC
CCGGTGAAGACCCCCATCCCCATGTCCTTCCAGCCCGTGCCTGGAGGCGTCCTTCCAGAGGCTCTGGA
CACCTCCCACGGGACCTCCATC**TAG**CTCGAGGGC

NOV2d, CG101396-02 Protein Sequence	SEQ ID NO: 12	1009 aa	MW at 112102.6kD

MEALTLWLLPWICQCVSVRADSIIHIGAIFEENAAKDDRVFQLAVSDLSLSDDILQSEKITYSIKVIE
ANNPFQAVQEACDLMTQGILALVTSTGCASANALQSLTDAMHIPHLFVQRNPGGSPRTACHLNPSPDG
EAYTLASRPPVRLNDVMLRLVTELRWQKFVMFYDSEYDIRGLQSFLDQASRLGLDVSLQKVDKNISHV
FTSLFTTMKTEELNRYRDTLRRAILLLSPQGAHSFINEAVETNLASKDSHWVFVNEEISDPEILDLVH
SALGRMTVVRQIFPSAKDNQKCTRNNHRISSLLCDPQEGYLQMLQISNLYLYDSVLMLANAFHRKLED
RKWHSMASLNCIRKSTKPWNGGRSMLDTIKKGHITGLTGVMEFREDSSNPYVQFEILGTTYSETFGKD
MRKLATWDSEKGLNGSLQERPMGSRLQGLTLKVVTVLEEPFVMVAENILGQPKRYKGFSIDVLDALAK
ALGFKYEIYQAPDGRYGHQLHNTSWNGMIGELISKRADLAISAITITPERESVVDFSKRYMDYSVGIL
IKKPEEKISIFSLFAPFDFAVWACIAAAIPVVGVLIFVLNRIQAVRAQSAAQPRPSASATLHSAIWIV
YGAFVQQGGESSVNSMAMRIVMGSWWLFTLIVCSSYTANLAAFLTVSRMDNPIRTFQDLSKQVEMSYG
TVRDSAVYEYFRAKGTNPLEQDSTFAELWRTISKNGGADNCVSSPSEGIRKAKKGNYAFLWDVAVVEY
AALTDDDCSVTVIGNSISSKGYGIALQHGSPYRDLFSQRILELQDTGDLDVLKQKWWPHMGRCDLTSH
ASAQADGKSLKLHSFAGVFCILAIGLLLACLVAALELWWNSNRCHQETPKEDKEVNLEQVHRRMNSLM
DEDIAHKQISPASIELSALEMGGLAPTQTLEPTREYQNTQLSVSTFLPEQSSHGTSRTLSSGPSSNLP
LPLSSSATMPSMQCKHRSPNGGLFRQSPVKTPIPMSFQPVPGGVLPEALDTSHGTSI

NOV2e, SNP13379211 of CG101396-01, DNA Sequence	SEQ ID NO: 13	3189 bp
	ORF Start: ATG at 86	ORF Stop: TGA at 3113
	SNP Pos: 3184	SNP Change: C to T

CTGCACCGGGACCAGCGCCTCCCCGCTTCGCGCTGCCCTCGGCCTCGCCCCGGGCCCGGGTGGATGAG
CCGCGCGCCCGGGGGAC**ATG**GAAGCGCTGACGCTGTGGCTTCTCCCCTGGATATGCCAGTGCGTGTCG
GTGCGGGCCGACTCCATCATCCACATCGGTGCCATCTTCGAGGAGAACGCGGCCAAGGACGACAGGGT
GTTCCAGTTGGCGGTATCCGACCTGAGCCTCAACGATGACATCCTGCAGAGCGAGAAGATCACCTACT
CCATCAAGGTCATCGAGGCCAACAACCCATTCCAGGCTGTGCAGGAAGCCTGTGACCTCATGACCCAG
GGGATTTTGGCCTTGGTCACGTCCACTGGCTGTGCATCTGCCAATGCCCTGCAGTCCCTCACGGATGC
CATGCACATCCCACACCTCTTTGTCCAGCGCAACCCGGGAGGGTCGCCACGCACCGCATGCCACCTGA
ACCCCAGCCCCGATGGTGAGGCCTACACACTGGCTTCGAGACCACCCGTCCGCCTCAATGATGTCATG
CTCAGGCTGGTGACGGAGCTGCGCTGGCAGAAGTTCGTCATGTTCTACGACAGCGAGTATGATATCCG
TGGGCTTCAAAGCTTTCTGGACCAGGCCTCGCGGCTGGGCCTTGACGTCTCTTTACAAAAGGTGGACA
AGAACATTAGCCACGTATTCACCAGCCTCTTCACCACGATGAAGACAGAGGAGCTGAATCGCTACCGG
GACACGCTTCGCCGCGCCATCCTGCTGCTCAGCCCACAGGGAGCCCACTCCTTCATCAACGAGGCCGT
GGAGACCAACCTGGCTTCCAAGGACAGCCACTGGGTCTTTGTGAATGAGGAAATCAGTGACCCGGAGA
TCCTGGATCTGGTCCATAGTGCCCTTGGAAGGATGACCGTGGTCCGGCAAATCTTTCCGTCTGCAAAG
GACAATCAGAAATGCACGAGGAACAACCACCGCATCTCCTCCCTGCTCTGCGACCCCCAGGAAGGCTA

CCTCCAGATGCTGCAGATCTCCAACCTCTATCTGTATGACAGTGTTCTGATGCTGGCCAACGCCTTTC
ACAGGAAGCTGGAGGACCGGAAGTGGCATAGCATGGCGAGCCTCAACTGCATACGGAAATCCACTAAG
CCATGGAATGGTGGGAGGTCCATGCTGGATACCATCAAAAAGGGCCACATCACTGGCCTCACTGGGGT
GATGGAGTTTCGGGAGGACAGTTCGAATCCCTATGTCCAGTTTGAAATCCTTGGCACTACCTATAGTG
AGACTTTTGGCAAAGACATGCGCAAGTTGGCGACATGGGACTCAGAGAAGGGCTTGAATGGCAGCTTG
CAAGAGAGGCCCATGGGCAGCCGCCTCCAAGGATTGACTCTTAAAGTGGTGACTGTCTTGGAAGAGCC
TTTCGTGATGGTGGCTGAGAACATCCTAGGACAGCCCAAGCGCTACAAAGGGTTCTCCATAGATGTCC
TGGATGCACTGGCCAAGGCTCTGGGCTTTAAATATGAGATTTACCAAGCCCCTGATGGCAGGTACGGT
CACCAGCTCCATAACACCTCCTGGAACGGGATGATCGGGGAGCTCATCAGCAAGAGAGCAGACTTGGC
CATCTCTGCCATCACCATCACCCCAGAGAGGGAGAGCGTTGTGGACTTCAGCAAGCGGTACATGGACT
ATTCAGTGGGGATTCTAATTAAGAAGCCCGAGGAGAAAATCAGCATCTTCTCCCTCTTTGCTCCATTT
GATTTCGCTGTGTGGGCCTGCATTGCAGCAGCCATCCCTGTGGTTGGTGTGCTGATATTTGTGTTGAA
CAGGATACAGGCTGTGAGGGCTCAGAGTGCTGCCCAGCCCAGGCCGTCAGCTTCTGCCACTCTGCACA
GCGCCATCTGGATTGTCTATGGAGCCTTCGTACAGCAAGGTGGCGAATCTTCCGTGAACTCCATGGCC
ATGCGCATCGTGATGGGCAGCTGGTGGCTCTTCACGCTCATTGTGTGCTCCTCCTACACAGCCAACCT
TGCTGCCTTCCTCACAGTGTCCAGGATGGACAACCCCATAAGGACTTTCCAGGACCTGTCCAAACAAG
TGGAAATGTCTTATGGCACTGTCCGGGATTCTGCTGTATATGAGTACTTCCGAGCCAAGGGCACCAAC
CCCCTGGAGCAGGACAGCACGTTTGCTGAACTCTGGCGGACCATCAGCAAGAACGGAGGGGCTGACAA
CTGCGTGTCCAGTCCTTCAGAAGGCATCAGGAAGGCAAAGAAGGGGAACTACGCCTTCCTGTGGGATG
TGGCCGTGGTGGAATACGCAGCCCTGACGGATGACGACTGCTCGGTGACTGTCATCGGCAACAGCATC
AGCAGCAAGGGTTACGGGATTGCCCTGCAGCATGGCAGCCCCTACAGGGACCTCTTCTCCCAGAGGAT
CCTGGAGCTGCAGGACACAGGGGACCTGGATGTGCTGAAGCAGAAGTGGTGGCCGCACATGGGCCGCT
GTGACCTCACCAGCCATGCCAGCGCCCAGGCCGACGGCAAATCCCTCAAGCTGCACAGCTTCGCCGGG
GTCTTCTGCATCCTGGCCATTGGCCTGCTCCTGGCCTGCCTGGTGGCTGCCCTGGAGTTGTGGTGGAA
CAGCAACCGGTGCCACCAGGAGACCCCCAAGGAGGACAAAGAAGTGAACTTGGAGCAGGTCCACCGGC
GCATGAACAGCCTCATGGATGAAGACATTGCTCACAAGCAGATTTCCCCAGCGTCGATTGAGCTCTCG
GCCCTGGAGATGGGGGGCCTGGCTCCCACCCAGACCTTGGAGCCGACACGGGAGTACCAGAACACCCA
GCTCTCGGTCAGCACCTTTCTGCCAGAGCAGAGCAGCCATGGCACCAGCCGGACACTCTCATCAGGGC
CCAGCAGCAACCTGCCGCTGCCGCTGAGCAGCTCGGCGACCATGCCCTCCATGCAGTGCAAACACAGG
TCACCCAACGGGGGGCTGTTCCGGCAGAGCCCGGTGAAGACCCCCATCCCCATGTCCTTCCAGCCCGT
GCCTGGAGGCGTCCTTCCAGAGGCTCTGGACACCTCCCACGGGACCTCCATCTGACTGCGCCGCCTGC
CCTCCTGCCCACCCTCCCACCCACCCGACCAGCAGAGCTTTTTAATACAAGAAAATAACAA

NOV2e, SNP13379211 of CG101396-01, Protein Sequence	SEQ ID NO: 14	1009 aa	MW at 112129.7kD
			SNP Change: no change

MEALTLWLLPWICQCVSVRADSIIHIGAIFEENAAKDDRVFQLAVSDLSLNDDILQSEKITYSIKVIE
ANNPFQAVQEACDLMTQGILALVTSTGCASANALQSLTDAMHIPHLFVQRNPGGSPRTACHLNPSPDG
EAYTLASRPPVRLNDVMLRLVTELRWQKFVMFYDSEYDIRGLQSFLDQASRLGLDVSLQKVDKNISHV
FTSLFTTMKTEELNRYRDTLRRAILLLSPQGAHSFINEAVETNLASKDSHWVFVNEEISDPEILDLVH
SALGRMTVVRQIFPSAKDNQKCTRNNHRISSLLCDPQEGYLQMLQISNLYLYDSVLMLANAFHRKLED
RKWHSMASLNCIRKSTKPWNGGRSMLDTIKKGHITGLTGVMEFREDSSNPYVQFEILGTTYSETFGKD
MRKLATWDSEKGLNGSLQERPMGSRLQGLTLKVVTVLEEPFVMVAENILGQPKRYKGFSIDVLDALAK
ALGFKYEIYQAPDGRYGHQLHNTSWNGMIGELISKRADLAISAITITPERESVVDFSKRYMDYSVGIL
IKKPEEKISIFSLFAPFDFAVWACIAAAIPVVGVLIFVLNRIQAVRAQSAAQPRPSASATLHSAIWIV
YGAFVQQGGESSVNSMAMRIVMGSWWLFTLIVCSSYTANLAAFLTVSRMDNPIRTFQDLSKQVEMSYG
TVRDSAVYEYFRAKGTNPLEQDSTFAELWRTISKNGGADNCVSSPSEGIRKAKKGNYAFLWDVAVVEY
AALTDDDCSVTVIGNSISSKGYGIALQHGSPYRDLFSQRILELQDTGDLDVLKQKWWPHMGRCDLTSH
ASAQADGKSLKLHSFAGVFCILAIGLLLACLVAALELWWNSNRCHQETPKEDKEVNLEQVHRRMNSLM
DEDIAHKQISPASIELSALEMGGLAPTQTLEPTREYQNTQLSVSTFLPEQSSHGTSRTLSSGPSSNLP
LPLSSSATMPSMQCKHRSPNGGLFRQSPVKTPIPMSFQPVPGGVLPEALDTSHGTSI

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 2B.

Table 2B. Comparison of the NOV2 protein sequences.

NOV2a	----MEALTLWLLPWICQCVSVRADSIIHIGAIFEENAAKDDRVFQLAVSDLSLNDDILQ

```
NOV2b   ----------TGTDSIIHIGAIFEENAAKDDRVFQLAVSDLSLNDDILQ
NOV2c   TGSTMEALTLWLLPWICQCVSVRADSIIHIGAIFEENAAKDDRVFQLAVSDLSLNDDILQ
NOV2d   ----MEALTLWLLPWICQCVSVRADSIIHIGAIFEENAAKDDRVFQLAVSDLSLSDDILQ

NOV2a   SEKITYSIKVIEANNPFQAVQEACDLMTQGILALVTSTGCASANALQSLTDAMHIPHLFV
NOV2b   SEKITYSIKVIEANNPFQAVQEACDLMTQGILALVTSTGCASANALQSLTDAMHIPHLFV
NOV2c   SEKITYSIKVIEANNPFQAVQEACDLMTQGILALVTSTGCASANALQSLTDAMHIPHLFV
NOV2d   SEKITYSIKVIEANNPFQAVQEACDLMTQGILALVTSTGCASANALQSLTDAMHIPHLFV

NOV2a   QRNPGGSPRTACHLNPSPDGEAYTLASRPPVRLNDVMLRLVTELRWQKFVMFYDSEYDIR
NOV2b   QRNPGGSPRTACHLNPSPDGEAYTLASRPPVRLNDVMLRLVTELRWQKFVMFYDSEYDIR
NOV2c   QRNPGGSPRTACHLNPSPDGEAYTLASRPPVRLNDVMLRLVTELRWQKFVMFYDSEYDIR
NOV2d   QRNPGGSPRTACHLNPSPDGEAYTLASRPPVRLNDVMLRLVTELRWQKFVMFYDSEYDIR

NOV2a   GLQSFLDQASRLGLDVSLQKVDKNISHVFTSLFTTMKTEELNRYRDTLRRAILLLSPQGA
NOV2b   GLQSFLDQASRLGLDVSLQKVDKNISHVFTSLFTTMKTEELNRYRDTLRRAILLLSPQGA
NOV2c   GLQSFLDQASRLGLDVSLQKVDKNISHVFTSLFTTMKTEELNRYRDTLRRAILLLSPQGA
NOV2d   GLQSFLDQASRLGLDVSLQKVDKNISHVFTSLFTTMKTEELNRYRDTLRRAILLLSPQGA

NOV2a   HSFINEAVETNLASKDSHWVFVNEEISDPEILDLVHSALGRMTVVRQIFPSAKDNQKCTR
NOV2b   HSFINEAVETNLASKDSHWVFVNEEISDPEILDLVHSALGRMTVVRQIFPSAKDNQKCTR
NOV2c   HSFINEAVETNLASKDSHWVFVNEEISDPEILDLVHSALGRMTVVRQIFPSAKDNQKCTR
NOV2d   HSFINEAVETNLASKDSHWVFVNEEISDPEILDLVHSALGRMTVVRQIFPSAKDNQKCTR

NOV2a   NNHRISSLLCDPQEGYLQMLQISNLYLYDSVLMLANAFHRKLEDRKWHSMASLNCIRKST
NOV2b   NNHRISSLLCDPQEGYLQMLQISNLYLYDSVLMLANAFHRKLEDRKWHSMASLNCIRKST
NOV2c   NNHRISSLLCDPQEGYLQMLQISNLYLYDSVLMLANAFHRKLEDRKWHSMASLNCIRKST
NOV2d   NNHRISSLLCDPQEGYLQMLQISNLYLYDSVLMLANAFHRKLEDRKWHSMASLNCIRKST

NOV2a   KPWNGGRSMLDTIKKGHITGLTGVMEFREDSSNPYVQFEILGTTYSETFGKDMRKLATWD
NOV2b   KPWNGGRSMLDTIKKGHITGLTGVMEFREDSSNPYVQFEILGTTYSETFGKDMRKLATWD
NOV2c   KPWNGGRSMLDTIKKGHITGLTGVMEFREDSSNPYVQFEILGTTYSETFGKDMRKLATWD
NOV2d   KPWNGGRSMLDTIKKGHITGLTGVMEFREDSSNPYVQFEILGTTYSETFGKDMRKLATWD

NOV2a   SEKGLNGSLQERPMGSRLQGLTLKVVTVLEEPFVMVAENILGQPKRYKGFSIDVLDALAK
NOV2b   SEKGLNGSLQERPMGSRLQGLTLKVVTVLEEPFVMVAENILGQPKRYKGFSIDVLDALAK
NOV2c   SEKGLNGSLQERPMGSRLQGLTLKVVTVLEEPFVMVAENILGQPKRYKGFSIDVLDALAK
NOV2d   SEKGLNGSLQERPMGSRLQGLTLKVVTVLEEPFVMVAENILGQPKRYKGFSIDVLDALAK

NOV2a   ALGFKYEIYQAPDGRYGHQLHNTSWNGMIGELISKRADLAISAITITPERESVVDFSKRY
NOV2b   ALGFKYEIYQAPDGRYGHQLHNTSWNGMIGELISKRADLAISAITITPERESVVDFSKRY
NOV2c   ALGFKYEIYQAPDGRYGHQLHNTSWNGMIGELISKRADLAISAITITPERESVVDFSKRY
NOV2d   ALGFKYEIYQAPDGRYGHQLHNTSWNGMIGELISKRADLAISAITITPERESVVDFSKRY

NOV2a   MDYSVGILIKKPEEKISIFSLFAPFDFAVWACIAAAIPVVGVLIFVLNRIQAVRAQSAAQ
NOV2b   MDYSVGILIKKPEEKISIFSLFAPFDFAVWACIAAAIPVVGVLIFVLNRIQAVRAQSAAQ
NOV2c   MDYSVGILIKKPEEKISIFSLFAPFDFAVWACIAAAIPVVGVLIFVLNRIQAVRAQSAAQ
NOV2d   MDYSVGILIKKPEEKISIFSLFAPFDFAVWACIAAAIPVVGVLIFVLNRIQAVRAQSAAQ

NOV2a   PRPSASATLHSAIWIVYGAFVQQGGESSVNSMAMRIVMGSWWLFTLIVCSSYTANLAAFL
NOV2b   PRPSASATLHSAIWIVYGAFVQQGGESSVNSMAMRIVMGSWWLFTLIVCSSYTANLAAFL
NOV2c   PRPSASATLHSAIWIVYGAFVQQGGESSVNSMAMRIVMGSWWLFTLIVCSSYTANLAAFL
NOV2d   PRPSASATLHSAIWIVYGAFVQQGGESSVNSMAMRIVMGSWWLFTLIVCSSYTANLAAFL
```

```
NOV2a   TVSRMDNPIRTFQDLSKQVEMSYGTVRDSAVYEYFRAKGTNPLEQDSTFAELWRTISKNG
NOV2b   TVSRMDNPIRTFQDLSKQVEMSYGTVRDSAVYEYFRAKGTNPLEQDSTFAELWRTISKNG
NOV2c   TVSRMDNPIRTFQDLSKQVEMSYGTVRDSAVYEYFRAKGTNPLEQDSTFAELWRTISKNG
NOV2d   TVSRMDNPIRTFQDLSKQVEMSYGTVRDSAVYEYFRAKGTNPLEQDSTFAELWRTISKNG

NOV2a   GADNCVSSPSEGIRKAKKGNYAFLWDVAVVEYAALTDDDCSVTVIGNSISSKGYGIALQH
NOV2b   GADNCVSSPSEGIRKAKKGNYAFLWDVAVVEYAALTDDDCSVTVIGNSISSKGYGIALQH
NOV2c   GADNCVSSPSEGIRKAKKGNYAFLWDVAVVEYAALTDDDCSVTVIGNSISSKGYGIALQH
NOV2d   GADNCVSSPSEGIRKAKKGNYAFLWDVAVVEYAALTDDDCSVTVIGNSISSKGYGIALQH

NOV2a   GSPYRDLFSQRILELQDTGDLDVLKQKWWPHMGRCDLTSHASAQADGKSLKLHSFAGVFC
NOV2b   GSPYRDLFSQRILELQDTGDLDVLKQKWWPHMGRCDLTSHASAQADGKSLKLHSFAGVFC
NOV2c   GSPYRDLFSQRILELQDTGDLDVLKQKWWPHMGRCDLTSHASAQADGKSLKLHSFAGVFC
NOV2d   GSPYRDLFSQRILELQDTGDLDVLKQKWWPHMGRCDLTSHASAQADGKSLKLHSFAGVFC

NOV2a   ILAIGLLLACLVAALELWWNSNRCHQETPKEDKEVNLEQVHRRMNSLMDEDIAHKQISPA
NOV2b   ILAIGLLLACLVAALELWWNSNRCHQETPKEDKEVNLEQVHRRMNSLMDEDIAHKQISPA
NOV2c   ILAIGLLLACLVAALELWWNSNRCHQETPKEDKEVNLEQVHRRMNSLMDEDIAHKQISPA
NOV2d   ILAIGLLLACLVAALELWWNSNRCHQETPKEDKEVNLEQVHRRMNSLMDEDIAHKQISPA

NOV2a   SIELSALEMGGLAPTQTLEPTREYQNTQLSVSTFLPEQSSHGTSRTLSSGPSSNLPLPLS
NOV2b   SIELSALEMGSLAPTQTLEPTREYQNTQLSVSTFLPEQSSHGTSRTLSSGPSSNLPLPLS
NOV2c   SIELSALEMGGLAPTQTLEPTREYQNTQLSVSTFLPEQSSHGTSRTLSSGPSSNLPLPLS
NOV2d   SIELSALEMGGLAPTQTLEPTREYQNTQLSVSTFLPEQSSHGTSRTLSSGPSSNLPLPLS

NOV2a   SSATMPSMQCKHRSPNGGLFRQSPVKTPIPMSFQPVPGGVLPEALDTSHGTSI---
NOV2b   SSATMPSMQCKHRSPNGGLFRQSPVKTPIPMSFQPVPGGVLPEALDTSHGTSILEG
NOV2c   SSATMPSMQCKHRSPNGGLFRQSPVKTPIPMSFQPVPGGVLPEALDTSHGTSILEG
NOV2d   SSATMPSMQCKHRSPNGGLFRQSPVKTPIPMSFQPVPGGVLPEALDTSHGTSI---

NOV2a  (SEQ ID NO:  6)
NOV2b  (SEQ ID NO:  8)
NOV2c  (SEQ ID NO:  10)
NOV2d  (SEQ ID NO:  12)
```

Further analysis of the NOV2a protein yielded the following properties shown in Table 2C.

Table 2C. Protein Sequence Properties NOV2a

SignalP analysis:	Cleavage site between residues 21 and 22
PSORT II analysis:	

```
PSG:  a new signal peptide prediction method
      N-region:  length 2;  pos.chg 0;  neg.chg 1
      H-region:  length 16;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -1.56
      possible cleavage site: between 17 and 18

>>> Seems to have no N-terminal signal peptide
```

```
ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   3
      INTEGRAL    Likelihood =-11.41   Transmembrane  567 - 583
      INTEGRAL    Likelihood = -0.43   Transmembrane  629 - 645
      INTEGRAL    Likelihood =-12.79   Transmembrane  834 - 850
      PERIPHERAL  Likelihood =  1.01 (at 437)
      ALOM score: -12.79  (number of TMSs: 3)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 574
      Charge difference:  5.0   C( 3.0) - N(-2.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  3.64
      Hyd Moment(95):  6.21    G content:       0
      D/E content:     2       S/T content:     2
      Score: -5.86

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 29  VRA|DS

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: PKRYKGF (5) at  460
      bipartite: RKWHSMASLNCIRKSTK at  341
      content of basic residues:   9.5%
      NLS Score:  0.45

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none
```

```
Prenylation motif: none



memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 70.6

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       44.4 %: endoplasmic reticulum
       22.2 %: vacuolar
       11.1 %: Golgi
       11.1 %: nuclear
       11.1 %: mitochondrial

>> prediction for CG101396-01 is end (k=9)
```

A search of the NOV2a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 2D.

Table 2D. Geneseq Results for NOV2a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV2a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAE13278	Human transporters and ion channels (TRICH)-5 - Homo sapiens, 1009 aa. [WO200177174-A2, 18-OCT-2001]	1..1009 1..1009	1008/1009 (99%) 1009/1009 (99%)	0.0
AAM48984	Rat glutamate receptor delta-1 subunit SEQ ID NO: 1 - Rattus norvegicus, 1009 aa. [WO200206313-A2, 24-JAN-2002]	1..1009 1..1009	993/1009 (98%) 1004/1009 (99%)	0.0
AAG77969	Human ion channel protein IC32391 - Homo sapiens, 1009 aa. [WO200183752-A2, 08-NOV-2001]	1..1009 1..1009	999/1009 (99%) 1001/1009 (99%)	0.0
AAB40361	Human ORFX ORF125 polypeptide sequence SEQ ID NO:250 - Homo sapiens, 927 aa. [WO200058473-A2, 05-OCT-2000]	83..1009 1..927	927/927 (100%) 927/927 (100%)	0.0
AAM48988	Human glutamate receptor delta-1 subunit SEQ ID NO: 8 - Homo sapiens, 791 aa. [WO200206313-A2, 24-JAN-2002]	219..1009 1..791	791/791 (100%) 791/791 (100%)	0.0

In a BLAST search of public sequence databases, the NOV2a protein was found to have homology to the proteins shown in the BLASTP data in Table 2E.

Table 2E. Public BLASTP Results for NOV2a				
Protein Accession Number	**Protein/Organism/Length**	**NOV2a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q62640	Glutamate receptor delta-1 subunit precursor - Rattus norvegicus (Rat), 1009 aa.	1..1009 1..1009	993/1009 (98%) 1004/1009 (99%)	0.0
CAD19379	Sequence 4 from Patent WO0183752 - Homo sapiens (Human), 1009 aa.	1..1009 1..1009	999/1009 (99%) 1001/1009 (99%)	0.0
S28857	glutamate receptor delta-1 chain precursor - rat, 1009 aa.	1..1009 1..1009	992/1009 (98%) 1004/1009 (99%)	0.0
Q61627	Glutamate receptor delta-1 subunit precursor - Mus musculus (Mouse), 1009 aa.	1..1009 1..1009	992/1009 (98%) 1002/1009 (98%)	0.0
Q9ULK0	Hypothetical protein KIAA1220 - Homo sapiens (Human), 791 aa (fragment).	219..1009 1..791	791/791 (100%) 791/791 (100%)	0.0

PFam analysis indicates that the NOV2a protein contains the domains shown in the Table 2F.

Table 2F. Domain Analysis of NOV2a			
Pfam Domain	**NOV2a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ANF_receptor	17..410	67/478 (14%) 261/478 (55%)	0.00034
SBP_bac_3	439..808	60/404 (15%) 221/404 (55%)	8.4e-07
lig_chan	562..852	123/325 (38%) 250/325 (77%)	6.9e-125

Example 3.

The NOV3 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 3A.

5

Table 3A. NOV3 Sequence Analysis

NOV3a, CG102348-01 DNA Sequence	SEQ ID NO: 15	3345 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 1479

CAGATGTCCAGTTCCAG**ATG**CCTGGACCCAGAGTGTGGGGGAAATATCTCTGGAGAAGCCCTCACTCC
AAAGGCTGTCCAGGCGCAATGTGGTGGCTGCTTCTCTGGGGAGTCCTCCAGGCTTGCCCAACCCGGGG
CTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGGGTACCCAGAGCCGTATGGCA
AAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTGTGAGGCTCGTCTTCCAGGAC
TTCGACCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATCTCATTCGTCGGTTCGGATCC
AAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGGTCAGAGGGAGTTTGTATCCT
CAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGAACAAGACTGCCCACCTTCAC
AAGGGCTTTCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAGCCCATCAGCGAGGCCAGCAG
GGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCAGAACCACTGCCAGGAGCCCT
ATTATCAGGCCGCGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGACCTGGAAAGACAGACAGGAT
GGGGAGGAGGTTCTTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACCCCCATTGCCCAGAATCAGAC
GACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGCCTTCACCAGTATCCACGGCC
GTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCCACACCGTCTACCCCAAGGAC
AGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCACACAGCCATAGATGAGATGCT
GAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTACCGTCAGAATGAGTCCCATA
ACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCCTGGGCCCCAACGTCCTCCCG
GTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGCTACGTCAGTGGGTTTGGCAT
GGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGTAGCTCCCAGGGAGGCCTGCA
ACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATATGTTCTGTGTTGGGGATGAG
ACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTGGTATGGGACAATCATGCCCA
TCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGAAGGGTATGACTTCTACACCA
AGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGAATT**GA**CCCTGGGGGCTTGAA
CAGGGACTGACCAGCACAGTGGAGGCCCCAGGCAACAGAGGGCCTGGAGTGAGGACTGAACACTGGGG
TAGGGGTTGGGGGTGGGGGGTTGGGGGAGGCAGGGAAATCCTATTCACATCACTGTTGCACCAAGCCA
CTGCAAGAGAAACCCCCACCCGGCAAGCCCGCCCCATCCCAGACAGGAAGCAGAGTCCCACAGACCGC
TCCTCCTCACCCTCTACCTCCCTGTGCTCATGCACTAGGCCCCGGGAAGCCTGTACATCTCAACAACT
TTCGCCTTGAATGTCCTTAGAACCACCTTCCCCTACTTCATCTGTTGACACAGCTTTTATACTCACCT
GTGGAAGAGTCAGCTACTCACCCGCTATTAGAGTATGGAGGAAGGGGTTTTCATTGCATTGCATTTCT
GAAACATTCCTAAGACCCTTTAGTTGACCTTCAAATATTCAAGCTATTCTGCAGCTCCAAGATGCAAT
TATAGAAACAGCTCCTTTTTTATTTTATGTCCTCTATATGCCAGGTGCTTCACCTGTTATTTCACTTA
ATCCTCATACCATATTTGCAAAGGATGTGTTATTATCTATGTGTGACAAATGAGGAAACTGAGGCTCA
GGGGATAAAGGGACTTGCCCAAGTCCCACAGCTGGTGTGTGACTGCAGAGACTGTGCTCTTCCCAGTG
TGCTGCAATACTTCTCAACCCTCCTCTAACCTGCTGTGTCACCCGCTTTCCCTCCCAGCCCCCACATC
CTTACCATTTTCCCTCCCTGGGAATTCCTGCTTCTGCGAAAATGGTATCCTCTAGCTCACACTTTCCT
AATGGCCCCATCTCCTGCAGAAGCCAGGTGAGCCCAGCACTGGACTGAAGTTCTTGCAGACACCCCAC
CTGTGCCCCTATCATCAGGGGAACTGCTCCACCTGAGAGGACCAACTCTTTAATTTTTAGTAAAACCT
GAAGGTGATGGGCCGGGCGCAGTGGCTCACGCCTGTAATCCCAACACCTTAGGAGTCCGAGGTGGGTG
GATCACGAGGTCAGGAGATCCAGCCCATCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATAC
AAAAATTAGCCGGGCGTGGTGACACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
CACTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCTAAGATCACGCCACTGCACTCCAGCCTGCGGACA
GACCAAGACTTCATCCCCCCCCAAAAAAAAAAGATTGGAGGTGATTTACAGTGAAAGACACAAATAAAA
TACAACTGTTCAATGGAAATAGAAAATAAACACCATAAAAGAGAGAAGAGAGGTAATTTGTTAGCATC
AAGAGTCAAGTTGCTATATGGTCAAAGGTTAAATTTATCTCTAAAAAATGGCAGGATTCAAAGTTGTA
CATACATGTGATTACTTCTGTTTTTTACACCCACATACAGTACAAAAGATTATTAAAAATATTCCCAA
AAGGCAGGTGCAATGATGCACACTTATACCCCCAGCCACTCAGGAGGCTGATGCAAGAGGATCGCTTG

AGCCCAGGAGTTGAAGTCCAGCCTAAGCAACATAGTGAAACCCCATCTCCAAAAATATAATAATAATT
CTCTCAAAATACTAAACAGAGGTGGTTTTATTGATAAGATTTTGGCTGTTTGGTTTTCCACTATTCTC
TATTGGCTAAAATTTGTTTAATGAGCATGAAATGTTTTTATTTTATTTTGCTTATTTTTATGATTGCA
AAAAATGATATGAGTTTCTCCCTGCCAAGGCAAAAAATATATATATACCTATAAAAAAAAAAAAAA
AAAAAAAAAAAA

NOV3a, CG102348-01 Protein Sequence	SEQ ID NO: 16	487 aa	MW at 53483.9kD

MPGPRVWGKYLWRSPHSKGCPGAMWWLLLWGVLQACPTRGSVLLAQELPQQLTSPGYPEPYGKGQESS
TDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLR
LTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASRGSEAINAPGDNPAKVQNHCQEPYYQAAA
AGALTCATPGTWKDRQDGEEVLQCMPVCGRPVTPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGAL
LGDRWILTAAHTVYPKDSVSLRKNQSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDI
ALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQK
RQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
DWIKGVMNGKN

NOV3b, 199842645 DNA Sequence	SEQ ID NO: 17	1368 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTTGCCCAACCCGGGGCTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGG
GTACCCAGAGCCGTATGGCAAAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTG
TGAGGTTCGTCTTCCAGGACTTCGACCTGGAGTCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATC
TCATTCGTCGGTTCGGATCCAAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGG
TCAGAGGGAGTTTGTATCCTCAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGA
ACAAGACTGCCCACCTCCACAAGGGCTTCCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAG
CCCATCAGCGAGGCCAGCAGGGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCA
GAACCACTGCCAGGAGCCCTATTATCAGGCCACGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGA
CCTGGAAAGACAGACAGGATGGGGAGGAGGTTCTTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACC
CCCATTGCCCAGAATCAGACGACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGC
CTTCACCAGTATCCACGGCCGTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCC
ACACCATCTACCCCAAGGACAGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCAC
ACAGCCATAGATGAGATGCTGAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTA
CCGTCAGAATGAGTCCCATAACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCC
TGGGCCCCAACGTCCTCCCGGTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGC
TACGTCAGTGGGTTTGGCATGGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGT
AGCTCCCAGGGAGGCCTGCAACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATA
TGTTCTGTGTTGGGGATGAGACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTG
GTATGGGACAATCATGCCCATCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGA
AGGGTATGACTTCTACACCAAGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGA
ATAAGCTT

NOV3b, 199842645 Protein Sequence	SEQ ID NO: 18	456 aa	MW at 49986.8kD

KLCPTRGSVLLAQELPQQLTSPGYPEPYGKGQESSTDIKAPEGFAVRFVFQDFDLESSQDCAGDSVTI
SFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQ
PISEASRGSEAINAPGDNPAKVQNHCQEPYYQATAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPVT
PIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKNQSVNVFLGH
TAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLG
YVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYV
VWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYVDWIKGVMNGKNKL

NOV3c, 198306343 DNA Sequence	SEQ ID NO: 19	741 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGCCTTCACCAGTATCCA
CGGCCGTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCCACACCATCTACCCCA
AGAACAGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCACACAGCCATAGATGAG
ATGCTGAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTACCGTCAGAATGAGTC
CCATAACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCCTGGGCCCCAACGTCC

TCCCGGTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGCTACGTCAGTGGGTTT
GGCATGGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGTAGCTCCCAGGGAGGC
CTGCAACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATATGTTCTGTGTTGGGG
ATGAGACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTGGTATGGGACAATCGT
GCCCATCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGAAGGGTATGACTTCTA
CACCAAGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGAATAAGCTT

NOV3c, 198306343 Protein Sequence	SEQ ID NO: 20	247 aa	MW at 27566.2kD

KLTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKNSVSLRKNQSVNVFLGHTAIDE
MLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGF
GMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVWDNR
AHHWVATGIVSWGIGCGEGYDFYTKVLSYVDWIKGVMNGKNKL

NOV3d, 199842665 DNA Sequence	SEQ ID NO: 21	1368 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTTGCCCAACCCGGGGCTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGG
GTACCCAGAGCCGTATGGCAAAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTG
TGAGGCTCGTCTTCCAGGACTTCGACCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATC
TCATTCGTCGGTTCGGATCCAAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGG
TCAGAGGGAGTTTGTATCCTCAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGA
ACAAGACTGCCCACCTCCACAAGGGCTTCCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAG
CCCATCAGCGAGGCCAGCAGGGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCA
GAACCACTGCCAGGAGCCCTATTATCAGGCCGCGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGA
CCTGGAAAGACAGACAGGATGGGGAGGAGGTTCTTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACC
CCCATTGCCCAGAATCAGACGACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGC
CTTCACCAGTATCCACGGCCGTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCC
ACACCGTCTACCCCAAGGACAGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCAC
ACAGCCATAGATGAGATGCTGAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTA
CCGTCAGAATGAGTCCCATAACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCC
TGGGCCCCAACGTCCTCCCGGTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGC
TACGTCAGTGGGTTTGGCATGGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGT
AGCTCCCAGGGAGGCCTGCAACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATA
TGTTCTGTGTTGGGGATGAGACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTG
GTATGGGACAATCATGCCCATCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGA
AGGGTATGACTTCTACACCAAGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGA
ATAAGCTT

NOV3d, 199842665 Protein Sequence	SEQ ID NO: 22	456 aa	MW at 49918.7kD

KLCPTRGSVLLAQELPQQLTSPGYPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTI
SFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQ
PISEASRGSEAINAPGDNPAKVQNHCQEPYYQAAAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPVT
PIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTVYPKDSVSLRKNQSVNVFLGH
TAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLG
YVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYV
VWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYVDWIKGVMNGKNKL

NOV3e, 199842661 DNA Sequence	SEQ ID NO: 23	1368 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTTGCCCAACCCGGGGCTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGG
GTACCCAGAGCCGTATGGCAAAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTG
TGAGGCTCGTCTTCCAGGACTTCGACCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATC
TCATTCGTCGGTTCGGATCCAAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGG
TCAGAGGGAGTTTGTATCCTCAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGG
ACAAGACTGCCCACCTCCACAAGGGCTTCCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAG
CCCATCAGCGAGGCCAGCAGGGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCA
GAACCACTGCCAGGAGCCCTATTATCAGGCCGCGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGA

CCTGGAAAGACAGACAGGATGGGGAGGAGGTTCTTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACC
CCCATTGCCCAGAATCAGACGACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGC
CTTCACCAGTATCCACGGCCGTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCC
ACACCATCTACCCCAAGGACAGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCAC
ACAGCCATAGATGAGATGCTGAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTA
CCGTCAGAATGAGTCCCATAACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCC
TGGGCCCCAACGTCCTCCCGGTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGC
TACGTCAGTGGGTTTGGCATGGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGT
AGCTCCCAGGGAGGCCTGCAACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATA
TGTTCTGTGTTGGGGATGAGACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTG
GTATGGGACAATCATGCCCATCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGA
AGGGTATGACTTCTACACCAAGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGA
ATAAGCTT

NOV3e, 199842661 Protein Sequence	SEQ ID NO: 24	456 aa	MW at 49933.7kD

KLCPTRGSVLLAQELPQQLTSPGYPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTI
SFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLRLTFRTQPSSEDKTAHLHKGFLALYQTVAVNYSQ
PISEASRGSEAINAPGDNPAKVQNHCQEPYYQAAAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPVT
PIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKNQSVNVFLGH
TAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLG
YVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYV
VWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYVDWIKGVMNGKNKL

NOV3f, 199597024 DNA Sequence	SEQ ID NO: 25	1479 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTCCCACCATGCCTGGACCCAGAGTGTGGGGGAAATATCTCTGGAGAAGCCCTCACTCCAAAGG
CTGTCCAGGCGCAATGTGGTGGCTGCTTCTCTGGGGAGTCCTCCAGGCTTGCCCAACCCGGGGCTCCG
TCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGGGTACCCAGAGCCGTATGGCAAAGGC
CAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTGTGAGGCTCGTCTTCCAGGACTTCGA
CCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAGTCTCATTCGTCGGTTCGGATCCAAGCC
AGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGGTCAGAGGGAGTTTGTATCCTCAGGG
AGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGAACAAGACTGCCCACCTCCACAAGGG
CTTCCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAGCCCATCAGCGAGGCCAGCAGGGGCT
CTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCAGAACCACTGCCAGGAGCCCTATTAT
CAGGCCGCGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGACCTGGAAAGACAGACAGGATGGGGA
GGAGGTTCTTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACCCCCATTGCCCAGAATCAGACGACCC
TCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGCCTTCACCAGTATCCACGGCCGTGGG
GGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCCACACCATCTACCCCAAGGACAGTGT
TTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCACACAGCCATAGATGAGATGCTGAAAC
TGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTACCGTCAGAATGAGTCCCATAACTTT
AGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCCTGGGCCCCAACGTCCTCCCGGTCTG
TCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGCTACGTCAGTGGGTTTGGCATGGAGA
TGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGTAGCTCCCAGGGAGGCCTGCAACGCC
TGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATATGTTCTGTGTTGGGGATGAGACGCA
AAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTGGTATGGGACAATCATGCCCATCACT
GGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGAAGGGTATGACTTCTACACCAAGGTG
CTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGAATAAGCTT

NOV3f, 199597024 Protein Sequence	SEQ ID NO: 26	493 aa	MW at 54164.8kD

KLPTMPGPRVWGKYLWRSPHSKGCPGAMWWLLLWGVLQACPTRGSVLLAQELPQQLTSPGYPEPYGKG
QESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTVSFVGSDPSQFCGQQGSPLGRPPGQREFVSSG
RSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASRGSEAINAPGDNPAKVQNHCQEPYY
QAAAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPVTPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRG
GGALLGDRWILTAAHTIYPKDSVSLRKNQSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNF
SGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNA
WLQKRQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKV
LSYVDWIKGVMNGKNKL

NOV3g, 199842653 DNA Sequence	SEQ ID NO: 27	1293 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTTGCCCAACCCGGGGCTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGG
GTACCCAGAGCCGTATGGCAAAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTG
TGAGGCTCGTCTTCCAGGACTTCGACCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATC
TCATTCGTCGGTTCGGATCCAAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGG
TCAGAGGGAGTTTGTATCCTCAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGA
ACAAGACTGCCCACCTCCACAAGGGCTTCCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAG
CCCATCAGCGAGGCCAGCAGGGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCA
GAACCACTGCCAGGAGCCCTATTATCAGGCCGCGGCAGCAGTCTGCGGACGGCCAGTCACCCCCATTG
CCCAGAATCAGACGACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGCCTTCACC
AGTATCCACGGCCGTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCCACACCAT
CTACCCCAAGGACAGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCACACAGCCA
TAGATGAGATGCTGAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTACCGTCAG
AATGAGTCCCATAACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCCTGGGCCC
CAACGTCCTCCCGGTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGCTACGTCA
GTGGGTTTGGCATGGAGATGGGCCGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGTAGCTCCC
AGGGAGGCCTGCAACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATATGTTCTG
TGTCGGGGATGAGACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTGGTATGGG
ACAATCATGCCCATCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGAAGGGTAT
GACTTCTACACCAAGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGAATAAGCT
T

NOV3g, 199842653 Protein Sequence	SEQ ID NO: 28	431 aa	MW at 47213.7kD

KLCPTRGSVLLAQELPQQLTSPGYPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTI
SFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQ
PISEASRGSEAINAPGDNPAKVQNHCQEPYYQAAAAVCGRPVTPIAQNQTTLGSSRAKLGNFPWQAFT
SIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKNQSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQ
NESHNFSGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMEMGRLTTELKYSRLPVAP
REACNAWLQKRQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGY
DFYTKVLSYVDWIKGVMNGKNKL

NOV3h, 199652830 DNA Sequence	SEQ ID NO: 29	1368 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTTGCCCAACCCGGGGCTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGG
GTACCCAGAGCCGTATGGCAAAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTG
TGAGGCTCGTCTTCCAGGACTTCGACCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATC
TCATTCGTCGGTTCGGATCCAAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGG
TCAGAGGGAGTTTGTATCCTCAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGA
ACAAGACTGCCCACCTCCACAAGGGCTTCCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAG
CCCATCAGCGAGGCCAGCAGGGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCCGCCAAGGTCCA
GAACCACTGCCAGGAGCCCTATTATCAGGCCGCGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGA
CCTGGGAAGACAGACAGGATGGGGAAGAGGTTCTTCAGTGTATGCCTGTCTGTGGACGGCCAGTCACC
CCCATTGCCCAGAATCAGACGACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGC
CTTCACCAGTATCCACGGCCGTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCC
ACACCATCTACCCCAAGGACAGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCAC
ACAGCCATAGATGAGATGCTGAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTA
CCGTCAGAATGAGTCCCATAACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCC
TGGGCCCCAACGTCCTCCCGGTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGC
TACGTCAGTGGGTTTGGCATGGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGT
AGCTCCCAGGGAGGCCTGCAACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATA
TGTTCTGTGTTGGGGATGAGACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTG
GTATGGGACAATCATGCCCATCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGA
AGGGTATGACTTCTACACCAAGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGA
ATAAGCTT

NOV3h, 199652830 Protein Sequence	SEQ ID NO: 30	456 aa	MW at 49933.7kD

KLCPTRGSVLLAQELPQQLTSPGYPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTI
SFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQ
PISEASRGSEAINAPGDNPAKVQNHCQEPYYQAAAAGALTCATPGTWEDRQDGEEVLQCMPVCGRPVT
PIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKNQSVNVFLGH
TAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLG
YVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYV
VWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYVDWIKGVMNGKNKL

NOV3i, 199652835 DNA Sequence	SEQ ID NO: 31	1368 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTTGCCCAACCCGGGGCTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGG
GTACCCAGAGCCGTATGGCAAAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTG
TGAGGCTCGTCTTCCAGGACTTCGACCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATC
TCATTCGTCGGTTCGGATCCAAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGG
TCAGAGGGAGTTTGTATCCTCAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGA
ACAAGACTGCCCACCTCCACAAGGGCTTCCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAG
CCCATCAGCGAGGCCAGCAGGGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCA
GAACCACTGCCAGGAGCCCTATTATCAGGCCGCGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGA
CCTGGAAAGACAGACAGGATGGGGAGGAGGTTCTTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACC
CCCATTGCCCAGAATCAGACGACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGC
CTTCACCAGTATCCACGGCCGTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCC
ACACCATCTACCCCAAGGACAGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCAC
ACAGCCATAGATGAGATGCTGAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTA
CCGTCAGAATGAGTCCCATAACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCGGCACAGCATCCCCC
TGGGCCCCAACGTCCTCCCGGTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGC
TACGTCAGTGGGTTTGGCATGGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGT
AGCTCCCAGGGAGGCCTGCAACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATA
TGTTCTGTGTTGGGGATGGGACACAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTG
GTATGGGACAATCATGCCCATCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGA
AGGGTATGACTTCTACACCAAGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGA
ATAAGCTT

NOV3i, 199652835 Protein Sequence	SEQ ID NO: 32	456 aa	MW at 49888.7kD

KLCPTRGSVLLAQELPQQLTSPGYPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTI
SFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQ
PISEASRGSEAINAPGDNPAKVQNHCQEPYYQAAAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPVT
PIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKNQSVNVFLGH
TAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELRHSIPLGPNVLPVCLPDNETLYRSGLLG
YVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFSDNMFCVGDGTQRHSVCQGDSGSVYV
VWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYVDWIKGVMNGKNKL

NOV3j, 198306308 DNA Sequence	SEQ ID NO: 33	387 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTTGCCCAACCCGGGGCTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGG
GTACCCAGAGCCGTATGGCAAAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTG
TGAGGCTCGTCTTCCAGGACTTCGACCTGGAGCCGTCCCAGGACTGTGCGGGGGACTCTGTCACAATC
TCATTCGTCGGTTCGGATCCAAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGG
TCAGAGGGAGTTTGTATCCTCAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGA
ACAAGACTGCCCACCTTCACAAGGGCTTTCTGGCCCTCTACAAGCTT

NOV3j, 198306308 Protein Sequence	SEQ ID NO: 34	129 aa	MW at 14001.5kD

KLCPTRGSVLLAQELPQQLTSPGYPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTI

SFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYKL		
NOV3k, CG102348-02 DNA Sequence	SEQ ID NO: 35 ORF Start: at 7	387 bp ORF Stop: at 382
AAGCTTTGCCCAACCCGGGGCTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGG GTACCCAGAGCCGTATGGCAAAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTG TGAGGCTCGTCTTCCAGGACTTCGACCTGGAGCCGTCCCAGGACTGTGCGGGGGACTCTGTCACAATC TCATTCGTCGGTTCGGATCCAAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGG TCAGAGGGAGTTTGTATCCTCAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGA ACAAGACTGCCCACCTTCACAAGGGCTTTCTGGCCCTCTACAAGCTT		
NOV3k, CG102348-02 Protein Sequence	SEQ ID NO: 36	125 aa MW at 13518.9kD
CPTRGSVLLAQELPQQLTSPGYPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISF VGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALY		
NOV3l, CG102348-03 DNA Sequence	SEQ ID NO: 37 ORF Start: at 7	741 bp ORF Stop: at 736
AAGCTTACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGCCTTCACCAGTATCCA CGGCCGTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCCACACCATCTACCCCA AGAACAGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCACACAGCCATAGATGAG ATGCTGAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTACCGTCAGAATGAGTC CCATAACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCCTGGGCCCCAACGTCC TCCCGGTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGCTACGTCAGTGGGTTT GGCATGGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGTAGCTCCCAGGGAGGC CTGCAACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATATGTTCTGTGTTGGGG ATGAGACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTGGTATGGGACAATCGT GCCCATCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGAAGGGTATGACTTCTA CACCAAGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGAATAAGCTT		
NOV3l, CG102348-03 Protein Sequence	SEQ ID NO: 38	243 aa MW at 27083.5kD
TLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKNSVSLRKNQSVNVFLGHTAIDEML KLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGM EMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVWDNRAH HWVATGIVSWGIGCGEGYDFYTKVLSYVDWIKGVMNGKN		
NOV3m, CG102348-04 DNA Sequence	SEQ ID NO: 39 ORF Start: at 7	1293 bp ORF Stop: at 1288
AAGCTTTGCCCAACCCGGGGCTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGG GTACCCAGAGCCGTATGGCAAAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTG TGAGGCTCGTCTTCCAGGACTTCGACCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATC TCATTCGTCGGTTCGGATCCAAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGG TCAGAGGGAGTTTGTATCCTCAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGA ACAAGACTGCCCACCTCCACAAGGGCTTCCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAG CCCATCAGCGAGGCCAGCAGGGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCA GAACCACTGCCAGGAGCCCTATTATCAGGCCGCGGCAGCAGTCTGCGGACGGCCAGTCACCCCCATTG CCCAGAATCAGACGACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGCCTTCACC AGTATCCACGGCCGTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCCACACCAT CTACCCCAAGGACAGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCACACAGCCA TAGATGAGATGCTGAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTACCGTCAG AATGAGTCCCATAACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCCTGGGCCC CAACGTCCTCCCGGTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGCTACGTCA GTGGGTTTGGCATGGAGATGGGCCGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGTAGCTCCC AGGGAGGCCTGCAACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATATGTTCTG TGTCGGGGATGAGACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTGGTATGGG ACAATCATGCCCATCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGAAGGGTAT GACTTCTACACCAAGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGAATAAGCT	

T

NOV3m, CG102348-04 Protein Sequence	SEQ ID NO: 40	427 aa	MW at 46731.0kD

CPTRGSVLLAQELPQQLTSPGYPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISF
VGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPI
SEASRGSEAINAPGDNPAKVQNHCQEPYYQAAAAVCGRPVTPIAQNQTTLGSSRAKLGNFPWQAFTSI
HGRGGGALLGDRWILTAAHTIYPKDSVSLRKNQSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNE
SHNFSGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMEMGRLTTELKYSRLPVAPRE
ACNAWLQKRQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDF
YTKVLSYVDWIKGVMNGKN

NOV3n, CG102348-05 DNA Sequence	SEQ ID NO: 41	1368 bp
	ORF Start: at 7	ORF Stop: at 1363

AAGCTTTGCCCAACCCGGGGCTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGG
GTACCCAGAGCCGTATGGCAAAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTG
TGAGGTTCGTCTTCCAGGACTTCGACCTGGAGTCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATC
TCATTCGTCGGTTCGGATCCAAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGG
TCAGAGGGAGTTTGTATCCTCAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGA
ACAAGACTGCCCACCTCCACAAGGGCTTCCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAG
CCCATCAGCGAGGCCAGCAGGGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCA
GAACCACTGCCAGGAGCCCTATTATCAGGCCACGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGA
CCTGGAAAGACAGACAGGATGGGGAGGAGGTTCTTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACC
CCCATTGCCCAGAATCAGACGACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGC
CTTCACCAGTATCCACGGCCGTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCC
ACACCATCTACCCCAAGGACAGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCAC
ACAGCCATAGATGAGATGCTGAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTA
CCGTCAGAATGAGTCCCATAACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCC
TGGGCCCCAACGTCCTCCCGGTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGC
TACGTCAGTGGGTTTGGCATGGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGT
AGCTCCCAGGGAGGCCTGCAACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATA
TGTTCTGTGTTGGGGATGAGACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTG
GTATGGGACAATCATGCCCATCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGA
AGGGTATGACTTCTACACCAAGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGA
ATAAGCTT

NOV3n, CG102348-05 Protein Sequence	SEQ ID NO: 42	452 aa	MW at 49504.1kD

CPTRGSVLLAQELPQQLTSPGYPEPYGKGQESSTDIKAPEGFAVRFVFQDFDLESSQDCAGDSVTISF
VGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPI
SEASRGSEAINAPGDNPAKVQNHCQEPYYQATAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPVTPI
AQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKNQSVNVFLGHTA
IDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYV
SGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVW
DNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYVDWIKGVMNGKN

NOV3o, CG102348-06 DNA Sequence	SEQ ID NO: 43	1479 bp
	ORF Start: ATG at 13	ORF Stop: at 1474

AAGCTTCCCACC**ATG**CCTGGACCCAGAGTGTGGGGGAAATATCTCTGGAGAAGCCCTCACTCCAAAGG
CTGTCCAGGCGCAATGTGGTGGCTGCTTCTCTGGGGAGTCCTCCAGGCTTGCCCAACCCGGGGCTCCG
TCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGGGTACCCAGAGCCGTATGGCAAAGGC
CAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTGTGAGGCTCGTCTTCCAGGACTTCGA
CCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATCTCATTCGTCGGTTCGGATCCAAGCC
AGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGGTCAGAGGGAGTTTGTATCCTCAGGG
AGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGAACAAGACTGCCCACCTCCACAAGGG
CTTCCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAGCCCATCAGCGAGGCCAGCAGGGGCT
CTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCAGAACCACTGCCAGGAGCCCTATTAT
CAGGCCACGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGACCTGGAAAGACAGACAGGATGGGGA

GGAGGTTCTTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACCCCCATTGCCCAGAATCAGACGACCC
TCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGCCTTCACCAGTATCCACGGCCGTGGG
GGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCCACACCATCTACCCCAAGGACAGTGT
TTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCACACAGCCATAGATGAGATGCTGAAAC
TGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTACCGTCAGAATGAGTCCCATAACTTT
AGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCCTGGGCCCCAACGTCCTCCCGGTCTG
TCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGCTACGTCAGTGGGTTTGGCATGGAGA
TGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGTAGCTCCCAGGGAGGCCTGCAACGCC
TGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATATGTTCTGTGTTGGGGATGAGACGCA
AAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTGGTATGGGACAATCATGCCCATCACT
GGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGAAGGGTATGACTTCTACACCAAGGTG
CTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGAATAAGCTT

NOV3o, CG102348-06 Protein Sequence	SEQ ID NO: 44	487 aa	MW at 53528.0kD

MPGPRVWGKYLWRSPHSKGCPGAMWWLLLWGVLQACPTRGSVLLAQELPQQLTSPGYPEPYGKGQESS
TDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLR
LTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASRGSEAINAPGDNPAKVQNHCQEPYYQATA
AGALTCATPGTWKDRQDGEEVLQCMPVCGRPVTPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGAL
LGDRWILTAAHTIYPKDSVSLRKNQSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDI
ALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQK
RQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
DWIKGVMNGKN

NOV3p, SNP13376570 of CG102348-01, DNA Sequence	SEQ ID NO: 45	3345 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 1479
	SNP Pos: 343	SNP Change: G to A

CAGATGTCCAGTTCCAG**ATG**CCTGGACCCAGAGTGTGGGGGAAATATCTCTGGAGAAGCCCTCACTCC
AAAGGCTGTCCAGGCGCAATGTGGTGGCTGCTTCTCTGGGGAGTCCTCCAGGCTTGCCCAACCCGGGG
CTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGGGTACCCAGAGCCGTATGGCA
AAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTGTGAGGCTCGTCTTCCAGGAC
TTCGACCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATCTCATTCGTCGGTTCGGATCC
AAACCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGGTCAGAGGGAGTTTGTATCCT
CAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGAACAAGACTGCCCACCTTCAC
AAGGGCTTTCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAGCCCATCAGCGAGGCCAGCAG
GGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCAGAACCACTGCCAGGAGCCCT
ATTATCAGGCCGCGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGACCTGGAAAGACAGACAGGAT
GGGGAGGAGGTTCTTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACCCCCATTGCCCAGAATCAGAC
GACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGCCTTCACCAGTATCCACGGCC
GTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCCACACCGTCTACCCCAAGGAC
AGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCACACAGCCATAGATGAGATGCT
GAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTACCGTCAGAATGAGTCCCATA
ACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCCTGGGCCCCAACGTCCTCCCG
GTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGCTACGTCAGTGGGTTTGGCAT
GGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGTAGCTCCCAGGGAGGCCTGCA
ACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATATGTTCTGTGTTGGGGATGAG
ACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTGGTATGGGACAATCATGCCCA
TCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGAAGGGTATGACTTCTACACCA
AGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGAAT**TGA**CCCTGGGGGCTTGAA
CAGGGACTGACCAGCACAGTGGAGGCCCCAGGCAACAGAGGGCCTGGAGTGAGGACTGAACACTGGGG
TAGGGGTTGGGGGTGGGGGGTTGGGGGAGGCAGGGAAATCCTATTCACATCACTGTTGCACCAAGCCA
CTGCAAGAGAAACCCCCACCCGGCAAGCCCGCCCCATCCCAGACAGGAAGCAGAGTCCCACAGACCGC
TCCTCCTCACCCTCTACCTCCCTGTGCTCATGCACTAGGCCCCGGGAAGCCTGTACATCTCAACAACT
TTCGCCTTGAATGTCCTTAGAACCACCTTCCCCTACTTCATCTGTTGACACAGCTTTTATACTCACCT
GTGGAAGAGTCAGCTACTCACCCGCTATTAGAGTATGGAGGAAGGGGTTTTCATTGCATTGCATTTCT
GAAACATTCCTAAGACCCTTTAGTTGACCTTCAAATATTCAAGCTATTCTGCAGCTCCAAGATGCAAT
TATAGAAACAGCTCCTTTTTTATTTTATGTCCTCTATATGCCAGGTGCTTCACCTGTTATTTCACTTA
ATCCTCATACCATATTTGCAAAGGATGTGTTATTATCTATGTGTGACAAATGAGGAAACTGAGGCTCA

GGGGATAAAGGGACTTGCCCAAGTCCCACAGCTGGTGTGTGACTGCAGAGACTGTGCTCTTCCCAGTG
TGCTGCAATACTTCTCAACCCTCCTCTAACCTGCTGTGTCACCCGCTTTCCCTCCCAGCCCCCACATC
CTTACCATTTTCCCTCCCTGGGAATTCCTGCTTCTGCGAAAATGGTATCCTCTAGCTCACACTTTCCT
AATGGCCCCATCTCCTGCAGAAGCCAGGTGAGCCCAGCACTGGACTGAAGTTCTTGCAGACACCCCAC
CTGTGCCCCTATCATCAGGGGAACTGCTCCACCTGAGAGGACCAACTCTTTAATTTTTAGTAAAACCT
GAAGGTGATGGGCCGGGCGCAGTGGCTCACGCCTGTAATCCCAACACCTTAGGAGTCCGAGGTGGGTG
GATCACGAGGTCAGGAGATCCAGCCCATCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATAC
AAAAATTAGCCGGGCGTGGTGACACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
CACTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCTAAGATCACGCCACTGCACTCCAGCCTGCGGACA
GACCAAGACTTCATCCCCCCCAAAAAAAAAAGATTGGAGGTGATTTACAGTGAAAGACACAAATAAAA
TACAACTGTTCAATGGAAATAGAAAATAAACACCATAAAAGAGAGAAGAGAGGTAATTTGTTAGCATC
AAGAGTCAAGTTGCTATATGGTCAAAGGTTAAATTTATCTCTAAAAAATGGCAGGATTCAAAGTTGTA
CATACATGTGATTACTTCTGTTTTTTACACCCACATACAGTACAAAAGATTATTAAAAATATTCCCAA
AAGGCAGGTGCAATGATGCACACTTATACCCCCAGCCACTCAGGAGGCTGATGCAAGAGGATCGCTTG
AGCCCAGGAGTTGAAGTCCAGCCTAAGCAACATAGTGAAACCCCATCTCCAAAAATATAATAATAATT
CTCTCAAAATACTAAACAGAGGTGGTTTTATTGATAAGATTTTGGCTGTTTGGTTTTCCACTATTCTC
TATTGGCTAAAATTTGTTTAATGAGCATGAAATGTTTTTATTTTATTTTGCTTATTTTTATGATTGCA
AAAAATGATATGAGTTTCTCCCTGCCAAGGCAAAAAATATATATATATACCTATAAAAAAAAAAAAAA
AAAAAAAAAAAAA

NOV3p, SNP13376570 of CG102348-01, Protein Sequence	SEQ ID NO: 46	487 aa	MW at 53510.9kD
	SNP Pos: 109		SNP Change: Ser to Asn

MPGPRVWGKYLWRSPHSKGCPGAMWWLLLWGVLQACPTRGSVLLAQELPQQLTSPGYPEPYGKGQESS
TDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPNQFCGQQGSPLGRPPGQREFVSSGRSLR
LTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASRGSEAINAPGDNPAKVQNHCQEPYYQAAA
AGALTCATPGTWKDRQDGEEVLQCMPVCGRPVTPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGAL
LGDRWILTAAHTVYPKDSVSLRKNQSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDI
ALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQK
RQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
DWIKGVMNGKN

NOV3q, SNP13376568 of CG102348-01, DNA Sequence	SEQ ID NO: 47	3345 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 1479
	SNP Pos: 564	SNP Change: G to A

CAGATGTCCAGTTCCAGATGCCTGGACCCAGAGTGTGGGGGAAATATCTCTGGAGAAGCCCTCACTCC
AAAGGCTGTCCAGGCGCAATGTGGTGGCTGCTTCTCTGGGGAGTCCTCCAGGCTTGCCCAACCCGGGG
CTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGGGTACCCAGAGCCGTATGGCA
AAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTGTGAGGCTCGTCTTCCAGGAC
TTCGACCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATCTCATTCGTCGGTTCGGATCC
AAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGGTCAGAGGGAGTTTGTATCCT
CAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGAACAAGACTGCCCACCTTCAC
AAGGGCTTTCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAGCCCATCAGCGAGGCCAGCAG
GGGCTCTGAGGCCATCAACACACCTGGAGACAACCCTGCCAAGGTCCAGAACCACTGCCAGGAGCCCT
ATTATCAGGCCGCGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGACCTGGAAAGACAGACAGGAT
GGGGAGGAGGTTCTTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACCCCCATTGCCCAGAATCAGAC
GACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGCCTTCACCAGTATCCACGGCC
GTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCCACACCGTCTACCCCAAGGAC
AGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCACACAGCCATAGATGAGATGCT
GAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTACCGTCAGAATGAGTCCCATA
ACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCCTGGGCCCCAACGTCCTCCCG
GTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGCTACGTCAGTGGGTTTGGCAT
GGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGTAGCTCCCAGGGAGGCCTGCA
ACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATATGTTCTGTGTTGGGGATGAG
ACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTGGTATGGGACAATCATGCCCA
TCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGAAGGGTATGACTTCTACACCA
AGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGAATTGACCCTGGGGGCTTGAA
CAGGGACTGACCAGCACAGTGGAGGCCCCAGGCAACAGAGGGCCTGGAGTGAGGACTGAACACTGGGG
TAGGGGTTGGGGGTGGGGGGTTGGGGGAGGCAGGGAAATCCTATTCACATCACTGTTGCACCAAGCCA

CTGCAAGAGAAACCCCCACCCGGCAAGCCCGCCCCATCCCAGACAGGAAGCAGAGTCCCACAGACCGC
TCCTCCTCACCCTCTACCTCCCTGTGCTCATGCACTAGGCCCCGGGAAGCCTGTACATCTCAACAACT
TTCGCCTTGAATGTCCTTAGAACCACCTTCCCCTACTTCATCTGTTGACACAGCTTTTATACTCACCT
GTGGAAGAGTCAGCTACTCACCCGCTATTAGAGTATGGAGGAAGGGGTTTTCATTGCATTGCATTTCT
GAAACATTCCTAAGACCCTTTAGTTGACCTTCAAATATTCAAGCTATTCTGCAGCTCCAAGATGCAAT
TATAGAAACAGCTCCTTTTTTATTTTATGTCCTCTATATGCCAGGTGCTTCACCTGTTATTTCACTTA
ATCCTCATACCATATTTGCAAAGGATGTGTTATTATCTATGTGTGACAAATGAGGAAACTGAGGCTCA
GGGGATAAAGGGACTTGCCCAAGTCCCACAGCTGGTGTGTGACTGCAGAGACTGTGCTCTTCCCAGTG
TGCTGCAATACTTCTCAACCCTCCTCTAACCTGCTGTGTCACCCGCTTTCCCTCCCAGCCCCCACATC
CTTACCATTTTCCCTCCCTGGGAATTCCTGCTTCTGCGAAAATGGTATCCTCTAGCTCACACTTTCCT
AATGGCCCCATCTCCTGCAGAAGCCAGGTGAGCCCAGCACTGGACTGAAGTTCTTGCAGACACCCCAC
CTGTGCCCCTATCATCAGGGGAACTGCTCCACCTGAGAGGACCAACTCTTTAATTTTTAGTAAAACCT
GAAGGTGATGGGCCGGGCGCAGTGGCTCACGCCTGTAATCCCAACACCTTAGGAGTCCGAGGTGGGTG
GATCACGAGGTCAGGAGATCCAGCCCATCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATAC
AAAAATTAGCCGGGCGTGGTGACACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
CACTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCTAAGATCACGCCACTGCACTCCAGCCTGCGGACA
GACCAAGACTTCATCCCCCCCAAAAAAAAAAGATTGGAGGTGATTTACAGTGAAAGACACAAATAAAA
TACAACTGTTCAATGGAAATAGAAAATAAACACCATAAAAGAGAGAAGAGAGGTAATTTGTTAGCATC
AAGAGTCAAGTTGCTATATGGTCAAAGGTTAAATTTATCTCTAAAAAATGGCAGGATTCAAAGTTGTA
CATACATGTGATTACTTCTGTTTTTTACACCCACATACAGTACAAAAGATTATTAAAAATATTCCCAA
AAGGCAGGTGCAATGATGCACACTTATACCCCCAGCCACTCAGGAGGCTGATGCAAGAGGATCGCTTG
AGCCCAGGAGTTGAAGTCCAGCCTAAGCAACATAGTGAAACCCCATCTCCAAAAATATAATAATAATT
CTCTCAAAATACTAAACAGAGGTGGTTTTATTGATAAGATTTTGGCTGTTTGGTTTTCCACTATTCTC
TATTGGCTAAAATTTGTTTAATGAGCATGAAATGTTTTTATTTTATTTTGCTTATTTTTATGATTGCA
AAAAATGATATGAGTTTCTCCCTGCCAAGGCAAAAAATATATATATATACCTATAAAAAAAAAAAAAA
AAAAAAAAAAAAA

NOV3q, SNP13376568 of CG102348-01, Protein Sequence	SEQ ID NO: 48	487 aa	MW at 53513.9kD
	SNP Pos: 183		SNP Change: Ala to Thr

MPGPRVWGKYLWRSPHSKGCPGAMWWLLLWGVLQACPTRGSVLLAQELPQQLTSPGYPEPYGKGQESS
TDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLR
LTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASRGSEAINTPGDNPAKVQNHCQEPYYQAAA
AGALTCATPGTWKDRQDGEEVLQCMPVCGRPVTPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGAL
LGDRWILTAAHTVYPKDSVSLRKNQSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDI
ALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQK
RQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
DWIKGVMNGKN

NOV3r, SNP13382463 of CG102348-01, DNA Sequence	SEQ ID NO: 49	3345 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 1479
	SNP Pos: 693	SNP Change: C to T

CAGATGTCCAGTTCCAG**ATG**CCTGGACCCAGAGTGTGGGGGAAATATCTCTGGAGAAGCCCTCACTCC
AAAGGCTGTCCAGGCGCAATGTGGTGGCTGCTTCTCTGGGGAGTCCTCCAGGCTTGCCCAACCCGGGG
CTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGGGTACCCAGAGCCGTATGGCA
AAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTGTGAGGCTCGTCTTCCAGGAC
TTCGACCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATCTCATTCGTCGGTTCGGATCC
AAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGGTCAGAGGGAGTTTGTATCCT
CAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGAACAAGACTGCCCACCTTCAC
AAGGGCTTTCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAGCCCATCAGCGAGGCCAGCAG
GGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCAGAACCACTGCCAGGAGCCCT
ATTATCAGGCCGCGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGACCTGGAAAGACAGACAGGAT
GGGGAGGAGGTT**T**TTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACCCCCATTGCCCAGAATCAGAC
GACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGCCTTCACCAGTATCCACGGCC
GTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCCACACCGTCTACCCCAAGGAC
AGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCACACAGCCATAGATGAGATGCT
GAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTACCGTCAGAATGAGTCCCATA
ACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCCTGGGCCCCAACGTCCTCCCG
GTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGCTACGTCAGTGGGTTTGGCAT

GGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGTAGCTCCCAGGGAGGCCTGCA
ACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATATGTTCTGTGTTGGGGATGAG
ACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTGGTATGGGACAATCATGCCCA
TCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGAAGGGTATGACTTCTACACCA
AGGTGCTCAGCTATGTGGACTGGATCAAGGGAGTGATGAATGGCAAGAATTGACCCTGGGGGCTTGAA
CAGGGACTGACCAGCACAGTGGAGGCCCCAGGCAACAGAGGGCCTGGAGTGAGGACTGAACACTGGGG
TAGGGGTTGGGGGTGGGGGGTTGGGGGAGGCAGGGAAATCCTATTCACATCACTGTTGCACCAAGCCA
CTGCAAGAGAAACCCCCACCCGGCAAGCCCGCCCCATCCCAGACAGGAAGCAGAGTCCCACAGACCGC
TCCTCCTCACCCTCTACCTCCCTGTGCTCATGCACTAGGCCCCGGGAAGCCTGTACATCTCAACAACT
TTCGCCTTGAATGTCCTTAGAACCACCTTCCCCTACTTCATCTGTTGACACAGCTTTTATACTCACCT
GTGGAAGAGTCAGCTACTCACCCGCTATTAGAGTATGGAGGAAGGGGTTTTCATTGCATTGCATTTCT
GAAACATTCCTAAGACCCTTTAGTTGACCTTCAAATATTCAAGCTATTCTGCAGCTCCAAGATGCAAT
TATAGAAACAGCTCCTTTTTTATTTTATGTCCTCTATATGCCAGGTGCTTCACCTGTTATTTCACTTA
ATCCTCATACCATATTTGCAAAGGATGTGTTATTATCTATGTGTGACAAATGAGGAAACTGAGGCTCA
GGGGATAAAGGGACTTGCCCAAGTCCCACAGCTGGTGTGTGACTGCAGAGACTGTGCTCTTCCCAGTG
TGCTGCAATACTTCTCAACCCTCCTCTAACCTGCTGTGTCACCCGCTTTCCCTCCCAGCCCCCACATC
CTTACCATTTTCCCTCCCTGGGAATTCCTGCTTCTGCGAAAATGGTATCCTCTAGCTCACACTTTCCT
AATGGCCCCATCTCCTGCAGAAGCCAGGTGAGCCCAGCACTGGACTGAAGTTCTTGCAGACACCCCAC
CTGTGCCCCTATCATCAGGGGAACTGCTCCACCTGAGAGGACCAACTCTTTAATTTTTAGTAAAACCT
GAAGGTGATGGGCCGGGCGCAGTGGCTCACGCCTGTAATCCCAACACCTTAGGAGTCCGAGGTGGGTG
GATCACGAGGTCAGGAGATCCAGCCCATCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATAC
AAAAATTAGCCGGGCGTGGTGACACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
CACTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCTAAGATCACGCCACTGCACTCCAGCCTGCGGACA
GACCAAGACTTCATCCCCCCCCAAAAAAAAAAAGATTGGAGGTGATTTACAGTGAAAGACACAAATAAAA
TACAACTGTTCAATGGAAATAGAAAATAAACACCATAAAAGAGAGAAGAGAGGTAATTTGTTAGCATC
AAGAGTCAAGTTGCTATATGGTCAAAGGTTAAATTTATCTCTAAAAAATGGCAGGATTCAAAGTTGTA
CATACATGTGATTACTTCTGTTTTTTACACCCACATACAGTACAAAAGATTATTAAAAATATTCCCAA
AAGGCAGGTGCAATGATGCACACTTATACCCCCAGCCACTCAGGAGGCTGATGCAAGAGGATCGCTTG
AGCCCAGGAGTTGAAGTCCAGCCTAAGCAACATAGTGAAACCCCATCTCCAAAAATATAATAATAATT
CTCTCAAAATACTAAACAGAGGTGGTTTTATTGATAAGATTTTGGCTGTTTGGTTTTCCACTATTCTC
TATTGGCTAAAATTTGTTTAATGAGCATGAAATGTTTTTATTTTATTTTGCTTATTTTTATGATTGCA
AAAAATGATATGAGTTTCTCCCTGCCAAGGCAAAAAATATATATATATACCTATAAAAAAAAAAAAAA
AAAAAAAAAAAAA

NOV3r, SNP13382463 of CG102348-01, Protein Sequence	SEQ ID NO: 50	487 aa	MW at 53517.9kD
	SNP Pos: 226		SNP Change: Leu to Phe

MPGPRVWGKYLWRSPHSKGCPGAMWWLLLWGVLQACPTRGSVLLAQELPQQLTSPGYPEPYGKGQESS
TDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLR
LTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASRGSEAINAPGDNPAKVQNHCQEPYYQAAA
AGALTCATPGTWKDRQDGEEVFQCMPVCGRPVTPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGAL
LGDRWILTAAHTVYPKDSVSLRKNQSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDI
ALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQK
RQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
DWIKGVMNGKN

NOV3s, SNP13374245 of CG102348-01, DNA Sequence	SEQ ID NO: 51	3345 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 1479
	SNP Pos: 1459	SNP Change: G to A

CAGATGTCCAGTTCCAGATGCCTGGACCCAGAGTGTGGGGGAAATATCTCTGGAGAAGCCCTCACTCC
AAAGGCTGTCCAGGCGCAATGTGGTGGCTGCTTCTCTGGGGAGTCCTCCAGGCTTGCCCAACCCGGGG
CTCCGTCCTCTTGGCCCAAGAGCTACCCCAGCAGCTGACATCCCCCGGGTACCCAGAGCCGTATGGCA
AAGGCCAAGAGAGCAGCACGGACATCAAGGCTCCAGAGGGCTTTGCTGTGAGGCTCGTCTTCCAGGAC
TTCGACCTGGAGCCGTCCCAGGACTGTGCAGGGGACTCTGTCACAATCTCATTCGTCGGTTCGGATCC
AAGCCAGTTCTGTGGTCAGCAAGGCTCCCCTCTGGGCAGGCCCCCTGGTCAGAGGGAGTTTGTATCCT
CAGGGAGGAGTTTGCGGCTGACCTTCCGCACACAGCCTTCCTCGGAGAACAAGACTGCCCACCTTCAC
AAGGGCTTTCTGGCCCTCTACCAAACCGTGGCTGTGAACTATAGTCAGCCCATCAGCGAGGCCAGCAG
GGGCTCTGAGGCCATCAACGCACCTGGAGACAACCCTGCCAAGGTCCAGAACCACTGCCAGGAGCCCT
ATTATCAGGCCGCGGCAGCAGGGGCACTCACCTGTGCAACCCCAGGGACCTGGAAAGACAGACAGGAT

GGGGAGGAGGTTCTTCAGTGTATGCCTGTCTGCGGACGGCCAGTCACCCCCATTGCCCAGAATCAGAC
GACCCTCGGTTCTTCCAGAGCCAAGCTGGGCAACTTCCCCTGGCAAGCCTTCACCAGTATCCACGGCC
GTGGGGGCGGGGCCCTGCTGGGGGACAGATGGATCCTCACTGCTGCCCACACCGTCTACCCCAAGGAC
AGTGTTTCTCTCAGGAAGAACCAGAGTGTGAATGTGTTCTTGGGCCACACAGCCATAGATGAGATGCT
GAAACTGGGGAACCACCCTGTCCACCGTGTCGTTGTGCACCCCGACTACCGTCAGAATGAGTCCCATA
ACTTTAGCGGGGACATCGCCCTCCTGGAGCTGCAGCACAGCATCCCCCTGGGCCCCAACGTCCTCCCG
GTCTGTCTGCCCGATAATGAGACCCTCTACCGCAGCGGCTTGTTGGGCTACGTCAGTGGGTTTGGCAT
GGAGATGGGCTGGCTAACTACTGAGCTGAAGTACTCGAGGCTGCCTGTAGCTCCCAGGGAGGCCTGCA
ACGCCTGGCTCCAAAAGAGACAGAGACCCGAGGTGTTTTCTGACAATATGTTCTGTGTTGGGGATGAG
ACGCAAAGGCACAGTGTCTGCCAGGGGGACAGTGGCAGCGTCTATGTGGTATGGGACAATCATGCCCA
TCACTGGGTGGCCACGGGCATTGTGTCCTGGGGCATAGGGTGTGGCGAAGGGTATGACTTCTACACCA
AGGTGCTCAGCTATGTGGACTGGATCAAGGAAGTGATGAATGGCAAGAATTGACCCTGGGGGCTTGAA
CAGGGACTGACCAGCACAGTGGAGGCCCCAGGCAACAGAGGGCCTGGAGTGAGGACTGAACACTGGGG
TAGGGGTTGGGGGTGGGGGGTTGGGGGAGGCAGGGAAATCCTATTCACATCACTGTTGCACCAAGCCA
CTGCAAGAGAAACCCCCACCCGGCAAGCCCGCCCCATCCCAGACAGGAAGCAGAGTCCCACAGACCGC
TCCTCCTCACCCTCTACCTCCCTGTGCTCATGCACTAGGCCCCGGGAAGCCTGTACATCTCAACAACT
TTCGCCTTGAATGTCCTTAGAACCACCTTCCCCTACTTCATCTGTTGACACAGCTTTTATACTCACCT
GTGGAAGAGTCAGCTACTCACCCGCTATTAGAGTATGGAGGAAGGGGTTTTCATTGCATTGCATTTCT
GAAACATTCCTAAGACCCTTTAGTTGACCTTCAAATATTCAAGCTATTCTGCAGCTCCAAGATGCAAT
TATAGAAACAGCTCCTTTTTTATTTTATGTCCTCTATATGCCAGGTGCTTCACCTGTTATTTCACTTA
ATCCTCATACCATATTTGCAAAGGATGTGTTATTATCTATGTGTGACAAATGAGGAAACTGAGGCTCA
GGGGATAAAGGGACTTGCCCAAGTCCCACAGCTGGTGTGTGACTGCAGAGACTGTGCTCTTCCCAGTG
TGCTGCAATACTTCTCAACCCTCCTCTAACCTGCTGTGTCACCCGCTTTCCCTCCCAGCCCCCACATC
CTTACCATTTTCCCTCCCTGGGAATTCCTGCTTCTGCGAAAATGGTATCCTCTAGCTCACACTTTCCT
AATGGCCCCATCTCCTGCAGAAGCCAGGTGAGCCCAGCACTGGACTGAAGTTCTTGCAGACACCCCAC
CTGTGCCCCTATCATCAGGGGAACTGCTCCACCTGAGAGGACCAACTCTTTAATTTTTAGTAAAACCT
GAAGGTGATGGGCCGGGCGCAGTGGCTCACGCCTGTAATCCCAACACCTTAGGAGTCCGAGGTGGGTG
GATCACGAGGTCAGGAGATCCAGCCCATCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATAC
AAAAATTAGCCGGGCGTGGTGACACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
CACTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCTAAGATCACGCCACTGCACTCCAGCCTGCGGACA
GACCAAGACTTCATCCCCCCCAAAAAAAAAAGATTGGAGGTGATTTACAGTGAAAGACACAAATAAAA
TACAACTGTTCAATGGAAATAGAAAATAAACACCATAAAAGAGAGAAGAGAGGTAATTTGTTAGCATC
AAGAGTCAAGTTGCTATATGGTCAAAGGTTAAATTTATCTCTAAAAAATGGCAGGATTCAAAGTTGTA
CATACATGTGATTACTTCTGTTTTTTACACCCACATACAGTACAAAAGATTATTAAAAATATTCCCAA
AAGGCAGGTGCAATGATGCACACTTATACCCCCAGCCACTCAGGAGGCTGATGCAAGAGGATCGCTTG
AGCCCAGGAGTTGAAGTCCAGCCTAAGCAACATAGTGAAACCCCATCTCCAAAAATATAATAATAATT
CTCTCAAAATACTAAACAGAGGTGGTTTTATTGATAAGATTTTGGCTGTTTGGTTTTCCACTATTCTC
TATTGGCTAAAATTTGTTTAATGAGCATGAAATGTTTTTATTTTATTTTGCTTATTTTTATGATTGCA
AAAAATGATATGAGTTTCTCCCTGCCAAGGCAAAAAATATATATATACCTATAAAAAAAAAAAAAA
AAAAAAAAAAAAA

NOV3s, SNP13374245 of CG102348-01, Protein Sequence	SEQ ID NO: 52	487 aa	MW at 53556.0kD
	SNP Pos: 481		SNP Change: Gly to Glu

MPGPRVWGKYLWRSPHSKGCPGAMWWLLLWGVLQACPTRGSVLLAQELPQQLTSPGYPEPYGKGQESS
TDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQGSPLGRPPGQREFVSSGRSLR
LTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASRGSEAINAPGDNPAKVQNHCQEPYYQAAA
AGALTCATPGTWKDRQDGEEVLQCMPVCGRPVTPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGAL
LGDRWILTAAHTVYPKDSVSLRKNQSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDI
ALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQK
RQRPEVFSDNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
DWIKEVMNGKN

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 3B.

Table 3B. Comparison of the NOV3 protein sequences.

```
NOV3a   ----MPGPRVWGKYLWRSPHSKGCPGAMWWLLLWGVLQACPTRGSVLLAQELPQQLTSPG
NOV3b   ----------------------------------KLCPTRGSVLLAQELPQQLTSPG
NOV3c   ------------------------------------------------------------
NOV3d   ----------------------------------KLCPTRGSVLLAQELPQQLTSPG
NOV3e   ----------------------------------KLCPTRGSVLLAQELPQQLTSPG
NOV3f   KLPTMPGPRVWGKYLWRSPHSKGCPGAMWWLLLWGVLQACPTRGSVLLAQELPQQLTSPG
NOV3g   ----------------------------------KLCPTRGSVLLAQELPQQLTSPG
NOV3h   ----------------------------------KLCPTRGSVLLAQELPQQLTSPG
NOV3i   ----------------------------------KLCPTRGSVLLAQELPQQLTSPG
NOV3j   ----------------------------------KLCPTRGSVLLAQELPQQLTSPG
NOV3k   ----------------------------------CPTRGSVLLAQELPQQLTSPG
NOV3l   ------------------------------------------------------------
NOV3m   ----------------------------------CPTRGSVLLAQELPQQLTSPG
NOV3n   ----------------------------------CPTRGSVLLAQELPQQLTSPG
NOV3o   ----MPGPRVWGKYLWRSPHSKGCPGAMWWLLLWGVLQACPTRGSVLLAQELPQQLTSPG

NOV3a   YPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQG
NOV3b   YPEPYGKGQESSTDIKAPEGFAVRFVFQDFDLESSQDCAGDSVTISFVGSDPSQFCGQQG
NOV3c   ------------------------------------------------------------
NOV3d   YPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQG
NOV3e   YPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQG
NOV3f   YPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTVSFVGSDPSQFCGQQG
NOV3g   YPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQG
NOV3h   YPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQG
NOV3i   YPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQG
NOV3j   YPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQG
NOV3k   YPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQG
NOV3l   ------------------------------------------------------------
NOV3m   YPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQG
NOV3n   YPEPYGKGQESSTDIKAPEGFAVRFVFQDFDLESSQDCAGDSVTISFVGSDPSQFCGQQG
NOV3o   YPEPYGKGQESSTDIKAPEGFAVRLVFQDFDLEPSQDCAGDSVTISFVGSDPSQFCGQQG

NOV3a   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASR
NOV3b   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASR
NOV3c   ------------------------------------------------------------
NOV3d   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASR
NOV3e   SPLGRPPGQREFVSSGRSLRLTFRTQPSSEDKTAHLHKGFLALYQTVAVNYSQPISEASR
NOV3f   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASR
NOV3g   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASR
NOV3h   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASR
NOV3i   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASR
NOV3j   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYKL--------------
NOV3k   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALY----------------
NOV3l   ------------------------------------------------------------
NOV3m   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASR
NOV3n   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASR
NOV3o   SPLGRPPGQREFVSSGRSLRLTFRTQPSSENKTAHLHKGFLALYQTVAVNYSQPISEASR

NOV3a   GSEAINAPGDNPAKVQNHCQEPYYQAAAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPV
NOV3b   GSEAINAPGDNPAKVQNHCQEPYYQATAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPV
NOV3c   ------------------------------------------------------------
```

```
NOV3d   GSEAINAPGDNPAKVQNHCQEPYYQAAAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPV
NOV3e   GSEAINAPGDNPAKVQNHCQEPYYQAAAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPV
NOV3f   GSEAINAPGDNPAKVQNHCQEPYYQAAAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPV
NOV3g   GSEAINAPGDNPAKVQNHCQEPYYQAAAA----------------------V--CGRPV
NOV3h   GSEAINAPGDNPAKVQNHCQEPYYQAAAAGALTCATPGTWEDRQDGEEVLQCMPVCGRPV
NOV3i   GSEAINAPGDNPAKVQNHCQEPYYQAAAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPV
NOV3j   ------------------------------------------------------------
NOV3k   ------------------------------------------------------------
NOV3l   ------------------------------------------------------------
NOV3m   GSEAINAPGDNPAKVQNHCQEPYYQAAAA----------------------V--CGRPV
NOV3n   GSEAINAPGDNPAKVQNHCQEPYYQATAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPV
NOV3o   GSEAINAPGDNPAKVQNHCQEPYYQATAAGALTCATPGTWKDRQDGEEVLQCMPVCGRPV

NOV3a   TPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTVYPKDSVSLRKN
NOV3b   TPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKN
NOV3c   ------KLTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKNSVSLRKN
NOV3d   TPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTVYPKDSVSLRKN
NOV3e   TPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKN
NOV3f   TPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKN
NOV3g   TPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKN
NOV3h   TPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKN
NOV3i   TPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKN
NOV3j   ------------------------------------------------------------
NOV3k   ------------------------------------------------------------
NOV3l   --------TLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKNSVSLRKN
NOV3m   TPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKN
NOV3n   TPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKN
NOV3o   TPIAQNQTTLGSSRAKLGNFPWQAFTSIHGRGGGALLGDRWILTAAHTIYPKDSVSLRKN

NOV3a   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL
NOV3b   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL
NOV3c   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL
NOV3d   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL
NOV3e   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL
NOV3f   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL
NOV3g   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL
NOV3h   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL
NOV3i   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELRHSIPLGPNVL
NOV3j   ------------------------------------------------------------
NOV3k   ------------------------------------------------------------
NOV3l   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL
NOV3m   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL
NOV3n   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL
NOV3o   QSVNVFLGHTAIDEMLKLGNHPVHRVVVHPDYRQNESHNFSGDIALLELQHSIPLGPNVL

NOV3a   PVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
NOV3b   PVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
NOV3c   PVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
NOV3d   PVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
NOV3e   PVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
NOV3f   PVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
NOV3g   PVCLPDNETLYRSGLLGYVSGFGMEMGRLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
NOV3h   PVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
NOV3i   PVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
```

```
NOV3j   ------------------------------------------------------------
NOV3k   ------------------------------------------------------------
NOV3l   PVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
NOV3m   PVCLPDNETLYRSGLLGYVSGFGMEMGRLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
NOV3n   PVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFS
NOV3o   PVCLPDNETLYRSGLLGYVSGFGMEMGWLTTELKYSRLPVAPREACNAWLQKRQRPEVFS

NOV3a   DNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3b   DNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3c   DNMFCVGDETQRHSVCQGDSGSVYVVWDNRAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3d   DNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3e   DNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3f   DNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3g   DNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3h   DNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3i   DNMFCVGDGTQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3j   ------------------------------------------------------------
NOV3k   ------------------------------------------------------------
NOV3l   DNMFCVGDETQRHSVCQGDSGSVYVVWDNRAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3m   DNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3n   DNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV
NOV3o   DNMFCVGDETQRHSVCQGDSGSVYVVWDNHAHHWVATGIVSWGIGCGEGYDFYTKVLSYV

NOV3a   DWIKGVMNGKN--
NOV3b   DWIKGVMNGKNKL
NOV3c   DWIKGVMNGKNKL
NOV3d   DWIKGVMNGKNKL
NOV3e   DWIKGVMNGKNKL
NOV3f   DWIKGVMNGKNKL
NOV3g   DWIKGVMNGKNKL
NOV3h   DWIKGVMNGKNKL
NOV3i   DWIKGVMNGKNKL
NOV3j   -------------
NOV3k   -------------
NOV3l   DWIKGVMNGKN--
NOV3m   DWIKGVMNGKN--
NOV3n   DWIKGVMNGKN--
NOV3o   DWIKGVMNGKN--

NOV3a  (SEQ ID NO:  16)
NOV3b  (SEQ ID NO:  18)
NOV3c  (SEQ ID NO:  20)
NOV3d  (SEQ ID NO:  22)
NOV3e  (SEQ ID NO:  24)
NOV3f  (SEQ ID NO:  26)
NOV3g  (SEQ ID NO:  28)
NOV3h  (SEQ ID NO:  30)
NOV3i  (SEQ ID NO:  32)
NOV3j  (SEQ ID NO:  34)
NOV3k  (SEQ ID NO:  36)
NOV3l  (SEQ ID NO:  38)
NOV3m  (SEQ ID NO:  40)
NOV3n  (SEQ ID NO:  42)
NOV3o  (SEQ ID NO:  44)
```

Further analysis of the NOV3a protein yielded the following properties shown in Table 3C.

Table 3C. Protein Sequence Properties NOV3a

SignalP analysis:	Cleavage site between residues 36 and 37

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 9;  pos.chg 2;  neg.chg 0
      H-region:  length 3;  peak value  -5.40
      PSG score:  -9.80

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   2.53
      possible cleavage site: between 35 and 36

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  2.97 (at 20)
      ALOM score:   2.97  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       3       Hyd Moment(75):  7.36
      Hyd Moment(95):  6.26    G content:       6
      D/E content:     1       S/T content:     4
      Score: -3.72

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 49  TRG|SV

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   8.4%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: PGPR


      KKXX-like motif in the C-terminus: MNGK

SKL: peroxisomal targeting signal in the C-terminus: none
```

```
PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 76.7

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       60.9 %: mitochondrial
       17.4 %: cytoplasmic
       17.4 %: nuclear
        4.3 %: peroxisomal

>> prediction for CG102348-01 is mit (k=23)
```

A search of the NOV3a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 3D.

Table 3D. Geneseq Results for NOV3a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV3a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU08683	Human FCTR5a polypeptide sequence - Homo sapiens, 487 aa. [WO200166747-A2, 13-SEP-2001]	1..487 1..487	485/487 (99%) 487/487 (99%)	0.0
AAU08684	Human FCTR5b polypeptide sequence - Homo sapiens, 487 aa. [WO200166747-A2, 13-SEP-2001]	1..487 1..487	484/487 (99%) 487/487 (99%)	0.0
AAB23624	Human secreted protein SEQ ID NO: 48 - Homo sapiens, 464 aa. [WO200049134-A1, 24-AUG-2000]	24..487 1..464	464/464 (100%) 464/464 (100%)	0.0
AAB95642	Human protein sequence SEQ ID NO:18384 - Homo sapiens, 438 aa. [EP1074617-A2, 07-FEB-2001]	1..425 1..425	423/425 (99%) 423/425 (99%)	0.0
ABG65427	Human albumin fusion protein #2102 - Homo sapiens, 396 aa. [WO200177137-A1, 18-OCT-2001]	24..413 1..390	389/390 (99%) 390/390 (99%)	0.0

In a BLAST search of public sequence databases, the NOV3a protein was found to have homology to the proteins shown in the BLASTP data in Table 3E.

Table 3E. Public BLASTP Results for NOV3a				
Protein Accession Number	**Protein/Organism/Length**	**NOV3a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9NZP8	Complement C1r-like proteinase - Homo sapiens (Human), 487 aa.	1..487 1..487	487/487 (100%) 487/487 (100%)	0.0
CAC88678	Sequence 16 from Patent WO0166747 - Homo sapiens (Human), 487 aa.	1..487 1..487	485/487 (99%) 487/487 (99%)	0.0
Q9H804	Hypothetical protein - Homo sapiens (Human), 438 aa.	1..425 1..425	423/425 (99%) 423/425 (99%)	0.0
AAH35220	Similar to complement component - Homo sapiens (Human), 705 aa.	189..483 400..701	176/303 (58%) 226/303 (74%)	e-108
Q8J012	Complement component 1, r subcomponent - Homo sapiens (Human), 449 aa (fragment).	189..483 144..445	176/303 (58%) 226/303 (74%)	e-108

PFam analysis indicates that the NOV3a protein contains the domains shown in the Table 3F.

Table 3F. Domain Analysis of NOV3a			
Pfam Domain	**NOV3a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
CUB	40..160	41/131 (31%) 88/131 (67%)	1.4e-06
trypsin	245..479	87/270 (32%) 174/270 (64%)	2.1e-43

Example 4.

The NOV4 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 4A.

Table 4A. NOV4 Sequence Analysis		
NOV4a, CG125860-02 DNA Sequence	SEQ ID NO: 53	1880 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 1563

TTACAACAGTGCCACTGACCCCT**ATG**AGCCTGATGCTGGATGACCAACCCCCTATGGAGGCCCAGTAT
GCAGAGGAGGGCCCAGGACCTGGGATCTTCAGAGCAGAGCCTGGAGACCAGCAGCATCCCAGTAGGCC
AGACTGGGCCATAGGGGAAATGACAGGGTGGGGACAGTGGAGGGCAATCATCCTACATTCCCCGGATC
CTCCTTGGGGTCAGCCCCACATGATTGATGTTTCTCAGGCAGTGTGCTGGCGTTCCATGCGACGTGGC
TGTGCAGTGCTGGGAGCCCTGGGGCTGCTGGCCGGTGCAGGTGTTGGCTCATGGCTCCTAGTGCTGTA
TCTGTGTCCTGCTGCCTCTCAGCCCATTTCCGGGACCTTGCAGGATGAGGAGATAACTTTGAGCTGCT
CAGAGGCCAGCGCTGAGGAAGCTCTGCTCCCTGCACTTCCCAAAACACCTGCACTTCCCAAAACAGTA
TCTTTCAGAATAAACAGCGAAGACTTCTTGCTGGAAGCGCAAGTGAGGGATCAGCCACGCTGGCTCCT
GGTCTGCCATGAGGGCTGGAGCCCCGCCCTGGGGCTGCAGATCTGCTGGAGCCTTGGGCATCTCAGAC
TCACTCACCACAAGGGAGTAAACCTCACTGACATCAAACTCAACAGTTCCCAGGAGTTTGCTCAGCTC
TCTCCTAGACTGGGAGGCTTCCTGGAGGAGGCGTGGCAGCCCAGTAGGACTACTGAGGCTGTTAGGAA
CAACTGCACTTCTGGTCAAGTTGTTTCCCTCAGATGCTCTGAGTGTGGAGCGAGGCCCCTGGCTTCCC
GGATAGTTGGTGGGCAGTCTGTGGCTCCTGGGCGCTGGCCGTGGCAGGCCAGCGTGGCCCTGGGCTTC
CGGCACACGTGTGGGGGCTCTGTGCTAGCGCCACGCTGGGTGGTGACTGCTGCACATTGTATGCACAG
TTTCAGGCTGGCCCGCCTGTCCAGCTGGCGGGTTCATGCGGGGCTGGTCAGCCACAGTGCCGTCAGGC
CCCACCAAGGGGCTCTGGTGGAGAGGATTATCCCACACCCCCTCTACAGTGCCCAGAATCATGACTAC
GACGTCGCCCTCCTGAGGCTCCAGACCGCTCTCAACTTCTCAGACACTGTGGGCGCTGTGTGCCTGCC
GGCCAAGGAACAGCATTTTCCGAAGGGCTCGCGGTGCTGGGTGTCTGGCTGGGGCCACACCCACCCTA
GCCATACTTACAGCTCGGATATGCTCCAGGACACGGTGGTGCCCCTGCTCAGCACTCAGCTCTGCAAC
AGCTCTTGCGTGTACAGCGGAGCCCTCACCCCCCGCATGCTTTGCGCTGGCTACCTGGACGGAAGGGC
TGATGCATGCCAGGGAGATAGCGGGGGCCCCCTAGTGTGCCCAGATGGGGACACATGGCGCCTAGTGG
GGGTGGTCAGCTGGGGGCGTGGCTGCGCAGAGCCCAATCACCCAGGTGTCTACGCCAAGGTAGCTGAG
TTTCTGGACTGGATCCATGACACTGCTCAGGTGAGTGTGGGGGCAGGAGTAGGGCAGGGAGATTTC**TA**
AAGGACCTGCCCTCGAATGCAAGGAACCTTACCCCTTAGGCCCGGGCCCTGCTGGGGACTGGGGAGGG
TGCTAGGACATATTCCCCAGAGTGAGTGGAGGAAGAAGTGAAGCTTAAACATGGAATCCATTGGATTT
CTATCAGTTTAAGGATGAACTGGGTAAGAGTATGCCTGAGTTTGTATCCCAGATCTACCATTTCCTGT
GTCGACCTTTGGCAAATTTCTAACTTTGTTAAACCTTAATTTCCTGATAATAACCATGATGGCTACTT
ATATGCTATTGTTATATGCTATTAAATAAGACCCGTACAATGCC

NOV4a, CG125860-02 Protein Sequence	SEQ ID NO: 54	513 aa	MW at 55531.5kD

MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPSRPDWAIGEMTGWGQWRAIILHSPDPPWGQPHM
IDVSQAVCWRSMRRGCAVLGALGLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEA
LLPALPKTPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVN
LTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAVRNNCTSGQVVSLRCSECGARPLASRIVGGQSV
APGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVE
RIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDM
LQDTVVPLLSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRG
CAEPNHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

NOV4b, CG125860-01 DNA Sequence	SEQ ID NO: 55	1787 bp
	ORF Start: ATG at 54	ORF Stop: TAA at 1470

GCGGAACATTGCCTAGTAGACCCTGAGGCTTTACAACAGTGCCACTGACCCCT**ATG**AGCCTGATGCTG
GATGACCAACCCCCTATGGAGGCCCAGTATGCAGAGGAGGGCCCAGGACCTGGGATCTTCAGAGCAGA
GCCTGGAGACCAGCAGCATCCCATTTCTCAGGCGGTGTGCTGGCGTTCCATGCGACGTGGCTGTGCAG
TGCTGGGAGCCCTGGGGCTGCTGGCCGGTGCAGGTGTTGGCTCATGGCTCCTAGTGCTGTATCTGTGT
CCTGCTGCCTCTCAGCCCATTTCCGGGACCTTGCAGGATGAGGAGATAACTTTGAGCTGCTCAGAGGC
CAGCGCTGAGGAAGCTCTGCTCCCTGCACTTCCCAAAACAGTATCTTTCAGAATAAACAGCGAAGACT
TCTTGCTGGAAGCGCAAGTGAGGGATCAGCCACGCTGGCTCCTGGTCTGCCATGAGGGCTGGAGCCCC
GCCCTGGGGCTGCAGATCTGCTGGAGCCTTGGGCATCTCAGACTCACTCACCACAAGGGAGTAAACCT
CACTGACATCAAACTCAACAGTTCCCAGGAGTTTGCTCAGCTCTCTCCTAGACTGGGAGGCTTCCTGG
AGGAGGCGTGGCAGCCCAGTAGGACTACTGAGGCTGTTAGGAACAACTGCACTTCTGGTCAAGTTGTT
TCCCTCAGATGCTCTGAGTGTGGAGCGAGGCCCCTGGCTTCCCGGATAGTTGGTGGGCAGTCTGTGGC

TCCTGGGCGCTGGCCGTGGCAGGCCAGCGTGGCCCTGGGCTTCCGGCACACGTGTGGGGGCTCTGTGC
TAGCGCCACGCTGGGTGGTGACTGCTGCACATTGTATGCACAGTTTCAGGCTGGCCCGCCTGTCCAGC
TGGCGGGTTCATGCGGGGCTGGTCAGCCACAGTGCCGTCAGGCCCCACCAAGGGGCTCTGGTGGAGAG
GATTATCCCACACCCCCTCTACAGTGCCCAGAATCATGACTACGACGTCGCCCTCCTGAGGCTCCAGA
CCGCTCTCAACTTCTCAGACACTGTGGGCGCTGTGTGCCTGCCGGCCAAGGAACAGCATTTTCCGAAG
GGCTCGCGGTGCTGGGTGTCTGGCTGGGGCCACACCCACCCTAGCCATACTTACAGCTCGGATATGCT
CCAGGACACGGTGGTGCCCCTGCTCAGCACTCAGCTCTGCAACAGCTCTTGCGTGTACAGCGGAGCCC
TCACCCCCCGCATGCTTTGCGCTGGCTACCTGGACGGAAGGGCTGATGCATGCCAGGGAGATAGCGGG
GGCCCCCTAGTGTGCCCAGATGGGGACACATGGCGCCTAGTGGGGGTGGTCAGCTGGGGGCGTGGCTG
CGCAGAGCCCAATCACCCAGGTGTCTACGCCAAGGTAGCTGAGTTTCTGGACTGGATCCATGACACTG
CTCAGGTGAGTGTGGGGGCAGGAGTAGGGCAGGGAGATTTC**TAA**AGGACCTGCCCTCGAATGCAAGGA
ACCTTACCCCTTAGGCCCGGGCCCTGCTGGGGACTGGGGAGGGTGCTAGGACATATTCCCCAGAGTGA
GTGGAGGAAGAAGTGAAGCTTAAACATGGAATCCATTGGATTTCTATCAGTTTAAGGATGAACTGGGT
AAGAGTATGCCTGAGTTTGTATCCCAGATCTACCATTTCCTGTGTCGACCTTTGGCAAATTTCTAACT
TTGTTAAACCTTAATTTCCTGATAATAACCATGATGGCTACTTATATGCTATTGTTATATGCTATTAA
ATAAGACCCGTACAATGCC

NOV4b, CG125860-01 Protein Sequence	SEQ ID NO: 56	472 aa	MW at 50916.3kD

MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPISQAVCWRSMRRGCAVLGALGLLAGAGVGSWLL
VLYLCPAASQPISGTLQDEEITLSCSEASAEEALLPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCH
EGWSPALGLQICWSLGHLRLTHHKGVNLTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAVRNNCT
SGQVVSLRCSECGARPLASRIVGGQSVAPGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRL
ARLSSWRVHAGLVSHSAVRPHQGALVERIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKE
QHFPKGSRCWVSGWGHTHPSHTYSSDMLQDTVVPLLSTQLCNSSCVYSGALTPRMLCAGYLDGRADAC
QGDSGGPLVCPDGDTWRLVGVVSWGRGCAEPNHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

NOV4c, SNP13376012 of CG125860-02, DNA Sequence	SEQ ID NO: 57	1880 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 1563
	SNP Pos: 36	SNP Change: C to T

TTACAACAGTGCCACTGACCCCT**ATG**AGCCTGATG**T**TGGATGACCAACCCCCTATGGAGGCCCAGTAT
GCAGAGGAGGGCCCAGGACCTGGGATCTTCAGAGCAGAGCCTGGAGACCAGCAGCATCCCAGTAGGCC
AGACTGGGCCATAGGGGAAATGACAGGGTGGGGACAGTGGAGGGCAATCATCCTACATTCCCCGGATC
CTCCTTGGGGTCAGCCCCACATGATTGATGTTTCTCAGGCAGTGTGCTGGCGTTCCATGCGACGTGGC
TGTGCAGTGCTGGGAGCCCTGGGGCTGCTGGCCGGTGCAGGTGTTGGCTCATGGCTCCTAGTGCTGTA
TCTGTGTCCTGCTGCCTCTCAGCCCATTTCCGGGACCTTGCAGGATGAGGAGATAACTTTGAGCTGCT
CAGAGGCCAGCGCTGAGGAAGCTCTGCTCCCTGCACTTCCCAAAACACCTGCACTTCCCAAAACAGTA
TCTTTCAGAATAAACAGCGAAGACTTCTTGCTGGAAGCGCAAGTGAGGGATCAGCCACGCTGGCTCCT
GGTCTGCCATGAGGGCTGGAGCCCCGCCCTGGGGCTGCAGATCTGCTGGAGCCTTGGGCATCTCAGAC
TCACTCACCACAAGGGAGTAAACCTCACTGACATCAAACTCAACAGTTCCCAGGAGTTTGCTCAGCTC
TCTCCTAGACTGGGAGGCTTCCTGGAGGAGGCGTGGCAGCCCAGTAGGACTACTGAGGCTGTTAGGAA
CAACTGCACTTCTGGTCAAGTTGTTTCCCTCAGATGCTCTGAGTGTGGAGCGAGGCCCCTGGCTTCCC
GGATAGTTGGTGGGCAGTCTGTGGCTCCTGGGCGCTGGCCGTGGCAGGCCAGCGTGGCCCTGGGCTTC
CGGCACACGTGTGGGGGCTCTGTGCTAGCGCCACGCTGGGTGGTGACTGCTGCACATTGTATGCACAG
TTTCAGGCTGGCCCGCCTGTCCAGCTGGCGGGTTCATGCGGGGCTGGTCAGCCACAGTGCCGTCAGGC
CCCACCAAGGGGCTCTGGTGGAGAGGATTATCCCACACCCCCTCTACAGTGCCCAGAATCATGACTAC
GACGTCGCCCTCCTGAGGCTCCAGACCGCTCTCAACTTCTCAGACACTGTGGGCGCTGTGTGCCTGCC
GGCCAAGGAACAGCATTTTCCGAAGGGCTCGCGGTGCTGGGTGTCTGGCTGGGGCCACACCCACCCTA
GCCATACTTACAGCTCGGATATGCTCCAGGACACGGTGGTGCCCCTGCTCAGCACTCAGCTCTGCAAC
AGCTCTTGCGTGTACAGCGGAGCCCTCACCCCCCGCATGCTTTGCGCTGGCTACCTGGACGGAAGGGC
TGATGCATGCCAGGGAGATAGCGGGGGCCCCCTAGTGTGCCCAGATGGGGACACATGGCGCCTAGTGG
GGGTGGTCAGCTGGGGGCGTGGCTGCGCAGAGCCCAATCACCCAGGTGTCTACGCCAAGGTAGCTGAG
TTTCTGGACTGGATCCATGACACTGCTCAGGTGAGTGTGGGGGCAGGAGTAGGGCAGGGAGATTTC**TA**
AAGGACCTGCCCTCGAATGCAAGGAACCTTACCCCTTAGGCCCGGGCCCTGCTGGGGACTGGGGAGGG
TGCTAGGACATATTCCCCAGAGTGAGTGGAGGAAGAAGTGAAGCTTAAACATGGAATCCATTGGATTT
CTATCAGTTTAAGGATGAACTGGGTAAGAGTATGCCTGAGTTTGTATCCCAGATCTACCATTTCCTGT
GTCGACCTTTGGCAAATTTCTAACTTTGTTAAACCTTAATTTCCTGATAATAACCATGATGGCTACTT
ATATGCTATTGTTATATGCTATTAAATAAGACCCGTACAATGCC

NOV4c, SNP13376012 of CG125860-02, Protein Sequence	SEQ ID NO: 58	513 aa	MW at 55531.5kD
	SNP Pos: 5		SNP Change: Leu to Leu

MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPSRPDWAIGEMTGWGQWRAIILHSPDPPWGQPHM
IDVSQAVCWRSMRRGCAVLGALGLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEA
LLPALPKTPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVN
LTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAVRNNCTSGQVVSLRCSECGARPLASRIVGGQSV
APGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVE
RIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDM
LQDTVVPLLSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRG
CAEPNHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

NOV4d, SNP13376014 of CG125860-02, DNA Sequence	SEQ ID NO: 59	1880 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 1563
	SNP Pos: 265	SNP Change: G to A

TTACAACAGTGCCACTGACCCCTATGAGCCTGATGCTGGATGACCAACCCCCTATGGAGGCCCAGTAT
GCAGAGGAGGGCCCAGGACCTGGGATCTTCAGAGCAGAGCCTGGAGACCAGCAGCATCCCAGTAGGCC
AGACTGGGCCATAGGGGAAATGACAGGGTGGGGACAGTGGAGGGCAATCATCCTACATTCCCCGGATC
CTCCTTGGGGTCAGCCCCACATGATTGATGTTTCTCAGGCAGTGTGCTGGCGTTCCATGCAACGTGGC
TGTGCAGTGCTGGGAGCCCTGGGGCTGCTGGCCGGTGCAGGTGTTGGCTCATGGCTCCTAGTGCTGTA
TCTGTGTCCTGCTGCCTCTCAGCCCATTTCCGGGACCTTGCAGGATGAGGAGATAACTTTGAGCTGCT
CAGAGGCCAGCGCTGAGGAAGCTCTGCTCCCTGCACTTCCCAAAACACCTGCACTTCCCAAAACAGTA
TCTTTCAGAATAAACAGCGAAGACTTCTTGCTGGAAGCGCAAGTGAGGGATCAGCCACGCTGGCTCCT
GGTCTGCCATGAGGGCTGGAGCCCCGCCCTGGGGCTGCAGATCTGCTGGAGCCTTGGGCATCTCAGAC
TCACTCACCACAAGGGAGTAAACCTCACTGACATCAAACTCAACAGTTCCCAGGAGTTTGCTCAGCTC
TCTCCTAGACTGGGAGGCTTCCTGGAGGAGGCGTGGCAGCCCAGTAGGACTACTGAGGCTGTTAGGAA
CAACTGCACTTCTGGTCAAGTTGTTTCCCTCAGATGCTCTGAGTGTGGAGCGAGGCCCCTGGCTTCCC
GGATAGTTGGTGGGCAGTCTGTGGCTCCTGGGCGCTGGCCGTGGCAGGCCAGCGTGGCCCTGGGCTTC
CGGCACACGTGTGGGGGCTCTGTGCTAGCGCCACGCTGGGTGGTGACTGCTGCACATTGTATGCACAG
TTTCAGGCTGGCCCGCCTGTCCAGCTGGCGGGTTCATGCGGGGCTGGTCAGCCACAGTGCCGTCAGGC
CCCACCAAGGGGCTCTGGTGGAGAGGATTATCCCACACCCCCTCTACAGTGCCCAGAATCATGACTAC
GACGTCGCCCTCCTGAGGCTCCAGACCGCTCTCAACTTCTCAGACACTGTGGGCGCTGTGTGCCTGCC
GGCCAAGGAACAGCATTTTCCGAAGGGCTCGCGGTGCTGGGTGTCTGGCTGGGGCCACACCCACCCTA
GCCATACTTACAGCTCGGATATGCTCCAGGACACGGTGGTGCCCCTGCTCAGCACTCAGCTCTGCAAC
AGCTCTTGCGTGTACAGCGGAGCCCTCACCCCCCGCATGCTTTGCGCTGGCTACCTGGACGGAAGGGC
TGATGCATGCCAGGGAGATAGCGGGGGCCCCCTAGTGTGCCCAGATGGGGACACATGGCGCCTAGTGG
GGGTGGTCAGCTGGGGGCGTGGCTGCGCAGAGCCCAATCACCCAGGTGTCTACGCCAAGGTAGCTGAG
TTTCTGGACTGGATCCATGACACTGCTCAGGTGAGTGTGGGGGCAGGAGTAGGGCAGGGAGATTTCTA
AAGGACCTGCCCTCGAATGCAAGGAACCTTACCCCTTAGGCCCGGGCCCTGCTGGGGACTGGGGAGGG
TGCTAGGACATATTCCCCAGAGTGAGTGGAGGAAGAAGTGAAGCTTAAACATGGAATCCATTGGATTT
CTATCAGTTTAAGGATGAACTGGGTAAGAGTATGCCTGAGTTTGTATCCCAGATCTACCATTTCCTGT
GTCGACCTTTGGCAAATTTCTAACTTTGTTAAACCTTAATTTCCTGATAATAACCATGATGGCTACTT
ATATGCTATTGTTATATGCTATTAAATAAGACCCGTACAATGCC

NOV4d, SNP13376014 of CG125860-02, Protein Sequence	SEQ ID NO: 60	513 aa	MW at 55503.5kD
	SNP Pos: 81		SNP Change: Arg to Gln

MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPSRPDWAIGEMTGWGQWRAIILHSPDPPWGQPHM
IDVSQAVCWRSMQRGCAVLGALGLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEA
LLPALPKTPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVN
LTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAVRNNCTSGQVVSLRCSECGARPLASRIVGGQSV
APGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVE
RIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDM
LQDTVVPLLSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRG
CAEPNHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

NOV4e, SNP13376015 of CG125860-02, DNA Sequence	SEQ ID NO: 61	1880 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 1563
	SNP Pos: 827	SNP Change: T to C

TTACAACAGTGCCACTGACCCCTATGAGCCTGATGCTGGATGACCAACCCCCTATGGAGGCCCAGTAT
GCAGAGGAGGGCCCAGGACCTGGGATCTTCAGAGCAGAGCCTGGAGACCAGCAGCATCCCAGTAGGCC
AGACTGGGCCATAGGGGAAATGACAGGGTGGGGACAGTGGAGGGCAATCATCCTACATTCCCCGGATC
CTCCTTGGGGTCAGCCCCACATGATTGATGTTTCTCAGGCAGTGTGCTGGCGTTCCATGCGACGTGGC
TGTGCAGTGCTGGGAGCCCTGGGGCTGCTGGCCGGTGCAGGTGTTGGCTCATGGCTCCTAGTGCTGTA
TCTGTGTCCTGCTGCCTCTCAGCCCATTTCCGGGACCTTGCAGGATGAGGAGATAACTTTGAGCTGCT
CAGAGGCCAGCGCTGAGGAAGCTCTGCTCCCTGCACTTCCCAAAACACCTGCACTTCCCAAAACAGTA
TCTTTCAGAATAAACAGCGAAGACTTCTTGCTGGAAGCGCAAGTGAGGGATCAGCCACGCTGGCTCCT
GGTCTGCCATGAGGGCTGGAGCCCCGCCCTGGGGCTGCAGATCTGCTGGAGCCTTGGGCATCTCAGAC
TCACTCACCACAAGGGAGTAAACCTCACTGACATCAAACTCAACAGTTCCCAGGAGTTTGCTCAGCTC
TCTCCTAGACTGGGAGGCTTCCTGGAGGAGGCGTGGCAGCCCAGTAGGACTACTGAGGCTGTTAGGAA
CAACTGCACTTCTGGTCAAGTTGTTTCCCTCAGATGCTCTGAGTGTGGAGCGAGGCCCCTGGCTTCCC
GGATAGTTGGCGGGCAGTCTGTGGCTCCTGGGCGCTGGCCGTGGCAGGCCAGCGTGGCCCTGGGCTTC
CGGCACACGTGTGGGGGCTCTGTGCTAGCGCCACGCTGGGTGGTGACTGCTGCACATTGTATGCACAG
TTTCAGGCTGGCCCGCCTGTCCAGCTGGCGGGTTCATGCGGGGCTGGTCAGCCACAGTGCCGTCAGGC
CCCACCAAGGGGCTCTGGTGGAGAGGATTATCCCACACCCCCTCTACAGTGCCCAGAATCATGACTAC
GACGTCGCCCTCCTGAGGCTCCAGACCGCTCTCAACTTCTCAGACACTGTGGGCGCTGTGTGCCTGCC
GGCCAAGGAACAGCATTTTCCGAAGGGCTCGCGGTGCTGGGTGTCTGGCTGGGGCCACACCCACCCTA
GCCATACTTACAGCTCGGATATGCTCCAGGACACGGTGGTGCCCCTGCTCAGCACTCAGCTCTGCAAC
AGCTCTTGCGTGTACAGCGGAGCCCTCACCCCCCGCATGCTTTGCGCTGGCTACCTGGACGGAAGGGC
TGATGCATGCCAGGGAGATAGCGGGGGCCCCCTAGTGTGCCCAGATGGGGACACATGGCGCCTAGTGG
GGGTGGTCAGCTGGGGGCGTGGCTGCGCAGAGCCCAATCACCCAGGTGTCTACGCCAAGGTAGCTGAG
TTTCTGGACTGGATCCATGACACTGCTCAGGTGAGTGTGGGGGCAGGAGTAGGGCAGGGAGATTTCTA
AAGGACCTGCCCTCGAATGCAAGGAACCTTACCCCTTAGGCCCGGGCCCTGCTGGGGACTGGGGAGGG
TGCTAGGACATATTCCCCAGAGTGAGTGGAGGAAGAAGTGAAGCTTAAACATGGAATCCATTGGATTT
CTATCAGTTTAAGGATGAACTGGGTAAGAGTATGCCTGAGTTTGTATCCCAGATCTACCATTTCCTGT
GTCGACCTTTGGCAAATTTCTAACTTTGTTAAACCTTAATTTCCTGATAATAACCATGATGGCTACTT
ATATGCTATTGTTATATGCTATTAAATAAGACCCGTACAATGCC

NOV4e, SNP13376015 of CG125860-02, Protein Sequence	SEQ ID NO: 62	513 aa	MW at 55531.5kD
	SNP Pos: 268		SNP Change: Gly to Gly

MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPSRPDWAIGEMTGWGQWRAIILHSPDPPWGQPHM
IDVSQAVCWRSMRRGCAVLGALGLLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEA
LLPALPKTPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVN
LTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAVRNNCTSGQVVSLRCSECGARPLASRIVGGQSV
APGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVE
RIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDM
LQDTVVPLLSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRG
CAEPNHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

NOV4f, SNP13376016 of CG125860-02, DNA Sequence	SEQ ID NO: 63	1880 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 1563
	SNP Pos: 830	SNP Change: G to T

TTACAACAGTGCCACTGACCCCTATGAGCCTGATGCTGGATGACCAACCCCCTATGGAGGCCCAGTAT
GCAGAGGAGGGCCCAGGACCTGGGATCTTCAGAGCAGAGCCTGGAGACCAGCAGCATCCCAGTAGGCC
AGACTGGGCCATAGGGGAAATGACAGGGTGGGGACAGTGGAGGGCAATCATCCTACATTCCCCGGATC
CTCCTTGGGGTCAGCCCCACATGATTGATGTTTCTCAGGCAGTGTGCTGGCGTTCCATGCGACGTGGC
TGTGCAGTGCTGGGAGCCCTGGGGCTGCTGGCCGGTGCAGGTGTTGGCTCATGGCTCCTAGTGCTGTA
TCTGTGTCCTGCTGCCTCTCAGCCCATTTCCGGGACCTTGCAGGATGAGGAGATAACTTTGAGCTGCT
CAGAGGCCAGCGCTGAGGAAGCTCTGCTCCCTGCACTTCCCAAAACACCTGCACTTCCCAAAACAGTA
TCTTTCAGAATAAACAGCGAAGACTTCTTGCTGGAAGCGCAAGTGAGGGATCAGCCACGCTGGCTCCT
GGTCTGCCATGAGGGCTGGAGCCCCGCCCTGGGGCTGCAGATCTGCTGGAGCCTTGGGCATCTCAGAC

TCACTCACCACAAGGGAGTAAACCTCACTGACATCAAACTCAACAGTTCCCAGGAGTTTGCTCAGCTC
TCTCCTAGACTGGGAGGCTTCCTGGAGGAGGCGTGGCAGCCCAGTAGGACTACTGAGGCTGTTAGGAA
CAACTGCACTTCTGGTCAAGTTGTTTCCCTCAGATGCTCTGAGTGTGGAGCGAGGCCCCTGGCTTCCC
GGATAGTTGGTGGTCAGTCTGTGGCTCCTGGGCGCTGGCCGTGGCAGGCCAGCGTGGCCCTGGGCTTC
CGGCACACGTGTGGGGGCTCTGTGCTAGCGCCACGCTGGGTGGTGACTGCTGCACATTGTATGCACAG
TTTCAGGCTGGCCCGCCTGTCCAGCTGGCGGGTTCATGCGGGGCTGGTCAGCCACAGTGCCGTCAGGC
CCCACCAAGGGGCTCTGGTGGAGAGGATTATCCCACACCCCCTCTACAGTGCCCAGAATCATGACTAC
GACGTCGCCCTCCTGAGGCTCCAGACCGCTCTCAACTTCTCAGACACTGTGGGCGCTGTGTGCCTGCC
GGCCAAGGAACAGCATTTTCCGAAGGGCTCGCGGTGCTGGGTGTCTGGCTGGGGCCACACCCACCCTA
GCCATACTTACAGCTCGGATATGCTCCAGGACACGGTGGTGCCCCTGCTCAGCACTCAGCTCTGCAAC
AGCTCTTGCGTGTACAGCGGAGCCCTCACCCCCCGCATGCTTTGCGCTGGCTACCTGGACGGAAGGGC
TGATGCATGCCAGGGAGATAGCGGGGGCCCCCTAGTGTGCCCAGATGGGGACACATGGCGCCTAGTGG
GGGTGGTCAGCTGGGGGCGTGGCTGCGCAGAGCCCAATCACCCAGGTGTCTACGCCAAGGTAGCTGAG
TTTCTGGACTGGATCCATGACACTGCTCAGGTGAGTGTGGGGGCAGGAGTAGGGCAGGGAGATTTC**TA**
AAGGACCTGCCCTCGAATGCAAGGAACCTTACCCCTTAGGCCCGGGCCCTGCTGGGGACTGGGGAGGG
TGCTAGGACATATTCCCCAGAGTGAGTGGAGGAAGAAGTGAAGCTTAAACATGGAATCCATTGGATTT
CTATCAGTTTAAGGATGAACTGGGTAAGAGTATGCCTGAGTTTGTATCCCAGATCTACCATTTCCTGT
GTCGACCTTTGGCAAATTTCTAACTTTGTTAAACCTTAATTTCCTGATAATAACCATGATGGCTACTT
ATATGCTATTGTTATATGCTATTAAATAAGACCCGTACAATGCC

NOV4f, SNP13376016 of CG125860-02, Protein Sequence	SEQ ID NO: 64	513 aa	MW at 55531.5kD
	SNP Pos: 269		SNP Change: Gly to Gly

MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPSRPDWAIGEMTGWGQWRAIILHSPDPPWGQPHM
IDVSQAVCWRSMRRGCAVLGALGLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEA
LLPALPKTPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVN
LTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAVRNNCTSGQVVSLRCSECGARPLASRIVGGQSV
APGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVE
RIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDM
LQDTVVPLLSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRG
CAEPNHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

NOV4g, SNP13376017 of CG125860-02, DNA Sequence	SEQ ID NO: 65	1880 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 1563
	SNP Pos: 836	SNP Change: T to C

TTACAACAGTGCCACTGACCCCT**ATG**AGCCTGATGCTGGATGACCAACCCCCTATGGAGGCCCAGTAT
GCAGAGGAGGGCCCAGGACCTGGGATCTTCAGAGCAGAGCCTGGAGACCAGCAGCATCCCAGTAGGCC
AGACTGGGCCATAGGGGAAATGACAGGGTGGGGACAGTGGAGGGCAATCATCCTACATTCCCCGGATC
CTCCTTGGGGTCAGCCCCACATGATTGATGTTTCTCAGGCAGTGTGCTGGCGTTCCATGCGACGTGGC
TGTGCAGTGCTGGGAGCCCTGGGGCTGCTGGCCGGTGCAGGTGTTGGCTCATGGCTCCTAGTGCTGTA
TCTGTGTCCTGCTGCCTCTCAGCCCATTTCCGGGACCTTGCAGGATGAGGAGATAACTTTGAGCTGCT
CAGAGGCCAGCGCTGAGGAAGCTCTGCTCCCTGCACTTCCCAAAACACCTGCACTTCCCAAAACAGTA
TCTTTCAGAATAAACAGCGAAGACTTCTTGCTGGAAGCGCAAGTGAGGGATCAGCCACGCTGGCTCCT
GGTCTGCCATGAGGGCTGGAGCCCCGCCCTGGGGCTGCAGATCTGCTGGAGCCTTGGGCATCTCAGAC
TCACTCACCACAAGGGAGTAAACCTCACTGACATCAAACTCAACAGTTCCCAGGAGTTTGCTCAGCTC
TCTCCTAGACTGGGAGGCTTCCTGGAGGAGGCGTGGCAGCCCAGTAGGACTACTGAGGCTGTTAGGAA
CAACTGCACTTCTGGTCAAGTTGTTTCCCTCAGATGCTCTGAGTGTGGAGCGAGGCCCCTGGCTTCCC
GGATAGTTGGTGGGCAGTCCGTGGCTCCTGGGCGCTGGCCGTGGCAGGCCAGCGTGGCCCTGGGCTTC
CGGCACACGTGTGGGGGCTCTGTGCTAGCGCCACGCTGGGTGGTGACTGCTGCACATTGTATGCACAG
TTTCAGGCTGGCCCGCCTGTCCAGCTGGCGGGTTCATGCGGGGCTGGTCAGCCACAGTGCCGTCAGGC
CCCACCAAGGGGCTCTGGTGGAGAGGATTATCCCACACCCCCTCTACAGTGCCCAGAATCATGACTAC
GACGTCGCCCTCCTGAGGCTCCAGACCGCTCTCAACTTCTCAGACACTGTGGGCGCTGTGTGCCTGCC
GGCCAAGGAACAGCATTTTCCGAAGGGCTCGCGGTGCTGGGTGTCTGGCTGGGGCCACACCCACCCTA
GCCATACTTACAGCTCGGATATGCTCCAGGACACGGTGGTGCCCCTGCTCAGCACTCAGCTCTGCAAC
AGCTCTTGCGTGTACAGCGGAGCCCTCACCCCCCGCATGCTTTGCGCTGGCTACCTGGACGGAAGGGC
TGATGCATGCCAGGGAGATAGCGGGGGCCCCCTAGTGTGCCCAGATGGGGACACATGGCGCCTAGTGG
GGGTGGTCAGCTGGGGGCGTGGCTGCGCAGAGCCCAATCACCCAGGTGTCTACGCCAAGGTAGCTGAG
TTTCTGGACTGGATCCATGACACTGCTCAGGTGAGTGTGGGGGCAGGAGTAGGGCAGGGAGATTTC**TA**
AAGGACCTGCCCTCGAATGCAAGGAACCTTACCCCTTAGGCCCGGGCCCTGCTGGGGACTGGGGAGGG

TGCTAGGACATATTCCCCAGAGTGAGTGGAGGAAGAAGTGAAGCTTAAACATGGAATCCATTGGATTT
CTATCAGTTTAAGGATGAACTGGGTAAGAGTATGCCTGAGTTTGTATCCCAGATCTACCATTTCCTGT
GTCGACCTTTGGCAAATTTCTAACTTTGTTAAACCTTAATTTCCTGATAATAACCATGATGGCTACTT
ATATGCTATTGTTATATGCTATTAAATAAGACCCGTACAATGCC

NOV4g, SNP13376017 of CG125860-02, Protein Sequence	SEQ ID NO: 66	513 aa	MW at 55531.5kD
	SNP Pos: 271		SNP Change: Ser to Ser

MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPSRPDWAIGEMTGWGQWRAIILHSPDPPWGQPHM
IDVSQAVCWRSMRRGCAVLGALGLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEA
LLPALPKTPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVN
LTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAVRNNCTSGQVVSLRCSECGARPLASRIVGGQSV
APGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVE
RIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDM
LQDTVVPLLSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRG
CAEPNHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

NOV4h, SNP13376011 of CG125860-02, DNA Sequence	SEQ ID NO: 67	1880 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 1563
	SNP Pos: 869	SNP Change: C to T

TTACAACAGTGCCACTGACCCCTATGAGCCTGATGCTGGATGACCAACCCCCTATGGAGGCCCAGTAT
GCAGAGGAGGGCCCAGGACCTGGGATCTTCAGAGCAGAGCCTGGAGACCAGCAGCATCCCAGTAGGCC
AGACTGGGCCATAGGGGAAATGACAGGGTGGGGACAGTGGAGGGCAATCATCCTACATTCCCCGGATC
CTCCTTGGGGTCAGCCCCACATGATTGATGTTTCTCAGGCAGTGTGCTGGCGTTCCATGCGACGTGGC
TGTGCAGTGCTGGGAGCCCTGGGGCTGCTGGCCGGTGCAGGTGTTGGCTCATGGCTCCTAGTGCTGTA
TCTGTGTCCTGCTGCCTCTCAGCCCATTTCCGGGACCTTGCAGGATGAGGAGATAACTTTGAGCTGCT
CAGAGGCCAGCGCTGAGGAAGCTCTGCTCCCTGCACTTCCCAAAACACCTGCACTTCCCAAAACAGTA
TCTTTCAGAATAAACAGCGAAGACTTCTTGCTGGAAGCGCAAGTGAGGGATCAGCCACGCTGGCTCCT
GGTCTGCCATGAGGGCTGGAGCCCCGCCCTGGGGCTGCAGATCTGCTGGAGCCTTGGGCATCTCAGAC
TCACTCACCACAAGGGAGTAAACCTCACTGACATCAAACTCAACAGTTCCCAGGAGTTTGCTCAGCTC
TCTCCTAGACTGGGAGGCTTCCTGGAGGAGGCGTGGCAGCCCAGTAGGACTACTGAGGCTGTTAGGAA
CAACTGCACTTCTGGTCAAGTTGTTTCCCTCAGATGCTCTGAGTGTGGAGCGAGGCCCCTGGCTTCCC
GGATAGTTGGTGGGCAGTCTGTGGCTCCTGGGCGCTGGCCGTGGCAGGCCAGTGTGGCCCTGGGCTTC
CGGCACACGTGTGGGGGCTCTGTGCTAGCGCCACGCTGGGTGGTGACTGCTGCACATTGTATGCACAG
TTTCAGGCTGGCCCGCCTGTCCAGCTGGCGGGTTCATGCGGGGCTGGTCAGCCACAGTGCCGTCAGGC
CCCACCAAGGGGCTCTGGTGGAGAGGATTATCCCACACCCCCTCTACAGTGCCCAGAATCATGACTAC
GACGTCGCCCTCCTGAGGCTCCAGACCGCTCTCAACTTCTCAGACACTGTGGGCGCTGTGTGCCTGCC
GGCCAAGGAACAGCATTTTCCGAAGGGCTCGCGGTGCTGGGTGTCTGGCTGGGGCCACACCCACCCTA
GCCATACTTACAGCTCGGATATGCTCCAGGACACGGTGGTGCCCCTGCTCAGCACTCAGCTCTGCAAC
AGCTCTTGCGTGTACAGCGGAGCCCTCACCCCCCGCATGCTTTGCGCTGGCTACCTGGACGGAAGGGC
TGATGCATGCCAGGGAGATAGCGGGGGCCCCCTAGTGTGCCCAGATGGGGACACATGGCGCCTAGTGG
GGGTGGTCAGCTGGGGGCGTGGCTGCGCAGAGCCCAATCACCCAGGTGTCTACGCCAAGGTAGCTGAG
TTTCTGGACTGGATCCATGACACTGCTCAGGTGAGTGTGGGGGCAGGAGTAGGGCAGGGAGATTTCTA
AAGGACCTGCCCTCGAATGCAAGGAACCTTACCCCTTAGGCCCGGGCCCTGCTGGGGACTGGGGAGGG
TGCTAGGACATATTCCCCAGAGTGAGTGGAGGAAGAAGTGAAGCTTAAACATGGAATCCATTGGATTT
CTATCAGTTTAAGGATGAACTGGGTAAGAGTATGCCTGAGTTTGTATCCCAGATCTACCATTTCCTGT
GTCGACCTTTGGCAAATTTCTAACTTTGTTAAACCTTAATTTCCTGATAATAACCATGATGGCTACTT
ATATGCTATTGTTATATGCTATTAAATAAGACCCGTACAATGCC

NOV4h, SNP13376011 of CG125860-02, Protein Sequence	SEQ ID NO: 68	513 aa	MW at 55531.5kD
	SNP Pos: 282		SNP Change: Ser to Ser

MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPSRPDWAIGEMTGWGQWRAIILHSPDPPWGQPHM
IDVSQAVCWRSMRRGCAVLGALGLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEA
LLPALPKTPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVN
LTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAVRNNCTSGQVVSLRCSECGARPLASRIVGGQSV
APGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVE
RIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDM
LQDTVVPLLSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRG
CAEPNHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

NOV4i, SNP13376018 of CG125860-02, DNA Sequence	SEQ ID NO: 69	1880 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 1563
	SNP Pos: 975	SNP Change: A to G

TTACAACAGTGCCACTGACCCCTATGAGCCTGATGCTGGATGACCAACCCCCTATGGAGGCCCAGTAT
GCAGAGGAGGGCCCAGGACCTGGGATCTTCAGAGCAGAGCCTGGAGACCAGCAGCATCCCAGTAGGCC
AGACTGGGCCATAGGGGAAATGACAGGGTGGGGACAGTGGAGGGCAATCATCCTACATTCCCCGGATC
CTCCTTGGGGTCAGCCCCACATGATTGATGTTTCTCAGGCAGTGTGCTGGCGTTCCATGCGACGTGGC
TGTGCAGTGCTGGGAGCCCTGGGGCTGCTGGCCGGTGCAGGTGTTGGCTCATGGCTCCTAGTGCTGTA
TCTGTGTCCTGCTGCCTCTCAGCCCATTTCCGGGACCTTGCAGGATGAGGAGATAACTTTGAGCTGCT
CAGAGGCCAGCGCTGAGGAAGCTCTGCTCCCTGCACTTCCCAAAACACCTGCACTTCCCAAAACAGTA
TCTTTCAGAATAAACAGCGAAGACTTCTTGCTGGAAGCGCAAGTGAGGGATCAGCCACGCTGGCTCCT
GGTCTGCCATGAGGGCTGGAGCCCCGCCCTGGGGCTGCAGATCTGCTGGAGCCTTGGGCATCTCAGAC
TCACTCACCACAAGGGAGTAAACCTCACTGACATCAAACTCAACAGTTCCCAGGAGTTTGCTCAGCTC
TCTCCTAGACTGGGAGGCTTCCTGGAGGAGGCGTGGCAGCCCAGTAGGACTACTGAGGCTGTTAGGAA
CAACTGCACTTCTGGTCAAGTTGTTTCCCTCAGATGCTCTGAGTGTGGAGCGAGGCCCCTGGCTTCCC
GGATAGTTGGTGGGCAGTCTGTGGCTCCTGGGCGCTGGCCGTGGCAGGCCAGCGTGGCCCTGGGCTTC
CGGCACACGTGTGGGGGCTCTGTGCTAGCGCCACGCTGGGTGGTGACTGCTGCACATTGTATGCACAG
TTTCAGGCTGGCCCGCCTGTCCGGCTGGCGGGTTCATGCGGGGCTGGTCAGCCACAGTGCCGTCAGGC
CCCACCAAGGGGCTCTGGTGGAGAGGATTATCCCACACCCCCTCTACAGTGCCCAGAATCATGACTAC
GACGTCGCCCTCCTGAGGCTCCAGACCGCTCTCAACTTCTCAGACACTGTGGGCGCTGTGTGCCTGCC
GGCCAAGGAACAGCATTTTCCGAAGGGCTCGCGGTGCTGGGTGTCTGGCTGGGGCCACACCCACCCTA
GCCATACTTACAGCTCGGATATGCTCCAGGACACGGTGGTGCCCCTGCTCAGCACTCAGCTCTGCAAC
AGCTCTTGCGTGTACAGCGGAGCCCTCACCCCCCGCATGCTTTGCGCTGGCTACCTGGACGGAAGGGC
TGATGCATGCCAGGGAGATAGCGGGGGCCCCCTAGTGTGCCCAGATGGGGACACATGGCGCCTAGTGG
GGGTGGTCAGCTGGGGGCGTGGCTGCGCAGAGCCCAATCACCCAGGTGTCTACGCCAAGGTAGCTGAG
TTTCTGGACTGGATCCATGACACTGCTCAGGTGAGTGTGGGGGCAGGAGTAGGGCAGGGAGATTTCTA
AAGGACCTGCCCTCGAATGCAAGGAACCTTACCCCTTAGGCCCGGGCCCTGCTGGGGACTGGGGAGGG
TGCTAGGACATATTCCCCAGAGTGAGTGGAGGAAGAAGTGAAGCTTAAACATGGAATCCATTGGATTT
CTATCAGTTTAAGGATGAACTGGGTAAGAGTATGCCTGAGTTTGTATCCCAGATCTACCATTTCCTGT
GTCGACCTTTGGCAAATTTCTAACTTTGTTAAACCTTAATTTCCTGATAATAACCATGATGGCTACTT
ATATGCTATTGTTATATGCTATTAAATAAGACCCGTACAATGCC

NOV4i, SNP13376018 of CG125860-02, Protein Sequence	SEQ ID NO: 70	513 aa	MW at 55501.5kD
	SNP Pos: 318		SNP Change: Ser to Gly

MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPSRPDWAIGEMTGWGQWRAIILHSPDPPWGQPHM
IDVSQAVCWRSMRRGCAVLGALGLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEA
LLPALPKTPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVN
LTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAVRNNCTSGQVVSLRCSECGARPLASRIVGGQSV
APGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSGWRVHAGLVSHSAVRPHQGALVE
RIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDM
LQDTVVPLLSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRG
CAEPNHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

NOV4j, SNP13382467 of CG125860-02, DNA Sequence	SEQ ID NO: 71	1880 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 1563
	SNP Pos: 1269	SNP Change: C to T

TTACAACAGTGCCACTGACCCCTATGAGCCTGATGCTGGATGACCAACCCCCTATGGAGGCCCAGTAT
GCAGAGGAGGGCCCAGGACCTGGGATCTTCAGAGCAGAGCCTGGAGACCAGCAGCATCCCAGTAGGCC
AGACTGGGCCATAGGGGAAATGACAGGGTGGGGACAGTGGAGGGCAATCATCCTACATTCCCCGGATC
CTCCTTGGGGTCAGCCCCACATGATTGATGTTTCTCAGGCAGTGTGCTGGCGTTCCATGCGACGTGGC
TGTGCAGTGCTGGGAGCCCTGGGGCTGCTGGCCGGTGCAGGTGTTGGCTCATGGCTCCTAGTGCTGTA
TCTGTGTCCTGCTGCCTCTCAGCCCATTTCCGGGACCTTGCAGGATGAGGAGATAACTTTGAGCTGCT
CAGAGGCCAGCGCTGAGGAAGCTCTGCTCCCTGCACTTCCCAAAACACCTGCACTTCCCAAAACAGTA
TCTTTCAGAATAAACAGCGAAGACTTCTTGCTGGAAGCGCAAGTGAGGGATCAGCCACGCTGGCTCCT
GGTCTGCCATGAGGGCTGGAGCCCCGCCCTGGGGCTGCAGATCTGCTGGAGCCTTGGGCATCTCAGAC
TCACTCACCACAAGGGAGTAAACCTCACTGACATCAAACTCAACAGTTCCCAGGAGTTTGCTCAGCTC
TCTCCTAGACTGGGAGGCTTCCTGGAGGAGGCGTGGCAGCCCAGTAGGACTACTGAGGCTGTTAGGAA

CAACTGCACTTCTGGTCAAGTTGTTTCCCTCAGATGCTCTGAGTGTGGAGCGAGGCCCCTGGCTTCCC
GGATAGTTGGTGGGCAGTCTGTGGCTCCTGGGCGCTGGCCGTGGCAGGCCAGCGTGGCCCTGGGCTTC
CGGCACACGTGTGGGGGCTCTGTGCTAGCGCCACGCTGGGTGGTGACTGCTGCACATTGTATGCACAG
TTTCAGGCTGGCCCGCCTGTCCAGCTGGCGGGTTCATGCGGGGCTGGTCAGCCACAGTGCCGTCAGGC
CCCACCAAGGGGCTCTGGTGGAGAGGATTATCCCACACCCCCTCTACAGTGCCCAGAATCATGACTAC
GACGTCGCCCTCCTGAGGCTCCAGACCGCTCTCAACTTCTCAGACACTGTGGGCGCTGTGTGCCTGCC
GGCCAAGGAACAGCATTTTCCGAAGGGCTCGCGGTGCTGGGTGTCTGGCTGGGGCCACACCCACCCTA
GCCATACTTACAGCTCGGATATGCTCCAGGACACGGTGGTGCCCTTGCTCAGCACTCAGCTCTGCAAC
AGCTCTTGCGTGTACAGCGGAGCCCTCACCCCCCGCATGCTTTGCGCTGGCTACCTGGACGGAAGGGC
TGATGCATGCCAGGGAGATAGCGGGGGCCCCCTAGTGTGCCCAGATGGGGACACATGGCGCCTAGTGG
GGGTGGTCAGCTGGGGGCGTGGCTGCGCAGAGCCCAATCACCCAGGTGTCTACGCCAAGGTAGCTGAG
TTTCTGGACTGGATCCATGACACTGCTCAGGTGAGTGTGGGGGCAGGAGTAGGGCAGGGAGATTTCTA
AAGGACCTGCCCTCGAATGCAAGGAACCTTACCCCTTAGGCCCGGGCCCTGCTGGGGACTGGGGAGGG
TGCTAGGACATATTCCCCAGAGTGAGTGGAGGAAGAAGTGAAGCTTAAACATGGAATCCATTGGATTT
CTATCAGTTTAAGGATGAACTGGGTAAGAGTATGCCTGAGTTTGTATCCCAGATCTACCATTTCCTGT
GTCGACCTTTGGCAAATTTCTAACTTTGTTAAACCTTAATTTCCTGATAATAACCATGATGGCTACTT
ATATGCTATTGTTATATGCTATTAAATAAGACCCGTACAATGCC

NOV4j, SNP13382467 of CG125860-02, Protein Sequence	SEQ ID NO: 72	513 aa	MW at 55531.5kD
	SNP Pos: 416		SNP Change: Leu to Leu

MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPSRPDWAIGEMTGWGQWRAIILHSPDPPWGQPHM
IDVSQAVCWRSMRRGCAVLGALGLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEA
LLPALPKTPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVN
LTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAVRNNCTSGQVVSLRCSECGARPLASRIVGGQSV
APGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVE
RIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDM
LQDTVVPLLSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRG
CAEPNHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

NOV4k, SNP13382466 of CG125860-02, DNA Sequence	SEQ ID NO: 73	1880 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 1563
	SNP Pos: 1272	SNP Change: C to T

TTACAACAGTGCCACTGACCCCTATGAGCCTGATGCTGGATGACCAACCCCCTATGGAGGCCCAGTAT
GCAGAGGAGGGCCCAGGACCTGGGATCTTCAGAGCAGAGCCTGGAGACCAGCAGCATCCCAGTAGGCC
AGACTGGGCCATAGGGGAAATGACAGGGTGGGGACAGTGGAGGGCAATCATCCTACATTCCCCGGATC
CTCCTTGGGGTCAGCCCCACATGATTGATGTTTCTCAGGCAGTGTGCTGGCGTTCCATGCGACGTGGC
TGTGCAGTGCTGGGAGCCCTGGGGCTGCTGGCCGGTGCAGGTGTTGGCTCATGGCTCCTAGTGCTGTA
TCTGTGTCCTGCTGCCTCTCAGCCCATTTCCGGGACCTTGCAGGATGAGGAGATAACTTTGAGCTGCT
CAGAGGCCAGCGCTGAGGAAGCTCTGCTCCCTGCACTTCCCAAAACACCTGCACTTCCCAAAACAGTA
TCTTTCAGAATAAACAGCGAAGACTTCTTGCTGGAAGCGCAAGTGAGGGATCAGCCACGCTGGCTCCT
GGTCTGCCATGAGGGCTGGAGCCCCGCCCTGGGGCTGCAGATCTGCTGGAGCCTTGGGCATCTCAGAC
TCACTCACCACAAGGGAGTAAACCTCACTGACATCAAACTCAACAGTTCCCAGGAGTTTGCTCAGCTC
TCTCCTAGACTGGGAGGCTTCCTGGAGGAGGCGTGGCAGCCCAGTAGGACTACTGAGGCTGTTAGGAA
CAACTGCACTTCTGGTCAAGTTGTTTCCCTCAGATGCTCTGAGTGTGGAGCGAGGCCCCTGGCTTCCC
GGATAGTTGGTGGGCAGTCTGTGGCTCCTGGGCGCTGGCCGTGGCAGGCCAGCGTGGCCCTGGGCTTC
CGGCACACGTGTGGGGGCTCTGTGCTAGCGCCACGCTGGGTGGTGACTGCTGCACATTGTATGCACAG
TTTCAGGCTGGCCCGCCTGTCCAGCTGGCGGGTTCATGCGGGGCTGGTCAGCCACAGTGCCGTCAGGC
CCCACCAAGGGGCTCTGGTGGAGAGGATTATCCCACACCCCCTCTACAGTGCCCAGAATCATGACTAC
GACGTCGCCCTCCTGAGGCTCCAGACCGCTCTCAACTTCTCAGACACTGTGGGCGCTGTGTGCCTGCC
GGCCAAGGAACAGCATTTTCCGAAGGGCTCGCGGTGCTGGGTGTCTGGCTGGGGCCACACCCACCCTA
GCCATACTTACAGCTCGGATATGCTCCAGGACACGGTGGTGCCCCTGTTCAGCACTCAGCTCTGCAAC
AGCTCTTGCGTGTACAGCGGAGCCCTCACCCCCCGCATGCTTTGCGCTGGCTACCTGGACGGAAGGGC
TGATGCATGCCAGGGAGATAGCGGGGGCCCCCTAGTGTGCCCAGATGGGGACACATGGCGCCTAGTGG
GGGTGGTCAGCTGGGGGCGTGGCTGCGCAGAGCCCAATCACCCAGGTGTCTACGCCAAGGTAGCTGAG
TTTCTGGACTGGATCCATGACACTGCTCAGGTGAGTGTGGGGGCAGGAGTAGGGCAGGGAGATTTCTA
AAGGACCTGCCCTCGAATGCAAGGAACCTTACCCCTTAGGCCCGGGCCCTGCTGGGGACTGGGGAGGG
TGCTAGGACATATTCCCCAGAGTGAGTGGAGGAAGAAGTGAAGCTTAAACATGGAATCCATTGGATTT
CTATCAGTTTAAGGATGAACTGGGTAAGAGTATGCCTGAGTTTGTATCCCAGATCTACCATTTCCTGT

GTCGACCTTTGGCAAATTTCTAACTTTGTTAAACCTTAATTTCCTGATAATAACCATGATGGCTACTT ATATGCTATTGTTATATGCTATTAAATAAGACCCGTACAATGCC			
NOV4k, SNP13382466 of CG125860-02, Protein Sequence	SEQ ID NO: 74	513 aa	MW at 55565.5kD
	SNP Pos: 417		SNP Change: Leu to Phe

MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPSRPDWAIGEMTGWGQWRAIILHSPDPPWGQPHM
IDVSQAVCWRSMRRGCAVLGALGLLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEA
LLPALPKTPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVN
LTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAVRNNCTSGQVVSLRCSECGARPLASRIVGGQSV
APGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVE
RIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDM
LQDTVVPLFSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRG
CAEPNHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 4B.

Table 4B. Comparison of the NOV4 protein sequences.

```
NOV4a   MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPSRPDWAIGEMTGWGQWRAIILHSPD
NOV4b   MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHP-------------------------

NOV4a   PPWGQPHMIDVSQAVCWRSMRRGCAVLGALGLLLAGAGVGSWLLVLYLCPAASQPISGTLQ
NOV4b   ----------ISQAVCWRSMRRGCAVLGALGLLLAGAGVGSWLLVLYLCPAASQPISGTLQ

NOV4a   DEEITLSCSEASAEEALLPALPKTPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGW
NOV4b   DEEITLSCSEASAEEALLPALPKT------VSFRINSEDFLLEAQVRDQPRWLLVCHEGW

NOV4a   SPALGLQICWSLGHLRLTHHKGVNLTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAV
NOV4b   SPALGLQICWSLGHLRLTHHKGVNLTDIKLNSSQEFAQLSPRLGGFLEEAWQPSRTTEAV

NOV4a   RNNCTSGQVVSLRCSECGARPLASRIVGGQSVAPGRWPWQASVALGFRHTCGGSVLAPRW
NOV4b   RNNCTSGQVVSLRCSECGARPLASRIVGGQSVAPGRWPWQASVALGFRHTCGGSVLAPRW

NOV4a   VVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVERIIPHPLYSAQNHDYDVALL
NOV4b   VVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVERIIPHPLYSAQNHDYDVALL

NOV4a   RLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDMLQDTVVPLLSTQ
NOV4b   RLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDMLQDTVVPLLSTQ

NOV4a   LCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRGCAEP
NOV4b   LCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLVGVVSWGRGCAEP

NOV4a   NHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF
NOV4b   NHPGVYAKVAEFLDWIHDTAQVSVGAGVGQGDF

NOV4a  (SEQ ID NO:  54)
NOV4b  (SEQ ID NO:  56)
```

Further analysis of the NOV4a protein yielded the following properties shown in Table 4C.

Table 4C. Protein Sequence Properties NOV4a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 7;  pos.chg 0;  neg.chg 2
      H-region:  length 4;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1): -12.80
      possible cleavage site: between 13 and 14

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -3.40   Transmembrane   89 - 105
      PERIPHERAL  Likelihood =  5.57 (at 173)
      ALOM score:  -3.40  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 96
      Charge difference: -8.0   C(-6.0) - N( 2.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 2   (cytoplasmic tail 1 to 89)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  8.83
      Hyd Moment(95):  8.90    G content:       0
      D/E content:     2       S/T content:     1
      Score: -6.04

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.2%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none
```

```
PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: too long tail

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 89

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       34.8 %: mitochondrial
       26.1 %: cytoplasmic
       17.4 %: Golgi
        8.7 %: endoplasmic reticulum
        4.3 %: extracellular, including cell wall
        4.3 %: nuclear
        4.3 %: vesicles of secretory system

>> prediction for CG125860-02 is mit (k=23)
```

A search of the NOV4a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 4D.

Table 4D. Geneseq Results for NOV4a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV4a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABG76906	Human hepsin/plasma transmembrane serine protease-like protein - Homo sapiens, 472 aa. [WO200233087-A2, 25-APR-2002]	1..513 1..472	470/513 (91%) 472/513 (91%)	0.0
AAU82752	Amino acid sequence of novel human protease #51 - Homo sapiens, 457 aa. [WO200200860-A2, 03-JAN-2002]	1..504 1..456	451/504 (89%) 453/504 (89%)	0.0
AAB11699	Human serine protease BSSP2 (hBSSP2), SEQ ID NO:10 - Homo sapiens, 457 aa. [WO200031272-A1, 02-JUN-2000]	1..504 1..456	451/504 (89%) 453/504 (89%)	0.0
AAB08950	Human secreted protein sequence encoded by gene 22 SEQ ID NO:107 - Homo sapiens, 480 aa. [WO200017222-A1, 30-MAR-2000]	69..504 57..479	417/436 (95%) 419/436 (95%)	0.0
AAB08912	Human secreted protein sequence encoded by gene 22 SEQ ID NO:69 - Homo sapiens, 414 aa. [WO200017222-A1, 30-MAR-2000]	80..504 1..412	409/425 (96%) 410/425 (96%)	0.0

In a BLAST search of public sequence databases, the NOV4a protein was found to have homology to the proteins shown in the BLASTP data in Table 4E.

Table 4E. Public BLASTP Results for NOV4a				
Protein Accession Number	**Protein/Organism/Length**	**NOV4a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9H3S3	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Homo sapiens (Human), 457 aa.	1..504 1..456	451/504 (89%) 453/504 (89%)	0.0
Q8CJ17	Adrenal mitochondrial protease long variant - Rattus norvegicus (Rat), 445 aa.	11..502 1..444	360/492 (73%) 393/492 (79%)	0.0
Q8CDR0	Transmembrane protease - Mus musculus (Mouse), 455 aa.	1..502 1..454	359/502 (71%) 394/502 (77%)	0.0
Q9ER04	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin) - Mus musculus (Mouse), 455 aa.	1..502 1..454	359/502 (71%) 394/502 (77%)	0.0
Q8CJ16	Adrenal mitochondrial protease short variant - Rattus norvegicus (Rat), 371 aa.	151..502 26..370	289/352 (82%) 312/352 (88%)	e-177

PFam analysis indicates that the NOV4a protein contains the domains shown in the Table 4F.

Table 4F. Domain Analysis of NOV4a			
Pfam Domain	**NOV4a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
trypsin	266..496	111/265 (42%) 198/265 (75%)	3.9e-87

Example 5.

The NOV5 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 5A.

Table 5A. NOV5 Sequence Analysis

NOV5a, CG155759-02 DNA Sequence	SEQ ID NO: 75	941 bp
	ORF Start: ATG at 4	ORF Stop: TAG at 931

GACATGGAACAGGATAATACAACATTGCTGACAGAGTTTGTTCTCACAGGACTTACATATCAGCCAGA
GTGGAAAATGCCCCTGTTCTTGGTGTTCTTGGTGATCTATCTCATCACTATTGTGTGGAACCTTGGTC
TGATTGCTCTTATCTGGAATGACCCACAACTTCACATCCCCATGTACTTTTTTCTTGGGAGTTTAGCC
TTTGTTGATGCTTGGATATCTTCCACAGTAACTCCCAAAATGTTGGTTAATTTCTTGGCCAAAAACAG
GATGATATCTCTGTCTGAATGCATGATTCAATTTTTTCCTTTGCATTTGGTGGAACTACAGAATGTT
TTCTCTTGGCAACAATGGCATATGATCGCTATGTAGCCATATGCAAACCTTTACTATATCCAGTGATT
ATGAACAATTCACTATGCATACGGCTGTTAGCCTTCTCATTTTTAGGTGGCTTCCTCCATGCCTTAAT
TCATGAAGTCCTTATATTCAGATTAACCTTCTGCAATTCTAACATAATACATCATTTTTACTGTGATA
TTATACCACTGTTTATGATTTCCTGTACTGACCCTTCTATTAATTTTCTAATGGTTTTTATTTTGTCT
GGCTCAATTCAGGTATTCACCATTGTGACAGTTCTTAATTCTTACACATTTGCTCTTTTCACAATCCT
AAAAAAGAAGTCTGTTAGAGGCGTAAGGAAAGCCTTTTCCACCTGTGGAGCCCATCTCTTATCTGTCT
CTTTATATTATGGCCCACTTATCTTCATGTATTTGCGCCCTGCATCTCCACAAGCAGATGACCAGGAT
ATGATAGACTCTGTCTTTTATACAATCATAATTCCTTTGCTAAATCCCATTATCTACAGTCTGAGAAA
TAAACAAGTAATAGATTCATTCACAAAAATGGTAAAAAGAAATGTTTAGATTTCATA

NOV5a, CG155759-02 Protein Sequence	SEQ ID NO: 76	309 aa	MW at 35396.8kD

MEQDNTTLLTEFVLTGLTYQPEWKMPLFLVFLVIYLITIVWNLGLIALIWNDPQLHIPMYFFLGSLAF
VDAWISSTVTPKMLVNFLAKNRMISLSECMIQFFSFAFGGTTECFLLATMAYDRYVAICKPLLYPVIM
NNSLCIRLLAFSFLGGFLHALIHEVLIFRLTFCNSNIIHHFYCDIIPLFMISCTDPSINFLMVFILSG
SIQVFTIVTVLNSYTFALFTILKKKSVRGVRKAFSTCGAHLLSVSLYYGPLIFMYLRPASPQADDQDM
IDSVFYTIIIPLLNPIIYSLRNKQVIDSFTKMVKRNV

NOV5b, CG155759-01 DNA Sequence	SEQ ID NO: 77		941 bp
	ORF Start: ATG at 4		ORF Stop: TAG at 931

GACATGGAACAGGATAATACAACATTGCTGACAGAGTTTGTTCTCACAGGACTTACATATCAGCCAGA
GTGGAAAATGCCCCTGTTCTTGGTGTTCTTGGTGATCTATCTCATCACTATTGTGTGGAACCTTGGTC
TGATTGCTCTTATCTGGAATGACCCACAACTTCACATCCCCATGTACTTTTTTCTTGGGAGTTTAGCC
TTTGTTGATGCTTGGATATCTTCCACAGTAACTCCCAAAATGTTGGTTAATTTCTTGGCCAAAAACAG
GATGATATCTCTGTCTGAATGCATGATTCAATTTTTTCCTTTGCATTTGGTGGAACTACAGAATGTT
TTCTCTTGGCAACAATGGCATATGATCGCTATGTAGCCATATGCAAACCTTTACTATATCCAGTGATT
ATGAACAATTCACTATGCATACGGCTGTTAGCCTTCTCATTTTTAGGTGGCTTCCTCCATGCCTTAAT
TCATGAAGTCCTTATATTCAGATTAACCTTCTGCAATTCTAACATAATACATCATTTTTACTGTGATA
TTATACCACTGTTTATGATTTCCTGTACTGACCCTTCTATTAATTTTCTAATGGTTTTTATTTTGTCT
GGCTCAATTCAGGTATTCACCATTGTGACAGTTCTTAATTCTTACACATTTGCTCTTTTCACAATCCT
AAAAAAGAAGTCTGTTAGAGGCGTAAGGAAAGCCTTTTCCACCTGTGGAGCCCATCTCTTATCTGTCT
CTTTATATTATGGCCCACTTATCTTCATGTATTTGCGCCCTGCATCTCCACAAGCAGATGACCAAGAT
ATGATAGACTCTGTCTTTTATACAATCATAATTCCTTTGCTAAATCCCATTATCTACAGTCTGAGAAA
TAAACAAGTAATAGATTCATTCACAAAAATGGTAAAAAGAAATGTTTAGATTTCATA

NOV5b, CG155759-01 Protein Sequence	SEQ ID NO: 78	309 aa	MW at 35396.8kD

MEQDNTTLLTEFVLTGLTYQPEWKMPLFLVFLVIYLITIVWNLGLIALIWNDPQLHIPMYFFLGSLAF
VDAWISSTVTPKMLVNFLAKNRMISLSECMIQFFSFAFGGTTECFLLATMAYDRYVAICKPLLYPVIM
NNSLCIRLLAFSFLGGFLHALIHEVLIFRLTFCNSNIIHHFYCDIIPLFMISCTDPSINFLMVFILSG
SIQVFTIVTVLNSYTFALFTILKKKSVRGVRKAFSTCGAHLLSVSLYYGPLIFMYLRPASPQADDQDM
IDSVFYTIIIPLLNPIIYSLRNKQVIDSFTKMVKRNV

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 5B.

Table 5B. Comparison of the NOV5 protein sequences.

```
NOV5a   MEQDNTTLLTEFVLTGLTYQPEWKMPLFLVFLVIYLITIVWNLGLIALIWNDPQLHIPMY
NOV5b   MEQDNTTLLTEFVLTGLTYQPEWKMPLFLVFLVIYLITIVWNLGLIALIWNDPQLHIPMY

NOV5a   FFLGSLAFVDAWISSTVTPKMLVNFLAKNRMISLSECMIQFFSFAFGGTTECFLLATMAY
NOV5b   FFLGSLAFVDAWISSTVTPKMLVNFLAKNRMISLSECMIQFFSFAFGGTTECFLLATMAY

NOV5a   DRYVAICKPLLYPVIMNNSLCIRLLAFSFLGGFLHALIHEVLIFRLTFCNSNIIHHFYCD
NOV5b   DRYVAICKPLLYPVIMNNSLCIRLLAFSFLGGFLHALIHEVLIFRLTFCNSNIIHHFYCD

NOV5a   IIPLFMISCTDPSINFLMVFILSGSIQVFTIVTVLNSYTFALFTILKKKSVRGVRKAFST
NOV5b   IIPLFMISCTDPSINFLMVFILSGSIQVFTIVTVLNSYTFALFTILKKKSVRGVRKAFST

NOV5a   CGAHLLSVSLYYGPLIFMYLRPASPQADDQDMIDSVFYTIIIPLLNPIIYSLRNKQVIDS
NOV5b   CGAHLLSVSLYYGPLIFMYLRPASPQADDQDMIDSVFYTIIIPLLNPIIYSLRNKQVIDS

NOV5a   FTKMVKRNV
NOV5b   FTKMVKRNV

NOV5a  (SEQ ID NO:  76)
NOV5b  (SEQ ID NO:  78)
```

Further analysis of the NOV5a protein yielded the following properties shown in Table 5C.

Table 5C. Protein Sequence Properties NOV5a

SignalP analysis:	Cleavage site between residues 52 and 53

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 0;  neg.chg 3
      H-region:  length 10;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.77
      possible cleavage site: between 38 and 39

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   5
      INTEGRAL    Likelihood = -9.24   Transmembrane   25 -  41
      INTEGRAL    Likelihood =  0.37   Transmembrane   57 -  73
      INTEGRAL    Likelihood = -2.02   Transmembrane  142 - 158
      INTEGRAL    Likelihood = -5.79   Transmembrane  196 - 212
      INTEGRAL    Likelihood = -1.70   Transmembrane  276 - 292
      PERIPHERAL  Likelihood =  0.90 (at 173)
      ALOM score:  -9.24  (number of TMSs: 5)
```

```
MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 32
      Charge difference:  0.5   C(-0.5) - N(-1.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  6.20
      Hyd Moment(95):  3.11     G content:       0
      D/E content:     2        S/T content:     0
      Score: -7.39

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   6.5%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: VKRN

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none
```

```
checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

---------------------------

Final Results (k = 9/23):

       33.3 %: endoplasmic reticulum
       11.1 %: mitochondrial
       11.1 %: Golgi
       11.1 %: vacuolar
       11.1 %: nuclear
       11.1 %: vesicles of secretory system
       11.1 %: cytoplasmic

>> prediction for CG155759-02 is end (k=9)
```

A search of the NOV5a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 5D.

Table 5D. Geneseq Results for NOV5a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV5a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABP72221	Human G-protein coupled receptor GCREC-15 (olfactory receptor) - Homo sapiens, 309 aa. [WO2003000859-A2, 03-JAN-2003]	1..309 1..309	309/309 (100%) 309/309 (100%)	e-179
AAU85253	G-coupled olfactory receptor #114 - Homo sapiens, 314 aa. [WO200198526-A2, 27-DEC-2001]	1..309 6..314	309/309 (100%) 309/309 (100%)	e-179
AAU95610	Human olfactory and pheromone G protein-coupled receptor #97 - Homo sapiens, 314 aa. [WO200224726-A2, 28-MAR-2002]	1..309 6..314	309/309 (100%) 309/309 (100%)	e-179
AAB71190	Human GPCRX protein SEQ ID 56 - Homo sapiens, 309 aa. [WO200250275-A2, 27-JUN-2002]	1..309 1..309	309/309 (100%) 309/309 (100%)	e-179
ABP95748	Human GPCR polypeptide SEQ ID NO 306 - Homo sapiens, 314 aa. [WO200216548-A2, 28-FEB-2002]	1..309 6..314	309/309 (100%) 309/309 (100%)	e-179

In a BLAST search of public sequence databases, the NOV5a protein was found to have homology to the proteins shown in the BLASTP data in Table 5E.

Table 5E. Public BLASTP Results for NOV5a				
Protein Accession Number	**Protein/Organism/Length**	**NOV5a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8NGV7	Seven transmembrane helix receptor - Homo sapiens (Human), 314 aa.	1..309 6..314	309/309 (100%) 309/309 (100%)	e-179
Q8VG48	Olfactory receptor MOR183-1 - Mus musculus (Mouse), 309 aa.	4..308 4..308	232/305 (76%) 272/305 (89%)	e-141
Q8VEX5	Olfactory receptor MOR183-9 - Mus musculus (Mouse), 309 aa.	1..309 1..309	239/309 (77%) 270/309 (87%)	e-139
CAD37583	Sequence 195 from Patent WO0224726 - Homo sapiens (Human), 325 aa.	1..309 17..325	241/309 (77%) 267/309 (85%)	e-139
Q8NGV6	Seven transmembrane helix receptor - Homo sapiens (Human), 309 aa.	1..309 1..309	241/309 (77%) 267/309 (85%)	e-139

PFam analysis indicates that the NOV5a protein contains the domains shown in the Table 5F.

Table 5F. Domain Analysis of NOV5a			
Pfam Domain	**NOV5a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	41..290	43/276 (16%) 169/276 (61%)	1.2e-19

Example 6.

The NOV6 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 6A.

Table 6A. NOV6 Sequence Analysis		
NOV6a, CG187667-01 DNA Sequence	SEQ ID NO: 79	2680 bp
	ORF Start: ATG at 873	ORF Stop: TAG at 2484

GGGCAGGTGTTAAGCGATAAAGGGGGGCAGAACAAGTCTTTTATCCGGCCATTTAATTGGTCACCTTA
AGAAAGAATCCTATCTTCCCACCGCAGCTGAGAAGACACTTCTAAGAGGTACAGCTAAAAGGTGGAGG
ACTGATAGACAACTTAGAGGTTGGTTTCGGGAACGATAAAACAAGTAACGGGCTGCCCGGCCTGCGCC
GCGGAGTCCCGAGGAGCCTGCGCTGTGCTCCTCTCGCGGTGTCTCGTCATCTCCGGGAAGACTCGGCG
CCTGGGTCCGCGCTCTCTGGGTAAGCTTTCCGGGAAGCTTTCCCGGGAGCTCGCTGGTCCTGGCCCCA
GAAGCCTGCGGACCCGCCCAGGGAGGATAAGCAGCTGAAAGACCGCGCGGTGCCGCTCCGAGGCCCCG
GGACGTGGGCCCATGGTCGGCCTGGCGCCACCTTTCCGGGGGAAGCCACGCGCACCAGGCATCGCACG
CGGCTCTGCACCCGCGCCGCCGGACCTGAAACCCGGCGGAGGGCACACGGGGCTGCCGCTGCGGGCCC
CGGACCAACCCATGCTTACTCCGGAGCCTGTACCGGCGCCGACGGGTCGGACCTCCCTGCGCGGTGTC
GCCCAGCGGGTTCGTGCGAAAGGCGGGGCCGACTACACGCGGTGCCGCGCCCTGAGACCGTTTATCTG
CAGTCAACGCAGCCTCCCGGCTCAGCCTGGGAAGATGCGCGAATCGGGAACCCCAGAGCGCGGTGGCT
AGACCGGGCTCCGCCGCCTCCCCCACAGCCCCTTTCCTAATCGTTCAGACGGAGCCTGGTCGACTTCG
CCGGAGACTGCCAGATCTCGTTCCTCTTCCCTGTGTCATCTTCTTAATTATAAATAATGGGGGATGAA
GATAAAAGAATTACATATGAAGATTCAGAACCATCCACAGGAATGAATTACACGCCCTCCATGCATCA
AGAAGCACAGGAGGAGACAGTTATGAAGCTCAAAGGTATAGATGCAAATGAACCAACAGAAGGAAGTA
TTCTTTTGAAAAGCAGTGAAAAAAAGCTACAAGAAACACCAACTGAAGCAAATCACGTACAAAGACTG
AGACAAATGCTGGCTTGCCCTCCACATGGTTTACTGGACAGGGTCATAACAAATGTTACCATCATTGT
TCTTCTGTGGGCTGTAGTTTGGTCAATTACTGGCAGTGAATGTCTTCCTGGAGGAAACCTATTTGGAA
TTATAATCCTATTCTATTGTGCCATCATTGGTGGTAAACTTTTGGGGCTTATTAAGTTACCTACATTG
CCTCCACTGCCTTCTCTTCTTGGCATGCTGCTTGCAGGGTTTCTCATCAGAAATATCCCAGTCATCAA
CGATAATGTGCAGATCAAGCACAAGTGGTCTTCCTCTTTGAGAAGCATAGCCCTGTCTATCATTCTGG
TTCGTGCTGGCCTTGGTCTGGATTCAAAGGCCCTGAAGAAGTTAAAGGGCGTTTGTGTAAGACTGTCC
ATGGGTCCCTGTATTGTGGAGGCGTGCACATCTGCTCTTCTTGCCCATTACCTGCTGGGTTTACCATG
GCAATGGGGATTTATACTGGGTTTTGTTTTAGGTGCTGTATCTCCAGCTGTTGTGGTGCCTTCAATGC
TCCTTTTGCAGGGAGGAGGCTATGGTGTTGAGAAGGGTGTCCCAACCTTGCTCATGGCAGCTGGCAGC
TTCGATGACATTCTGGCCATCACTGGCTTCAACACATGCTTGGGCATAGCCTTTTCCACAGGCTCTAC
TGTCTTTAATGTCCTCAGAGGAGTTTTGGAGGTGGTAATTGGTGTGGCAACTGGATCTGTTCTTGGAT
TTTTCATTCAGTACTTTCCAAGCCGTGACCAGGACAAACTTGTGTGTAAGAGAACATTCCTTGTGTTG
GGGTTGTCTGTGCTAGCTGTGTTCAGCAGTGTGCATTTTGGTTTCCCTGGATCAGGAGGACTGTGCAC
GTTGGTCATGGCTTTCCTTGCAGGCATGGGATGGACCAGCGAAAAGGCAGAGGTTGAAAAGATAATTG
CAGTTGCCTGGGACATTTTTCAGCCCCTTCTTTTTGGACTAATTGGAGCAGAGGTATCTATTGCATCT
CTCAGACCAGAAACTGTAGGCCTTTGTGTTGCCACCGTAGGCATTGCAGTATTGATACGAATTTTGAC
TACATTTCTGATGGTGTGTTTTGCTGGTTTTAACTTAAAAGAAAAGATATTTATTTCTTTTGCATGGC
TTCCAAAGGCCACAGTTCAGGCTGCAATAGGATCTGTGGCTTTGGACACAGCAAGGTCACATGGAGAG
AAACAATTAGAAGACTATGGAATGGATGTGTTGACAGTGGCATTTTTGTCCATCCTCATCACAGCCCC

AATTGGAAGTCTGCTTATTGGTTTACTGGGCCCCAGGCTTCTGCAGAAAGTTGAACATCAAAATAAAG
ATGAAGAAGTTCAAGGAGAGACTTCTGTGCAAGTTTAGGAAGCGCGGATTCTATTACTGGAAACTTTG
GGACTGAAAGGCCAAAGCTTCTGGGCCCACCATCAACGCAGCTCCGCTTTCATTTCTTTCACATACAA
CTTTCCACATAAGATTTCATGCGGAAAAAAAAAAAAAACTCACAAAGGTTTTATACTGATAACAGTAT
ATTAAGTGTTTACTTTGTACACAGCGTC

NOV6a, CG187667-01 Protein Sequence	SEQ ID NO: 80	537 aa	MW at 57563.3kD

MGDEDKRITYEDSEPSTGMNYTPSMHQEAQEETVMKLKGIDANEPTEGSILLKSSEKKLQETPTEANH
VQRLRQMLACPPHGLLDRVITNVTIIVLLWAVVWSITGSECLPGGNLFGIIILFYCAIIGGKLLGLIK
LPTLPPLPSLLGMLLAGFLIRNIPVINDNVQIKHKWSSSLRSIALSIILVRAGLGLDSKALKKLKGVC
VRLSMGPCIVEACTSALLAHYLLGLPWQWGFILGFVLGAVSPAVVVPSMLLLQGGGYGVEKGVPTLLM
AAGSFDDILAITGFNTCLGIAFSTGSTVFNVLRGVLEVVIGVATGSVLGFFIQYFPSRDQDKLVCKRT
FLVLGLSVLAVFSSVHFGFPGSGGLCTLVMAFLAGMGWTSEKAEVEKIIAVAWDIFQPLLFGLIGAEV
SIASLRPETVGLCVATVGIAVLIRILTTFLMVCFAGFNLKEKIFISFAWLPKATVQAAIGSVALDTAR
SHGEKQLEDYGMDVLTVAFLSILITAPIGSLLIGLLGPRLLQKVEHQNKDEEVQGETSVQV

NOV6b, CG187667-02 DNA Sequence	SEQ ID NO: 81	2285 bp
	ORF Start: ATG at 873	ORF Stop: TAA at 2235

GGGCAGGTGTTAAGCGATAAAGGGGGGCAGAACAAGTCTTTTATCCGGCCATTTAATTGGTCACCTTA
AGAAAGAATCCTATCTTCCCACCGCAGCTGAGAAGACACTTCTAAGAGGTACAGCTAAAAGGTGGAGG
ACTGATAGACAACTTAGAGGTTGGTTTCGGGAACGATAAAACAAGTAACGGGCTGCCCGGCCTGCGCC
GCGGAGTCCCGAGGAGCCTGCGCTGTGCTCCTCTCGCGGTGTCTCGTCATCTCCGGGAAGACTCGGCG
CCTGGGTCCGCGCTCTCTGGGTAAGCTTTCCGGGAAGCTTTCCCGGGAGCTCGCTGGTCCTGGCCCCA
GAAGCCTGCGGACCCGCCCAGGGAGGATAAGCAGCTGAAAGACCGCGCGGTGCCGCTCCGAGGCCCCG
GGACGTGGGCCCATGGTCGGCCTGGCGCCACCTTTCCGGGGGAAGCCACGCGCACCAGGCATCGCACG
CGGCTCTGCACCCGCGCCGCCGGACCTGAAACCCGGCGGAGGGCACACGGGGCTGCCGCTGCGGGCCC
CGGACCAACCCATGCTTACTCCGGAGCCTGTACCGGCGCCGACGGGTCGGACCTCCCTGCGCGGTGTC
GCCCAGCGGGTTCGTGCGAAAGGCGGGGCCGACTACACGCGGTGCCGCGCCCTGAGACCGTTTATCTG
CAGTCAACGCAGCCTCCCGGCTCAGCCTGGGAAGATGCGCGAATCGGGAACCCCAGAGCGCGGTGGCT
AGACCGGGCTCCGCCGCCTCCCCCACAGCCCCTTTCCTAATCGTTCAGACGGAGCCTGGTCGACTTCG
CCGGAGACTGCCAGATCTCGTTCCTCTTCCCTGTGTCATCTTCTTAATTATAAATAATGGGGGATGAA
GATAAAAGAATTACATATGAAGATTCAGAACCATCCACAGGAATGAATTACACGCCCTCCATGCATCA
AGAAGCACAGGAGGAGACAGTTATGAAGCTCAAAGGTATAGATGCAAATGAACCAACAGAAGGAAGTA
TTCTTTTGAAAAGCAGTGAAAAAAAGCTACAAGAAACACCAACTGAAGCAAATCACGTACAAAGACTG
AGACAAATGCTGGCTTGCCCTCCACATGGTTTACTGGACAGGGTCATAACAAATGTTACCATCATTGT
TCTTCTGTGGGCTGTAGTTTGGTCAATTACTGGCAGTGAATGTCTTCCTGGAGGAAACCTATTTGGAA
TTATAATCCTATTCTATTGTGCCATCATTGGTGGTAAACTTTTGGGGCTTATTAAGTTACCTACATTG
CCTCCACTGCCTTCTCTTCTTGGCATGCTGCTTGCAGGGTTTCTCATCAGAAATATCCCAGTCATCAA
CGATAATGTGCAGATCAAGCACAAGTGGTCTTCCTCTTTGAGAAGCATAGCCCTGTCTATCATTCTGG
TTCGTGCTGGCCTTGGTCTGGATTCAAAGGCCCTGAAGAAGTTAAAGGGCGTTTGTGTAAGACTGTCC
ATGGGTCCCTGTATTGTGGAGGCGTGCACATCTGCTCTTCTTGCCCATTACCTGCTGGGTTTACCATG
GCAATGGGGATTTATACTGGGTTTTGTTTTAGGTGCTGTATCTCCAGCTGTTGTGGTGCCTTCAATGC
TCCTTTTGCAGGGAGGAGGCTATGGTGTTGAGAAGGGTGTCCCAACCTTGCTCATGGCAGCTGGCAGC
TTCGATGACATTCTGGCCATCACTGGCTTCAACACATGCTTGGGCATAGCCTTTTCCACAGGCTCTAC
TGTCTTTAATGTCCTCAGAGGAGTTTTGGAGGTGGTAATTGGTGTGGCAACTGGATCTGTTCTTGGAT
TTTTCATTCAGTACTTTCCAAGCCGTGACCAGGACAAACTTGTGTGTAAGAGAACATTCCTTGTGTTG
GGGTTGTCTGTGCTAGCTGTGTTCAGCAGTGTGCATTTTGGTTTCCCTGGATCAGGAGGACTGTGCAC
GTTGGTCATGGCTTTCCTTGCAGGCATGGGATGGACCAGCGAAAAGGCAGAGGTTGAAAAGATAATTG
CAGTTGCCTGGGACATTTTTCAGCCCCTTCTTTTTGGACTAATTGGAGCAGAGGTATCTATTGCATCT
CTCAGACCAGAAACTGTAGGAAGCGCGGATTCTATTACTGGAAACTTTGGGACTGAAAGGCCAAAGCT
TCTGGGCCCACCATCAACGCAGCTCCGCTTTCATTTCTTTCACATACAACTTTCCACATAAGATTTCA
TGCGGAAAAAAAAAAAAAACTCACAAAGGTTTTATACTGAT

NOV6b, CG187667-02 Protein Sequence	SEQ ID NO: 82	454 aa	MW at 48709.8kD

MGDEDKRITYEDSEPSTGMNYTPSMHQEAQEETVMKLKGIDANEPTEGSILLKSSEKKLQETPTEANH
VQRLRQMLACPPHGLLDRVITNVTIIVLLWAVVWSITGSECLPGGNLFGIIILFYCAIIGGKLLGLIK
LPTLPPLPSLLGMLLAGFLIRNIPVINDNVQIKHKWSSSLRSIALSIILVRAGLGLDSKALKKLKGVC

VRLSMGPCIVEACTSALLAHYLLGLPWQWGFILGFVLGAVSPAVVVPSMLLLQGGGYGVEKGVPTLLM
AAGSFDDILAITGFNTCLGIAFSTGSTVFNVLRGVLEVVIGVATGSVLGFFIQYFPSRDQDKLVCKRT
FLVLGLSVLAVFSSVHFGFPGSGGLCTLVMAFLAGMGWTSEKAEVEKIIAVAWDIFQPLLFGLIGAEV
SIASLRPETVGSADSITGNFGTERPKLLGPPSTQLRFHFFHIQLST

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 6B.

Table 6B. Comparison of the NOV6 protein sequences.

```
NOV6a   MGDEDKRITYEDSEPSTGMNYTPSMHQEAQEETVMKLKGIDANEPTEGSILLKSSEKKLQ
NOV6b   MGDEDKRITYEDSEPSTGMNYTPSMHQEAQEETVMKLKGIDANEPTEGSILLKSSEKKLQ

NOV6a   ETPTEANHVQRLRQMLACPPHGLLDRVITNVTIIVLLWAVVWSITGSECLPGGNLFGIII
NOV6b   ETPTEANHVQRLRQMLACPPHGLLDRVITNVTIIVLLWAVVWSITGSECLPGGNLFGIII

NOV6a   LFYCAIIGGKLLGLIKLPTLPPLPSLLGMLLAGFLIRNIPVINDNVQIKHKWSSSLRSIA
NOV6b   LFYCAIIGGKLLGLIKLPTLPPLPSLLGMLLAGFLIRNIPVINDNVQIKHKWSSSLRSIA

NOV6a   LSIILVRAGLGLDSKALKKLKGVCVRLSMGPCIVEACTSALLAHYLLGLPWQWGFILGFV
NOV6b   LSIILVRAGLGLDSKALKKLKGVCVRLSMGPCIVEACTSALLAHYLLGLPWQWGFILGFV

NOV6a   LGAVSPAVVVPSMLLLQGGGYGVEKGVPTLLMAAGSFDDILAITGFNTCLGIAFSTGSTV
NOV6b   LGAVSPAVVVPSMLLLQGGGYGVEKGVPTLLMAAGSFDDILAITGFNTCLGIAFSTGSTV

NOV6a   FNVLRGVLEVVIGVATGSVLGFFIQYFPSRDQDKLVCKRTFLVLGLSVLAVFSSVHFGFP
NOV6b   FNVLRGVLEVVIGVATGSVLGFFIQYFPSRDQDKLVCKRTFLVLGLSVLAVFSSVHFGFP

NOV6a   GSGGLCTLVMAFLAGMGWTSEKAEVEKIIAVAWDIFQPLLFGLIGAEVSIASLRPETVGL
NOV6b   GSGGLCTLVMAFLAGMGWTSEKAEVEKIIAVAWDIFQPLLFGLIGAEVSIASLRPETVGS

NOV6a   CVATVGIAVLIRILTTFLMVCFAGFNLKEKIFISFAWLPKATVQAAIGSVALDTARSHGE
NOV6b   ADSITGNFGTERPK--LLG-----------P-------PSTQLRFHFFHIQLST------

NOV6a   KQLEDYGMDVLTVAFLSILITAPIGSLLIGLLGPRLLQKVEHQNKDEEVQGETSVQV
NOV6b   ------------------------------------------------------------

NOV6a  (SEQ ID NO:  80)
NOV6b  (SEQ ID NO:  82)
```

Further analysis of the NOV6a protein yielded the following properties shown in Table 6C.

Table 6C. Protein Sequence Properties NOV6a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 2;  neg.chg 4
      H-region:  length 0;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1): -13.82
      possible cleavage site: between 48 and 49

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:  10
      INTEGRAL    Likelihood = -5.20   Transmembrane   88 - 104
      INTEGRAL    Likelihood = -6.10   Transmembrane  118 - 134
      INTEGRAL    Likelihood = -3.13   Transmembrane  146 - 162
      INTEGRAL    Likelihood = -1.06   Transmembrane  176 - 192
      INTEGRAL    Likelihood = -6.26   Transmembrane  234 - 250
      INTEGRAL    Likelihood = -4.09   Transmembrane  308 - 324
      INTEGRAL    Likelihood = -4.94   Transmembrane  341 - 357
      INTEGRAL    Likelihood = -3.19   Transmembrane  388 - 404
      INTEGRAL    Likelihood = -7.91   Transmembrane  418 - 434
      INTEGRAL    Likelihood = -6.69   Transmembrane  493 - 509
      PERIPHERAL  Likelihood =  0.63 (at 213)
      ALOM score:  -7.91  (number of TMSs: 10)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 95
      Charge difference: -2.5   C(-1.0) - N( 1.5)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  9.18
      Hyd Moment(95):  7.68    G content:       1
      D/E content:     2       S/T content:     0
      Score: -6.91

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.4%
```

```
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

-------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       11.1 %: nuclear
       11.1 %: vesicles of secretory system
       11.1 %: mitochondrial

>> prediction for CG187667-01 is end (k=9)
```

A search of the NOV6a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 6D.

Table 6D. Geneseq Results for NOV6a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABU11723	Human MDDT polypeptide SEQ ID 670 - Homo sapiens, 538 aa. [WO200279449-A2, 10-OCT-2002]	1..537 2..538	537/537 (100%) 537/537 (100%)	0.0
ABP69843	Human polypeptide SEQ ID NO 1890 - Homo sapiens, 537 aa. [WO200270539-A2, 12-SEP-2002]	1..537 1..537	526/537 (97%) 530/537 (97%)	0.0
AAY94918	Human secreted protein clone dd504_18 protein sequence SEQ ID NO:42 - Homo sapiens, 396 aa. [WO200009552-A1, 24-FEB-2000]	100..464 1..365	231/365 (63%) 290/365 (79%)	e-137
AAU01672	Human secreted protein encoded by gene #13 - Homo sapiens, 370 aa. [WO200123409-A2, 05-APR-2001]	153..519 1..366	224/367 (61%) 287/367 (78%)	e-130
AAU01642	Human secreted protein immunogenic epitope encoded by gene #2 - Homo sapiens, 323 aa. [WO200123409-A2, 05-APR-2001]	214..519 1..305	185/306 (60%) 236/306 (76%)	e-106

In a BLAST search of public sequence databases, the NOV6a protein was found to have homology to the proteins shown in the BLASTP data in Table 6E.

Table 6E. Public BLASTP Results for NOV6a				
Protein Accession Number	**Protein/Organism/Length**	**NOV6a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAH47447	LOC133308 protein - Homo sapiens (Human), 519 aa.	19..537 1..519	519/519 (100%) 519/519 (100%)	0.0
Q96D95	Hypothetical protein - Homo sapiens (Human), 354 aa (fragment).	101..426 1..326	320/326 (98%) 321/326 (98%)	0.0
Q8C0X2	Hypothetical glutamic acid-rich region/Na+/H+ exchanger containing protein - Mus musculus (Mouse), 565 aa.	53..521 83..551	263/470 (55%) 359/470 (75%)	e-159
Q95JS4	Hypothetical 55.5 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 511 aa.	73..519 53..498	261/447 (58%) 347/447 (77%)	e-157
Q8CDX4	Hypothetical Na+/H+ exchanger containing protein - Mus musculus (Mouse), 295 aa.	253..537 1..284	247/285 (86%) 264/285 (91%)	e-138

PFam analysis indicates that the NOV6a protein contains the domains shown in the Table 6F.

Table 6F. Domain Analysis of NOV6a			
Pfam Domain	**NOV6a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
LrgA	114..199	26/114 (23%) 60/114 (53%)	0.7
Na_H_Exchanger	119..520	85/470 (18%) 286/470 (61%)	1.3e-16

Example 7.

The NOV7 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 7A.

Table 7A. NOV7 Sequence Analysis		
NOV7a, CG187676-01 DNA Sequence	SEQ ID NO: 83	2474 bp
	ORF Start: ATG at 5	ORF Stop: TGA at 2270

AAAAATGAAAAAACAGAGGAAAATTCTATGGAGGAAAGGAATCCACTTAGCCTTTTCTGAGAAATGGA
ATACTGGGTTTGGAGGCTTTAAGAAGTTTTATTTTCACCAACACTTGTGCATTCTGAAAGCTAAGCTG
GGAAGGCCAGTTACTTGGAATAGACAGTTGAGACATTTCCAGGGTAGAAAGAAAGCTCTTCAAATCCA
GAAAACGTGGATCAAGGATGAACCCCTTTGTGCTAAGACCAAGTTCAATGTGGCTACTCAAAATGTTA
GTACTTTGTCCTCTAAAGTGAAAAGAAAGGACGCTAAACACTTCATTTCCTCCTCAAAGACTCTCCTG
AGACTCCAAGCAGAGAAGCTGTTGTCATCAGCAAAGAATTCTGACCATGAATACTGCAGAGAGAAAAA
TCTCTTGAAGGCAGTTACTGACTTTCCATCAAATAGTGCTTTAGGTCAGGCCAATGGTCACAGACCTA
GGACAGACCCACAACCTTCTGACTTTCCCATGAAGTTCAATGGGGAGAGCCAAAGTCCAGGTGAGAGT
GGCACGATTGTGGTCACCTTGAACAACCATAAGAGAAAGGGCTTTTGTTACGGCTGCTGCCAAGGGCC
GGAGCACCACAGGAATGGGGGACCCTTGATTCCAAAAAAGTTCCAACTTAACCAACATAGAAGGATAA
AATTATCTCCTCTTATGATGTATGAGAAATTATCCATGATTAGATTTCGGTACAGGATTCTCAGATCC
CAGCACTTCAGAACCAAAAGCAAGGTTTGCAAGCTAAGAAAAGCCCAGCGAAGCTGGGTACAGAAAGT
CACTGGGGACCATCAAGAGACCCGTAGGGAGAACGGTGAGGGTGGCAGTTGCAGCCCATTTCCTTCCC
CAGAACCTAAAGACCCTTCTTGTCGGCATCAGCCGTACTTTCCAGATATGGACAGCAGTGCTGTGGTG
AAGGGGACGAACTCTCATGTGCCTGATTGCCACACTAAAGGAAGCTCTTTCTTGGGCAAGGAGCTTAG
TTTAGACGAAGCATTCCCTGACCAACAGAATGGCAGTGCCACAAACGCCTGGGACCAGTCATCCTGTT
CTTCTCCTAAGTGGGAGTGTACAGAGCTGATTCATGACATCCCCTTACCAGAACATCGTTCTAATACC
ATGTTCATTTCAGAAACTGAAAGAGAAATTATGACTCTGGGTCAGGAAAATCAGACAAGTTCTGTCAG
TGATGACAGAGTAAAACTGTCAGTGTCTGGAGCAGATACATCTGTGAGTAGCGTAGATGGGCCTGTGT
CCCAAAAGGCTGTTCAAAATGAGAACTCATACCAGATGGAGGAGGATGGATCTCTCAAGCAGAGCATT
CTTAGTTCTGAGTTGCTGGACCACCCTTACTGTAAAAGTCCACTGGAGGCTCCCTTGGTGTGCAGTGG
ACTCAAACTAGAAAATCAAGTAGGAGGTGGAAAGAACAGTCAGAAAGCCTCTCCAGTGGATGATGAAC
AGCTGTCAGTCTGTCTTTCTGGATTCCTAGATGAGGTTATGAAGAAGTATGGCAGTTTGGTTCCACTC
AGTGAAAAAGAAGTCCTTGGAAGATTAAAAGATGTCTTTAATGAAGACTTTTCTAATAGAAAACCATT
TATCAATAGGGAAATAACAAACTATCGGGCCAGACATCAAAAATGTAACTTCCGTATCTTCTATAATA
AACACATGCTGGATATGGACGACCTGGCGACTCTGGATGGTCAGAACTGGCTGAATGACCAGGTCATT
AATATGTATGGTGAGCTGATAATGGATGCAGTCCCAGACAAAGTTCACTTCTTCAACAGCTTTTTTCA
TAGACAGCTGGTAACCAAAGGATATAATGGAGTAAAAAGATGGACTAAAAAGGTGGATTTGTTTAAAA
AGAGTCTTCTGTTGATTCCTATTCACCTGGAAGTCCACTGGTCTCTCATTACTGTGACACTCTCTAAT
CGAATTATTTCATTTTATGATTCCCAAGGCATTCATTTTAAGTTTTGTGTAGAGAATATAAGAAAGTA
TTTGCTGACTGAAGCCAGAGAAAAAAATAGACCTGAATTTCTTCAGGGTTGGCAGACTGCTGTTACGA
AGTGTATTCCACAACAGAAAAACGACAGTGACTGTGGAGTCTTTGTGCTCCAGTACTGCAAGTGCCTC
GCCTTAGAGCAGCCTTTCCAGTTTTCACAAGAAGACATGCCCCGAGTGCGGAAGAGGATTTACAAGGA
GCTATGTGAGTGCCGGCTCATGGACTGAAACTCAGCAGGGACTCTGGGAAGTCTGACCAAGTTGGAGC
AGATGGTTTGTTACTTGAATCTCCAAACACTTAGTTGAATTTTTACAGATATTTCAGATCAGTGGTGT
TGGGCCACTATTGTTACCTCAAATTTATTTTTTGCCCTTATTCATTTCTCCAGCTACCATGTACTATT
GTTTAATGTTCAGTTTGGTTTCAAAA

NOV7a, CG187676-01 Protein Sequence	SEQ ID NO: 84	755 aa	MW at 86692.3kD

MKKQRKILWRKGIHLAFSEKWNTGFGGFKKFYFHQHLCILKAKLGRPVTWNRQLRHFQGRKKALQIQK
TWIKDEPLCAKTKFNVATQNVSTLSSKVKRKDAKHFISSSKTLLRLQAEKLLSSAKNSDHEYCREKNL
LKAVTDFPSNSALGQANGHRPRTDPQPSDFPMKFNGESQSPGESGTIVVTLNNHKRKGFCYGCCQGPE
HHRNGGPLIPKKFQLNQHRRIKLSPLMMYEKLSMIRFRYRILRSQHFRTKSKVCKLRKAQRSWVQKVT
GDHQETRRENGEGGSCSPFPSPEPKDPSCRHQPYFPDMDSSAVVKGTNSHVPDCHTKGSSFLGKELSL
DEAFPDQQNGSATNAWDQSSCSSPKWECTELIHDIPLPEHRSNTMFISETEREIMTLGQENQTSSVSD
DRVKLSVSGADTSVSSVDGPVSQKAVQNENSYQMEEDGSLKQSILSSELLDHPYCKSPLEAPLVCSGL
KLENQVGGGKNSQKASPVDDEQLSVCLSGFLDEVMKKYGSLVPLSEKEVLGRLKDVFNEDFSNRKPFI
NREITNYRARHQKCNFRIFYNKHMLDMDDLATLDGQNWLNDQVINMYGELIMDAVPDKVHFFNSFFHR
QLVTKGYNGVKRWTKKVDLFKKSLLLIPIHLEVHWSLITVTLSNRIISFYDSQGIHFKFCVENIRKYL
LTEAREKNRPEFLQGWQTAVTKCIPQQKNDSDCGVFVLQYCKCLALEQPFQFSQEDMPRVRKRIYKEL
CECRLMD

Further analysis of the NOV7a protein yielded the following properties shown in Table 7B.

Table 7B. Protein Sequence Properties NOV7a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 6;  neg.chg 0
      H-region:  length 7;  peak value   0.84
      PSG score:  -3.56

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1): -10.14
      possible cleavage site: between 18 and 19

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   0
      PERIPHERAL  Likelihood =  6.31 (at 710)
      ALOM score:  -0.06  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       2       Hyd Moment(75):  3.21
      Hyd Moment(95): 10.20    G content:       1
      D/E content:     1       S/T content:     1
      Score: -3.27

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 70  GRK|KA

NUCDISC: discrimination of nuclear localization signals
      pat4: HKRK (3) at  190
      pat7: PRVRKRI (5) at  738
      bipartite: none
      content of basic residues:  15.1%
      NLS Score:  0.15

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: KKQR

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none
```

```
VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 89

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       65.2 %: nuclear
       26.1 %: mitochondrial
        4.3 %: cytoplasmic
        4.3 %: peroxisomal

>> prediction for CG187676-01 is nuc (k=23)
```

A search of the NOV7a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 7C.

Table 7C. Geneseq Results for NOV7a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV7a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABJ26650	Human protein modification + maintenance molecule protein SEQ ID No 4 - Homo sapiens, 755 aa. [WO2003000844-A2, 03-JAN-2003]	1..755 1..755	755/755 (100%) 755/755 (100%)	0.0
AAB71412	Human HIPHUM106 splice variant SEQ ID 4 - Homo sapiens, 755 aa. [GB2371801-A, 07-AUG-2002]	1..755 1..755	754/755 (99%) 755/755 (99%)	0.0
AAB71411	Human HIPHUM106 protein SEQ ID 2 - Homo sapiens, 727 aa. [GB2371801-A, 07-AUG-2002]	1..755 1..727	727/755 (96%) 727/755 (96%)	0.0
ABP69066	Human polypeptide SEQ ID NO 1113 - Homo sapiens, 419 aa. [WO200270539-A2, 12-SEP-2002]	310..752 1..415	413/443 (93%) 413/443 (93%)	0.0
AAM25617	Human protein sequence SEQ ID NO:1132 - Homo sapiens, 270 aa. [WO200153455-A2, 26-JUL-2001]	485..755 1..270	267/271 (98%) 268/271 (98%)	e-157

In a BLAST search of public sequence databases, the NOV7a protein was found to have homology to the proteins shown in the BLASTP data in Table 7D.

Table 7D. Public BLASTP Results for NOV7a				
Protein Accession Number	**Protein/Organism/Length**	**NOV7a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96HI0	Sentrin-specific protease 5 (EC 3.4.22.-) (Sentrin/SUMO-specific protease SENP5) (Protease FKSG45) - Homo sapiens (Human), 755 aa.	1..755 1..755	754/755 (99%) 754/755 (99%)	0.0
Q8WP32	Sentrin-specific protease 5 (EC 3.4.22.-) (Sentrin/SUMO-specific protease SENP5) (QtsA-16408) - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 755 aa.	1..755 1..755	714/755 (94%) 735/755 (96%)	0.0
Q8BXW0	Hypothetical SUMO/sentrin/Ubl1 specific protease containing protein - Mus musculus (Mouse), 749 aa.	1..755 1..749	619/756 (81%) 667/756 (87%)	0.0
Q8BRD5	Sentrin/SUMO-specific protease - Mus musculus (Mouse), 568 aa.	485..753 294..566	148/274 (54%) 202/274 (73%)	1e-82
Q9EP97	Sentrin-specific protease 3 (EC 3.4.22.-) (Sentrin/SUMO-specific protease SENP3) (SUMO-1 specific protease 3) (Smt3-specific isopeptidase 1) (Smt3ip1) - Mus musculus (Mouse), 568 aa.	485..753 294..566	148/274 (54%) 202/274 (73%)	1e-82

PFam analysis indicates that the NOV7a protein contains the domains shown in the Table 7E.

Table 7E. Domain Analysis of NOV7a			
Pfam Domain	**NOV7a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Peptidase_C48	563..752	62/248 (25%) 149/248 (60%)	7e-43

Example 8.

The NOV8 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 8A.

Table 8A. NOV8 Sequence Analysis		
NOV8a, CG50235-04 DNA Sequence	SEQ ID NO: 85	3120 bp
	ORF Start: ATG at 256	ORF Stop: TGA at 2719

TAACCTGTCTCACGACGGTCTAAACCCAGGCAGCCTCGGCCGCCGGGCAAGTAGCTCCGAGCGGCTGC
TTCCCGGTTGCCTCGAAGAAGACAGGGGGCGCCGCGCTCCGCTTGCTCCGCGCCTGAGCCATGCCCAG
CAGCCCTGTGTAACCACCGAGTCCCGGCCGGAGCCGACCGACCCAGTGTGCGCCGTCTTTCGGCCGAG
CTGAGCTTTCGTGCACGCAACTCCCTCTGCCCCAGCCGGCCCCGCGCCACCATGCCCCGGGCGACTGC
ACTTGGGGCCCTGGTGTCACTGCTGCTGCTGCTGCCGCTGCCTCGCGGCGCCGGGGGACTCGGGGAGC
GCCCGGACGCCACCGCAGACTACTCAGAGCTGGACGGCGAGGAGGGCACGGAGCAGCAGCTGGAGCAT
TACCACGACCCTTGCAAAGCCGCTGTCTTTTGGGGAGACATTGCCTTAGATGAAGATGACTTGAAGCT
GTTTCACATTGACAAAGCCAGAGACTGGACCAAGCAGACAGTGGGGGCAACAGGACACAGCACAGGTG
GGCTTGAAGAGCAGGCATCTGAGAGCAGCCCAGACACCACAGCCATGGACACTGGCACCAAGGAAGCT
GGAAAGGATGGCCGGGAGAATACCACACTCCTGCACAGCCCTGGGACCTTGCATGCCGCAGCCAAGAC
CTTCTCTCCCCGGGTCCGAAGAGCCACAACCTCAAGGACAGAGAGGATATGGCCTGGAGGAGTCATCC
CCTACGTCATTGGAGGGAACTTCACTGGGAGCCAGAGGGCCATTTTTAAGCAGGCCATGAGACACTGG
GAGAAGCACACCTGTGTGACCTTCATAGAAAGGACGGATGAGGAAAGCTTTATTGTATTCAGTTACAG
AACCTGTGGCTGTTGCTCCTATGTTGGGCGCCGAGGAGGAGGCCCACAGGCCATATCCATTGGGAAGA
ACTGTGACAAGTTTGGCATTGTGGCTCACGAGCTGGGCCATGTGGTTGGGTTTTGGCATGAACACACC
CGGCCAGACAGAGACCAACATGTCACCATCATCAGGGAAAACATCCAGCCAGGTCAGGAGTATAATTT
CTTAAAAATGGAAGCTGGGGAAGTGAGCTCTCTGGGAGAGACATACGACTTTGACAGCATCATGCACT
ACGCCCGGAACACCTTCTCAAGAGGAGTTTTCTTAGACACCATCCTTCCCCGTCAAGATGACAATGGC
GTCAGGCCAACCATTGGCCAGCGCGTGCGGCTCAGTCAGGGAGACATAGCTCAAGCCCGGAAGCTGTA
CAAATGCCCAGCGTGTGGGGAGACCCTGCAGGACACAACGGGAAACTTTTCTGCACCTGGTTTCCCAA
ATGGGTACCCATCTTACTCCCACTGCGTCTGGAGGATCTCGGTCACCCCAGGGGAAAAGATCGTATTA
AACTTCACATCCATGGATTTGTTTAAAAGCCGACTGTGCTGGTATGATTACGTGGAGGTCCGGGATGG
TTACTGGAGAAAAGCCCCCCTTTTGGGCAGGTTTTGTGGCGATAAGATCCCGGAGCCCCTCGTCTCCA
CGGACAGCCGGCTCTGGGTGGAGTTCCGCAGCAGCAGCAACATCTTGGGCAAGGGCTTCTTTGCAGCG
TACGAAGCTACCTGCGGGGGAGACATGAACAAAGATGCCGGTCAGATTCAATCTCCCAACTATCCGGA
TGACTACAGACCTTCCAAGGAATGTGTCTGGAGGATTACGGTTTCGGAGGGGTTTCACGTGGGACTTA
CCTTCCAAGCTTTTGAGATTGAAAGGCACGACAGCTGTGCATATGACTACCTGGAAGTCCGGGATGGC
CCCACGGAAGAGAGTGCCCTGATCGGCCACTTTTGTGGCTATGAGAAGCCGGAGGATGTGAAATCGAG
CTCCAACAGACTGTGGATGAAGTTTGTGTCCGATGGCTCTATCAATAAAGCGGGCTTTGCAGCCAATT
TTTTCAAGGAGGTGGATGAGTGTTCCTGGCCAGATCACGGCGGGTGCGAACATCGCTGTGTGAACACG
CTGGGCAGCTACAAGTGTGCCTGTGACCCTGGCTACGAGCTGGCCGCCGATAAGAAGATGTGTGAAGT
GGCCTGTGGCGGTTTCATTACCAAGCTGAATGGAACCATCACCAGCCCTGGGTGGCCGAAGGAGTATC
CCACAAACAAAAACTGTGTCTGGCAGGTGGTGGCCCCCGCTCAGTACCGGATCTCCCTTCAGTTTGAA
GTGTTTGAACTGGAAGGCAATGACGTCTGTAAGTACGACTTTGTAGAGGTGCGCAGCGGCCTGTCCCC
CGACGCCAAGCTGCACGGCAGGTTCTGCGGCTCTGAGACGCCGGAAGTCATCACCTCGCAGAGCAACA

ACATGCGCGTGGAGTTCAAGTCCGACAACACCGTCTCCAAGCGCGGCTTCAGGGCCCACTTCTTCTCA
GATAAGGACGAGTGTGCCAAGGACAACGGCGGGTGTCAGCATGAGTGCGTCAACACCTTCGGGAGCTA
CCTGTGCAGGTGCAGAAACGGCTACTGGCTCCACGAGAATGGGCATGACTGCAAAGAGGCTGGCTGTG
CACACAAGATCAGCAGTGTGGAGGGGACCCTGGCGAGCCCCAACTGGCCTGACAAATACCCCAGCCGG
AGGGAGTGTACCTGGAACATCTCTTCGACTGCAGGCCACAGAGTAAAACTCAGTGCGGGGTCAGGC**TG**
AAGGCTGAAGTGCAGACCAAAGAGCTCTATTCCCACGCCCAGTTTGGGGACAACAACTACCCGAGCGA
GGCCCGCTGTGACTGGGTGATCGTGGCAGAGGACGGCTACGGCGTGGAGCTGACATTCCGGACCTTTG
AGGTTGAGGAGGAGGCCGACTGCGGCTACGACTACATGGAAGCCTACGACGGCTACGACAGCTCAGCG
CCCAGGCTCGGCCGCTTCTGTGGCTCTGGGCCATTAGAAGAAATCTACTCTGCAGGTGATTCCCTGAT
GATTCGATTCCGCACAGATGACACCATCAACAAGAAAGGCTTTCATGCCCGATACACCAGCACCAAGT
TCCAGGATGCCCTGCACATGAAGAAATAGTGCTGATGTTCTTGAAAGACAGAAACTGAGA

NOV8a, CG50235-04 Protein Sequence	SEQ ID NO: 86	821 aa	MW at 91404.2kD

MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAVFWGDIALD
EDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAGKDGRENTTLLHSPGTL
HAAAKTFSPRVRRATTSRTERIWPGGVIPYVIGGNFTGSQRAIFKQAMRHWEKHTCVTFIERTDEESF
IVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIVAHELGHVVGFWHEHTRPDRDQHVTIIRENIQP
GQEYNFLKMEAGEVSSLGETYDFDSIMHYARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIA
QARKLYKCPACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIVLNFTSMDLFKSRLCWYDY
VEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSNILGKGFFAAYEATCGGDMNKDAGQIQ
SPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFEIERHDSCAYDYLEVRDGPTEESALIGHFCGYEKP
EDVKSSSNRLWMKFVSDGSINKAGFAANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAAD
KKMCEVACGGFITKLNGTITSPGWPKEYPTNKNCVWQVVAPAQYRISLQFEVFELEGNDVCKYDFVEV
RSGLSPDAKLHGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECV
NTFGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTAGHRVKL
SAGSG

NOV8b, CG50235-01 DNA Sequence	SEQ ID NO: 87	3350 bp
	ORF Start: ATG at 365	ORF Stop: TAG at 3341

CGCCCATTGGCTCCTCAGCCAAGCACGTACACCAAATGTCTGAACCTGCGGTTCCTCTCGTACTGAGC
AGGATTACCATGGCAACAACACATCATCAGTAGGGTAAAACTAACCTGTCTCACGACGGTCTAAACCC
AGGCAGCCTCGGCCGCCGGGCAAGTAGCTCCGAGCGGCTGCTTCCCGGTTGCCTCGAAGAAGACAGGG
GGCGCCGCGCTCCGCTTGCTCCGCGCCTGAGCCATGCCCAGCAGCCCTGTGTAACCACCGAGTCCCGG
CCGGAGCCGACCGACCCAGTGTGCGCCGTCTTTCGGCCGAGCTGAGCTTTCGTGCACGCAACTCCCTC
TGCCCCAGCCGGCCCCGCGCCACC**ATG**CCCCGGGCGACTGCACTTGGGGCCCTGGTGTCACTGCTGCT
GCTGCTGCCGCTGCCTCGCGGCGCCGGGGGACTCGGGGAGCGCCCGGACGCCACCGCAGACTACTCAG
AGCTGGACGGCGAGGAGGGCACGGAGCAGCAGCTGGAGCATTACCACGACCCTTGCAAAGCCGCTGTC
TTTTGGGGAGACATTGCCTTAGATGAAGATGACTTGAAGCTGTTTCACATTGACAAAGCCAGAGACTG
GACCAAGCAGACAGTGGGGGCAACAGGACACAGCACAGGTGGGCTTGAAGAGCAGGCATCTGAGAGCA
GCCCAGACACCACAGCCATGGACACTGGCACCAAGGAAGCTGGAAAGGATGGCCGGGAGAATACCACA
CTCCTGCACAGCCCTGGGACCTTGCATGCCGCAGCCAAGACCTTCTCTCCCCGGGTCCGAAGAGCCAC
AACCTCAAGGACAGAGAGGATATGGCCTGGAGGAGTCATCCCCTACGTCATTGGAGGGAACTTCACTG
GGAGCCAGAGGGCCATTTTTAAGCAGGCCATGAGACACTGGGAGAAGCACACCTGTGTGACCTTCATA
GAAAGGACGGATGAGGAAAGCTTTATTGTATTCAGTTACAGAACCTGTGGCTGTTGCTCCTATGTTGG
GCGCCGAGGAGGAGGCCCACAGGCCATATCCATTGGGAAGAACTGTGACAAGTTTGGCATTGTGGCTC
ACGAGCTGGGCCATGTGGTTGGGTTTTGGCATGAACACACCCGGCCAGACAGAGACCAACATGTCACC
ATCATCAGGGAAAACATCCAGCCAGGTCAGGAGTATAATTTCTTAAAAATGGAAGCTGGGGAAGTGAG
CTCTCTGGGAGAGACATACGACTTTGACAGCATCATGCACTACGCCCGGAACACCTTCTCAAGAGGAG
TTTTCTTAGACACCATCCTTCCCCGTCAAGATGACAATGGCGTCAGGCCAACCATTGGCCAGCGCGTG
CGGCTCAGTCAGGGAGACATAGCTCAAGCCCGGAAGCTGTACAAATGCCCAGCGTGTGGGGAGACCCT
GCAGGACACAACGGGAAACTTTTCTGCACCTGGTTTCCCAAATGGGTACCCATCTTACTCCCACTGCG
TCTGGAGGATCTCGGTCACCCCAGGGGAAAAGATCGTATTAAACTTCACATCCATGGATTTGTTTAAA
AGCCGACTGTGCTGGTATGATTACGTGGAGGTCCGGGATGGTTACTGGAGAAAAGCCCCCCTTTTGGG
CAGGTTTTGTGGCGATAAGATCCCGGAGCCCCTCGTCTCCACGGACAGCCGGCTCTGGGTGGAGTTCC
GCAGCAGCAGCAACATCTTGGGCAAGGGCTTCTTTGCAGCGTACGAAGCTACCTGCGGGGGAGACATG
AACAAAGATGCCGGTCAGATTCAATCTCCCAACTATCCGGATGACTACAGACCTTCCAAGGAATGTGT
CTGGAGGATTACGGTTTCAGAGGGGTTTCACGTGGGACTTACCTTCCAAGCTTTTGAGATTGAAAGGC
ACGACAGCTGTGCATATGACTACCTGGAAGTCCGGGATGGCCCCACGGAAGAGAGTGCCCTGATCGGC

CACTTTTGTGGCTATGAGAAGCCGGAGGATGTGAAATCGAGCTCCAACAGACTGTGGATGAAGTTTGT
GTCCGATGGCTCTATCAATAAAGCGGGCTTTGCAGCCAATTTTTTCAAGGAGGTGGATGAGTGTTCCT
GGCCAGATCACGGCGGGTGCGAACATCGCTGTGTGAACACGCTGGGCAGCTACAAGTGTGCCTGTGAC
CCTGGCTACGAGCTGGCCGCCGATAAGAAGATGTGTGAAGTGGCCTGTGGCGGTTTCATTACCAAGCT
GAATGGAACCATCACCAGCCCTGGGTGGCCGAAGGAGTATCCCACAAACAAAAACTGTGTCTGGCAGG
TGGTGGCCCCCACTCAGTACCGGATCTCCCTTCAGTTTGAAGTGTTTGAACTGGAAGGCAATGACGTC
TGTAAGTACGACTTTGTAGAGGTGCGCAGCGGCCTGTCCCCCGACGCCAAGCTGCACGGCAGGTTCTG
CGGCTCTGAGACGCCGGAAGTCATCACCTCGCAGAGCAACAACATGCGCGTGGAGTTCAAGTCCGACA
ACACCGTCTCCAAGCGCGGCTTCAGGGCCCACTTCTTCTCAGATAAGGACGAGTGTGCCAAGGACAAC
GGCGGGTGTCAGCATGAGTGCGTCAACACCTTCGGGAGCTACCTGTGCAGGTGCAGAAACGGCTACTG
GCTCCACGAGAATGGGCATGACTGCAAAGAGGCTGGCTGTGCACACAAGATCAGCAGTGTGGAGGGGA
CCCTGGCGAGCCCCAACTGGCCTGACAAATACCCCAGCCGGAGGGAGTGTACCTGGAACATCTCTTCG
ACTGCAGGCCACAGAGTGAAACTCACCTTTAATGAGTTTGAGATCGAGCAGCACCAGGAATGTGCCTA
TGACCACCTGGAAATGTATGACGGGCCGGACAGCCTGGCCCCCATTCTGGGCCGTTTCTGCGGCAGCA
AGAAACCAGACCCCACGGTGGCTTCCGGCAGCAAGTGCGGGGGCAGGCTGAAGGCTGAAGTGCAGACC
AAAGAGCTCTATTCCCACGCCCAGTTTGGGGACAACAACTACCCGAGCGAGGCCCGCTGTGACTGGGT
GATCGTGGCAGAGGACGGCTACGGCGTGGAGCTGACATTCCGGACCTTTGAGGTTGAGGAGGAGGCCG
ACTGCGGCTACGACTACATGGAAGCCTACGACGGCTACGACAGCTCAGCGCCCAGGCTCGGCCGCTTC
TGTGGCTCTGGGCCATTAGAAGAAATCTACTCTGCAGGTGATTCCCTGATGATTCGATTCCGCACAGA
TGACACCATCAACAAGAAAGGCTTTCATGCCCGATACACCAGCACCAAGTTCCAGGATGGCCTGCACA
TGAAGAAA**TAG**TGCTGAT

NOV8b, CG50235-01 Protein Sequence	SEQ ID NO: 88	992 aa	MW at 110925.6kD

MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAVFWGDIALD
EDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAGKDGRENTTLLHSPGTL
HAAAKTFSPRVRRATTSRTERIWPGGVIPYVIGGNFTGSQRAIFKQAMRHWEKHTCVTFIERTDEESF
IVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIVAHELGHVVGFWHEHTRPDRDQHVTIIRENIQP
GQEYNFLKMEAGEVSSLGETYDFDSIMHYARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIA
QARKLYKCPACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIVLNFTSMDLFKSRLCWYDY
VEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSNILGKGFFAAYEATCGGDMNKDAGQIQ
SPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFEIERHDSCAYDYLEVRDGPTEESALIGHFCGYEKP
EDVKSSSNRLWMKFVSDGSINKAGFAANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAAD
KKMCEVACGGFITKLNGTITSPGWPKEYPTNKNCVWQVVAPTQYRISLQFEVFELEGNDVCKYDFVEV
RSGLSPDAKLHGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECV
NTFGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTAGHRVKL
TFNEFEIEQHQECAYDHLEMYDGPDSLAPILGRFCGSKKPDPTVASGSKCGGRLKAEVQTKELYSHAQ
FGDNNYPSEARCDWVIVAEDGYGVELTFRTFEVEEEADCGYDYMEAYDGYDSSAPRLGRFCGSGPLEE
IYSAGDSLMIRFRTDDTINKKGFHARYTSTKFQDGLHMKK

NOV8c, CG50235-02 DNA Sequence	SEQ ID NO: 89	5006 bp
	ORF Start: ATG at 365	ORF Stop: TAG at 3410

CGCCCATTGGCTCCTCAGCCAAGCACGTACACCAAATGTCTGAACCTGCGGTTCCTCTCGTACTGAGC
AGGATTACCATGGCAACAACACATCATCAGTAGGGTAAAACTAACCTGTCTCACGACGGTCTAAACCC
AGGCAGCCTCGGCCGCCGGGCAAGTAGCTCCGAGCGGCTGCTTCCCGGTTGCCTCGAAGAAGACAGGG
GGCGCCGCGCTCCGCTTGCTCCGCGCCTGAGCCATGCCCAGCAGCCCTGTGTAACCACCGAGTCCCGG
CCGGAGCCGACCGACCCAGTGTGCGCCGTCTTTCGGCCGAGCTGAGCTTTCGTGCACGCAACTCCCTC
TGCCCCAGCCGGCCCCGCGCCACC**ATG**CCCCGGGCGACTGCACTTGGGGCCCTGGTGTCACTGCTGCT
GCTGCTGCCGCTGCCTCGCGGCGCCGGGGGACTCGGGGAGCGCCCGGACGCCACCGCAGACTACTCAG
AGCTGGACGGCGAGGAGGGCACGGAGCAGCAGCTGGAGCATTACCACGACCCTTGCAAAGCCGCTGTC
TTTTGGGGAGACATTGCCTTAGATGAAGATGACTTGAAGCTGTTTCACATTGACAAAGCCAGAGACTG
GACCAAGCAGACAGTGGGGGCAACAGGACACAGCACAGGTGGGCTTGAAGAGCAGGCATCTGAGAGCA
GCCCAGACACCACAGCCATGGACACTGGCACCAAGGAAGCTGGAAAGGATGGCCGGGAGAATACCACA
CTCCTGCACAGCCCTGGGACCTTGCATGCCGCAGCCAAGACCTTCTCTCCCCGGGTCCGAAGAGCCAC
AACCTCAAGGACAGAGAGGATATGGCCTGGAGGAGTCATCCCCTACGTCATTGGAGGGAACTTCACTG
GGAGCCAGAGGGCCATTTTTAAGCAGGCCATGAGACACTGGGAGAAGCACACCTGTGTGACCTTCATA
GAAAGGACGGATGAGGAAAGCTTTATTGTATTCAGTTACAGAACCTGTGGCTGTTGCTCCTATGTTGG
GCGCCGAGGAGGAGGCCCACAGGCCATATCCATTGGGAAGAACTGTGACAAGTTTGGCATTGTGGCTC
ACGAGCTGGGCCATGTGGTTGGGTTTTGGCATGAACACACCCGGCCAGACAGAGACCAACATGTCACC

ATCATCAGGGAAAACATCCAGCCAGGTCAGGAGTATAATTTCTTAAAAATGGAAGCTGGGGAAGTGAG
CTCTCTGGGAGAGACATACGACTTTGACAGCATCATGCACTACGCCCGGAACACCTTCTCAAGAGGAG
TTTTCTTAGACACCATCCTTCCCCGTCAAGATGACAATGGCGTCAGGCCAACCATTGGCCAGCGCGTG
CGGCTCAGTCAGGGAGACATAGCTCAAGCCCGGAAGCTGTACAAATGCCCAGCGTGTGGGGAGACCCT
GCAGGACACAACGGGAAACTTTTCTGCACCTGGTTTCCCAAATGGGTACCCATCTTACTCCCACTGCG
TCTGGAGGATCTCGGTCACCCCAGGGGAAAAGATCGTATTAAACTTCACATCCATGGATTTGTTTAAA
AGCCGACTGTGCTGGTATGATTACGTGGAGGTCCGGGATGGTTACTGGAGAAAAGCCCCCCTTTTGGG
CAGGTTTTGTGGCGATAAGATCCCGGAGCCCCTCGTCTCCACGGACAGCCGGCTCTGGGTGGAGTTCC
GCAGCAGCAGCAACATCTTGGGCAAGGGCTTCTTTGCAGCGTACGAAGCTACCTGCGGGGGAGACATG
AACAAAGATGCCGGTCAGATTCAATCTCCCAACTATCCGGATGACTACAGACCTTCCAAGGAATGTGT
CTGGAGGATTACGGTTTCGGAGGGGTTTCACGTGGGACTTACCTTCCAAGCTTTTGAGATTGAAAGGC
ACGACAGCTGTGCATATGACTACCTGGAAGTCCGGGATGGCCCCACGGAAGAGAGTGCCCTGATCGGC
CACTTTTGTGGCTATGAGAAGCCGGAGGATGTGAAATCGAGCTCCAACAGACTGTGGATGAAGTTTGT
GTCCGATGGCTCTATCAATAAAGCGGGCTTTGCAGCCAATTTTTTCAAGGAGGTGGATGAGTGTTCCT
GGCCAGATCACGGCGGGTGCGAACATCGCTGTGTGAACACGCTGGGCAGCTACAAGTGTGCCTGTGAC
CCTGGCTACGAGCTGGCCGCCGATAAGAAGATGTGTGAAGTGGCCTGTGGCGGTTTCATTACCAAGCT
GAATGGAACCATCACCAGCCCTGGGTGGCCGAAGGAGTATCCCACAAACAAAAACTGTGTCTGGCAGG
TGGTGGCCCCCCGCTCAGTACCGGATCTCCCTTCAGTTTGAAGTGTTTGAACTGGAAGGCAATGACGTC
TGTAAGTACGACTTTGTAGAGGTGCGCAGCGGCCTGTCCCCCGACGCCAAGCTGCACGGCAGGTTCTG
CGGCTCTGAGACGCCGGAAGTCATCACCTCGCAGAGCAACAACATGCGCGTGGAGTTCAAGTCCGACA
ACACCGTCTCCAAGCGCGGCTTCAGGGCCCACTTCTTCTCAGATAAGGACGAGTGTGCCAAGGACAAC
GGCGGGTGTCAGCATGAGTGCGTCAACACCTTCGGGAGCTACCTGTGCAGGTGCAGAAACGGCTACTG
GCTCCACGAGAATGGGCATGACTGCAAAGAGGCTGGCTGTGCACACAAGATCAGCAGTGTGGAGGGGA
CCCTGGCGAGCCCCAACTGGCCTGACAAATACCCCAGCCGGAGGGAGTGTACCTGGAACATCTCTTCG
ACTGCAGGCCACAGAGTGAAACTCACCTTTAATGAGTTTGAGATCGAGCAGCACCAGGAATGTGCCTA
TGACCACCTGGAAATGTATGACGGGCCGGACAGCCTGGCCCCCATTCTGGGCCGTTTCTGCGGCAGCA
AGAAACCAGACCCCACGGTGGCTTCCGGCAGCAGTATGTTTCTCAGGTTTTATTCGGATGCCTCAGTG
CAGAGGAAAGGCTTCCAGGCAGTGCACAGCACAGAGTGCGGGGGCAGGCTGAAGGCTGAAGTGCAGAC
CAAAGAGCTCTATTCCCACGCCCAGTTTGGGGACAACAACTACCCGAGCGAGGCCCGCTGTGACTGGG
TGATCGTGGCAGAGGACGGCTACGGCGTGGAGCTGACATTCCGGACCTTTGAGGTTGAGGAGGAGGCC
GACTGCGGCTACGACTACATGGAAGCCTACGACGGCTACGACAGCTCAGCGCCCAGGCTCGGCCGCTT
CTGTGGCTCTGGGCCATTAGAAGAAATCTACTCTGCAGGTGATTCCCTGATGATTCGATTCCGCACAG

ATGACACCATCAACAAGAAAGGCTTTCATGCCCGATACACCAGCACCAAGTTCCAGGATGCCCTGCAC
ATGAAGAAATAGTGCTGATGTTCTTGAAAGACAGAAACTGAGAATTTTTTTGTTTTGTTTTGTTTTTA
ACAACAATAGCACCTTGAAAATCTGCCCTAAAACAGTGTACAGTATTTTTCTCAAACAAAAACTCAGA
ATCCAGCCTTAGAGGTATATATTTGAATGAAAGTCTTGTAAGTTTGGCCAACAAGGTGGAGAAAAAAA
TGTTCTTTTGCTTCTGTCTGCAATGTTGTCATTCATGAACTGTTAAAGTGTTAAAGATTAGGATTGGA
GTCACTGACCATTCCGGCTATGCTTCTTCATACCATTCTCCTTGTTGTCCCTTGCTCCTATGTGGCAA
AAGGTCAGCCTTGGGGTTGGCCGTTCCTCTAATCTGGACTTGCTTGCAAAGGTGCCAGGCTGTCTTCT
GTCCATGTTGGGCATAAGGGATGAAAACTTGGCCGAGACTAATGTGTGGCCCACAGCTTTGGCTGGAA
TCATTTTCTTTCTCTCTGCCAGGGACATGTCAACCAAGAAACCTGAAAATATGGATGGATGTCAGGAC
TAAAAAAAGGCATCACAGTGAGCAGTGAGCACAGAGGGAGTTTCGAGTATAAGAATCATTGTCATGAA
GTTAGGAGACCACAAAGCCATTTCTCAGAGTCATTCACTCTCCTTGTCCCTTTGGTTTCCCCCCTTCC
TTAATTGCAGTGGGGGCTAAGGTATCCATTATGAATACAGCAGAACATTTGCTGGCGAGAGTCCTGTC
TGCTGAGAAGACAATATTGTGGCTCGTCCTGATATTTTTTCATTCATTGACTTTGAGAAGACTCCACC
TGTGCTTGGAATTCCATGGGCTTCAAAGAACATTTCTTCTTTTAGCTTTGGAGGCACTTGCCGTGGCA
CACCTGGACTCCTTGACATCCAATTCAAACTGCATTTGCAAAATGTGCAAAGACCTCTTATGAGGGAC
CAATTCAGGTCCCTTATGGGGTGAACACTGTTGAAGACTGGTTAATTATAAGTTATGTAAGAATCATC
GCCTTGTGGAACAAGTCAATCAGTGACTAGCTTCCTGTAGCCAATCAGGTTAAAGAGGGCGTTGGTAA
TTTTGTTCTGATTTAACTAGTATTCAATCACCAACTTGCAAACAGAATTCATAACACTTGGCACTTGT
TCTAGAGAAGTGTAGAGGATGATGTTAACATAATTTTAGCACTTCAAGGTATAATTTAAACAGTGAGG
TAGTTTTGAATGGCATTTCATTAAGGCATCTATGGGCATTATGAGCTAAAAGCTGTGGTATGTTAGCT
TTAAAAGAGTATTTATGTTGGAATAATTTTTAAATAATGTTTACATAACTGTAAGTCCTGTTTGGTTG
TTGTTGGACGCAGGGCGGCACATGAGTGTTTTTGGTTAGAGCCAAGATAGCTCCCATGCACCGGAATT
CCTTTGGGATGAATCAGCATCATTTTAAACAAAGTATATGTAAAAGGTGAAAGGTTATATTTTTTACA
GATCAGAATGTGGCACCAGAGGACTGTGTCTCATTAAAGTGATTGCTGGGAGCAAAAACTAGAATGAT
ACAAAGAAAGGTCAGAGAAATGCATGGGAATATTTTTTCTTT

NOV8c, CG50235-02 Protein Sequence	SEQ ID NO: 90	1015 aa	MW at 113555.5kD

MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAVFWGDIALD
EDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAGKDGRENTTLLHSPGTL
HAAAKTFSPRVRRATTSRTERIWPGGVIPYVIGGNFTGSQRAIFKQAMRHWEKHTCVTFIERTDEESF
IVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIVAHELGHVVGFWHEHTRPDRDQHVTIIRENIQP
GQEYNFLKMEAGEVSSLGETYDFDSIMHYARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIA
QARKLYKCPACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIVLNFTSMDLFKSRLCWYDY
VEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSNILGKGFFAAYEATCGGDMNKDAGQIQ
SPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFEIERHDSCAYDYLEVRDGPTEESALIGHFCGYEKP
EDVKSSSNRLWMKFVSDGSINKAGFAANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAAD
KKMCEVACGGFITKLNGTITSPGWPKEYPTNKNCVWQVVAPAQYRISLQFEVFELEGNDVCKYDFVEV
RSGLSPDAKLHGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECV
NTFGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTAGHRVKL
TFNEFEIEQHQECAYDHLEMYDGPDSLAPILGRFCGSKKPDPTVASGSSMFLRFYSDASVQRKGFQAV
HSTECGGRLKAEVQTKELYSHAQFGDNNYPSEARCDWVIVAEDGYGVELTFRTFEVEEEADCGYDYME
AYDGYDSSAPRLGRFCGSGPLEEIYSAGDSLMIRFRTDDTINKKGFHARYTSTKFQDALHMKK

NOV8d, CG50235-03 DNA Sequence	SEQ ID NO: 91	3146 bp
	ORF Start: ATG at 227	ORF Stop: TAG at 3137

GCAGCCTCGGCCGCCGGGCAAGTAGCTCCGAGCGGCTGCTTCCCGGTTGCCTCGACGAAGACAGGGGG
CGCCGCGCTCCGCTTGCTCCGCGCCTGAGCCATGCCCAGCAGCCCTGTGTAACCACCGAGTCCCGGCC
GGAGCCGACCGACCCAGTGTGCGCCGTCTTTCGGCCGAGCTGAGCTTTCGTGCACGCAACTCCCTCTG
CCCCAGCCGGCCCCGCGCCACC**ATG**CCCCGGGCGACTGCACTTGGGGCCCTGGTGTCACTGCTGCTGC
TGCTGCCGCTGCCTCGCGGCGCCGGGGGACTCGGGGAGCGCCCGGACGCCACCGCAGACTACTCAGAG
CTGGACGGCGAGGAGGGCACGGAGCAGCAGCTGGAGCATTACCACGACCCTTGCAAAGCCGCTGTCTT
TTGGGGAGACATTGCCTTAGATGAAGATGACTTGAAGCTGTTTCACATTGACAAAGCCAGAGACTGGA
CCAAGCAGACAGTGGGGGCAACAGGACACAGCACAGGTGGGCTTGAAGAGCAGGCATCTGAGAGCAGC
CCAGACACCACAGCCATGGACACTGGCACCAAGGAAGCTGGAAAGGGGAGCCAGAGGGCCATTTTTAA
GCAGGCCATGAGACACTGGGAGAAGCACACCTGTGTGACCTTCATAGAAAGGACGGATGAGGAAAGCT
TTATTGTATTCAGTTACAGAACCTGTGGCTGTTGCTCCTATGTTGGGCGCCGAGGAGGAGGCCCACAG
GCCATATCCATTGGGAAGAACTGTGACAAGTTTGGCATTGTGGCTCACGAGCTGGGCCATGTGGTTGG
GTTTTGGCATGAACACACCCGGCCAGACAGAGACCAACATGTCACCATCATCAGGGAAAACATCCAGC
CAGGTCAGGAGTATAATTTCTTAAAAATGGAAGCTGGGGAAGTGAGCTCTCTGGGAGAGACATACGAC
TTTGACAGCATCATGCACTACGCCCGGAACACCTTCTCAAGAGGAGTTTTTTTAGACACCATCCTTCC
CCGTCAAGATGACAATGGCGTCAGGCCAACCATTGGCCAGCGCGTGCGGCTCAGTCAGGGAGACATAG
CTCAAGCCCGGAAGCTGTACAAATGCCCAGGTCCTACTTGTGCTTTTGTTAGCCAGAAAACATCAATC
TGCTTGCTACACTTCTCACCAACCTGTTCCGAGGGCTTTGGCTGGCAAAGGGCGTGTGGGGAGACCCT
GCAGGACACAACGGGAAACTTTTCTGCACCTGGTTTCCCAAATGGGTACCCATCTTACTCCCACTGCG
TCTGGAGGATCTCGGTCACCCCAGGGGAAAAGATCGTATTAAACTTCACATCCATGGATTTGTTTAAA
AGCCGACTGTGCTGGTATGATTACGTGGAGGTCCGGGATGGTTACTGGAGAAAAGCCCCCCTTTTGGG
CAGGTTTTGTGGCGATAAGATCCCGGAGCCCCTCGTCTCCACGGACAGCCGGCTCTGGGTGGAGTTCC
GCAGCAGCAGCAACATCTTGGGCAAGGGCTTCTTTGCAGCGTACGAAGCTACCTGCGGGGGAGACATG
AACAAAGATGCCGGTCAGATTCAATCTCCCAACTATCCGGATGACTACAGACCTTCCAAGGAATGTGT
CTGGAGGATTACGGTTTCAGAGGGGTTTCACGTGGGACTTACCTTCCAAGCTTTTGAGATTGAAAGGC
ACGACAGCTGTGCATATGACTACCTGGAAGTCCGGGATGGCCCCACGGAAGAGAGTGCCCTGATCGGC
CACTTTTGTGGCTATGAGAAGCCGGAGGATGTGAAATCGAGCTCCAACAGACTGTGGATGAAGTTTGT
GTCCGATGGCTCTATCAATAAAGCGGGCTTTGCAGCCAATTTTTTCAAGGAGGTGGATGAGTGTTCCT
GGCCAGATCACGGCGGGTGCGAGCATCGCTGTGTGAACACGCTGGGCAGCTACAAGTGTGCCTGTGAC
CCTGGCTACGAGCTGGCCGCCGATAAGAAGATGTGTGAAGTGGCCTGTGGCGGTTTCATTACCAAGCT
GAATGGAACCATCACCAGCCCTGGGTGGCCGAAGGAGTATCCCACAAACAAAAACTGTGTCTGGCAGG
TGGTGGCCCCCACTCAGTACCGGATCTCCCTTCAGTTTGAAGTGTTTGAACTGGAAGGCAATGACGTC
TGTAAGTACGACTTTGTAGAGGTGCGCAGCGGCCTGTCCCCCGACGCCAAGCTGCACGGCAGGTTCTG
CGGCTCTGAGACGCCGGAGGTCATCACCTCGCAGAGCAACAACATGCGCGTGGAGTTCAAGTCCGACA
ACACCGTCTCCAAGCGCGGCTTCAGGGCCCACTTCTTCTCAGATAAGGACGAGTGTGCCAAGGACAAC
GGCGGGTGTCAGCATGAGTGCGTCAACACCTTCGGGAGCTACCTGTGCAGGTGCAGAAACGGCTACTG

GCTCCACGAGAATGGGCATGACTGCAAAGAGGCTGGCTGTGCACACAAGATCAGCAGTGTGGAGGGGA
CCCTGGCGAGCCCCAACTGGCCTGACAAATACCCCAGCCGGAGGGAGTGTACCTGGAACATCTCTTCG
ACTGCAGGCCACAGAGTGAAACTCACCTTTAATGAGTTTGAGATCGAGCAGCACCAGGAATGTGCCTA
TGACCACCTGGAAATGTATGACGGGCCGGACAGCCTGGCCCCCATTCTGGGCCGTTTCTGCGGTAGCA
AGAAACCAGACCCCACGGTGGCTTCCGGCAGCAAGTGCGGGGGCAGGCTGAAGGCTGAAGTGCAGACC
AAAGAGCTCTATTCCCACGCCCAGTTTGGGGACAACAACTACCCGAGCGAGGCCCGCTGTGACTGGGT
GATCGTGGCAGAGGACGGCTACGGCGTGGAGCTGACATTCCGGACCTTTGAGGTTGAGGAGGAGGCCG
ACTGCGGCTACGACTACATGGAAGCCTACGACGGCTACGACAGCTCAGCGCCCAGGCTCGGCCGCTTC
TGTGGCTCTGGGCCATTAGAAGAAATCTACTCTGCAGGTGATTCCCTGATGATTCGATTCCGCACAGA
TGACACCATCAACAAGAAAGGCTTTCATGCCCGATACACCAGCACCAAGTTCCAGGATGCCCTGCACA
TGAAGAAATAGTGCTGAT

NOV8d, CG50235-03 Protein Sequence	SEQ ID NO: 92	970 aa	MW at 108564.0kD

MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAVFWGDIALD
EDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAGKGSQRAIFKQAMRHWE
KHTCVTFIERTDEESFIVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIVAHELGHVVGFWHEHTR
PDRDQHVTIIRENIQPGQEYNFLKMEAGEVSSLGETYDFDSIMHYARNTFSRGVFLDTILPRQDDNGV
RPTIGQRVRLSQGDIAQARKLYKCPGPTCAFVSQKTSICLLHFSPTCSEGFGWQRACGETLQDTTGNF
SAPGFPNGYPSYSHCVWRISVTPGEKIVLNFTSMDLFKSRLCWYDYVEVRDGYWRKAPLLGRFCGDKI
PEPLVSTDSRLWVEFRSSSNILGKGFFAAYEATCGGDMNKDAGQIQSPNYPDDYRPSKECVWRITVSE
GFHVGLTFQAFEIERHDSCAYDYLEVRDGPTEESALIGHFCGYEKPEDVKSSSNRLWMKFVSDGSINK
AGFAANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAADKKMCEVACGGFITKLNGTITSP
GWPKEYPTNKNCVWQVVAPTQYRISLQFEVFELEGNDVCKYDFVEVRSGLSPDAKLHGRFCGSETPEV
ITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECVNTFGSYLCRCRNGYWLHENGHD
CKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTAGHRVKLTFNEFEIEQHQECAYDHLEMYD
GPDSLAPILGRFCGSKKPDPTVASGSKCGGRLKAEVQTKELYSHAQFGDNNYPSEARCDWVIVAEDGY
GVELTFRTFEVEEEADCGYDYMEAYDGYDSSAPRLGRFCGSGPLEEIYSAGDSLMIRFRTDDTINKKG
FHARYTSTKFQDALHMKK

NOV8e, SNP13377383 of CG50235-04, DNA Sequence	SEQ ID NO: 93	3120 bp
	ORF Start: ATG at 256	ORF Stop: TGA at 2719
	SNP Pos: 2201	SNP Change: A to G

TAACCTGTCTCACGACGGTCTAAACCCAGGCAGCCTCGGCCGCCGGGCAAGTAGCTCCGAGCGGCTGC
TTCCCGGTTGCCTCGAAGAAGACAGGGGGCGCCGCGCTCCGCTTGCTCCGCGCCTGAGCCATGCCCAG
CAGCCCTGTGTAACCACCGAGTCCCGGCCGGAGCCGACCGACCCAGTGTGCGCCGTCTTTCGGCCGAG
CTGAGCTTTCGTGCACGCAACTCCCTCTGCCCCAGCCGGCCCCGCGCCACCATGCCCCGGGCGACTGC
ACTTGGGGCCCTGGTGTCACTGCTGCTGCTGCTGCCGCTGCCTCGCGGCGCCGGGGGACTCGGGGAGC
GCCCGGACGCCACCGCAGACTACTCAGAGCTGGACGGCGAGGAGGGCACGGAGCAGCAGCTGGAGCAT
TACCACGACCCTTGCAAAGCCGCTGTCTTTTGGGGAGACATTGCCTTAGATGAAGATGACTTGAAGCT
GTTTCACATTGACAAAGCCAGAGACTGGACCAAGCAGACAGTGGGGGCAACAGGACACAGCACAGGTG
GGCTTGAAGAGCAGGCATCTGAGAGCAGCCCAGACACCACAGCCATGGACACTGGCACCAAGGAAGCT
GGAAAGGATGGCCGGGAGAATACCACACTCCTGCACAGCCCTGGGACCTTGCATGCCGCAGCCAAGAC
CTTCTCTCCCCGGGTCCGAAGAGCCACAACCTCAAGGACAGAGAGGATATGGCCTGGAGGAGTCATCC
CCTACGTCATTGGAGGGAACTTCACTGGGAGCCAGAGGGCCATTTTTAAGCAGGCCATGAGACACTGG
GAGAAGCACACCTGTGTGACCTTCATAGAAAGGACGGATGAGGAAAGCTTTATTGTATTCAGTTACAG
AACCTGTGGCTGTTGCTCCTATGTTGGGCGCCGAGGAGGAGGCCCACAGGCCATATCCATTGGGAAGA
ACTGTGACAAGTTTGGCATTGTGGCTCACGAGCTGGGCCATGTGGTTGGGTTTTGGCATGAACACACC
CGGCCAGACAGAGACCAACATGTCACCATCATCAGGGAAAACATCCAGCCAGGTCAGGAGTATAATTT
CTTAAAAATGGAAGCTGGGGAAGTGAGCTCTCTGGGAGAGACATACGACTTTGACAGCATCATGCACT
ACGCCCGGAACACCTTCTCAAGAGGAGTTTTCTTAGACACCATCCTTCCCCGTCAAGATGACAATGGC
GTCAGGCCAACCATTGGCCAGCGCGTGCGGCTCAGTCAGGGAGACATAGCTCAAGCCCGGAAGCTGTA
CAAATGCCCAGCGTGTGGGGAGACCCTGCAGGACACAACGGGAAACTTTTCTGCACCTGGTTTCCCAA
ATGGGTACCCATCTTACTCCCACTGCGTCTGGAGGATCTCGGTCACCCCAGGGGAAAAGATCGTATTA
AACTTCACATCCATGGATTTGTTTAAAAGCCGACTGTGCTGGTATGATTACGTGGAGGTCCGGGATGG
TTACTGGAGAAAAGCCCCCCTTTTGGGCAGGTTTTGTGGCGATAAGATCCCGGAGCCCCTCGTCTCCA
CGGACAGCCGGCTCTGGGTGGAGTTCCGCAGCAGCAGCAACATCTTGGGCAAGGGCTTCTTTGCAGCG
TACGAAGCTACCTGCGGGGGAGACATGAACAAAGATGCCGGTCAGATTCAATCTCCCAACTATCCGGA

TGACTACAGACCTTCCAAGGAATGTGTCTGGAGGATTACGGTTTCGGAGGGGTTTCACGTGGGACTTA
CCTTCCAAGCTTTTGAGATTGAAAGGCACGACAGCTGTGCATATGACTACCTGGAAGTCCGGGATGGC
CCCACGGAAGAGAGTGCCCTGATCGGCCACTTTTGTGGCTATGAGAAGCCGGAGGATGTGAAATCGAG
CTCCAACAGACTGTGGATGAAGTTTGTGTCCGATGGCTCTATCAATAAAGCGGGCTTTGCAGCCAATT
TTTTCAAGGAGGTGGATGAGTGTTCCTGGCCAGATCACGGCGGGTGCGAACATCGCTGTGTGAACACG
CTGGGCAGCTACAAGTGTGCCTGTGACCCTGGCTACGAGCTGGCCGCCGATAAGAAGATGTGTGAAGT
GGCCTGTGGCGGTTTCATTACCAAGCTGAATGGAACCATCACCAGCCCTGGGTGGCCGAAGGAGTATC
CCACAAACAAAAACTGTGTCTGGCGGGTGGTGGCCCCCGCTCAGTACCGGATCTCCCTTCAGTTTGAA
GTGTTTGAACTGGAAGGCAATGACGTCTGTAAGTACGACTTTGTAGAGGTGCGCAGCGGCCTGTCCCC
CGACGCCAAGCTGCACGGCAGGTTCTGCGGCTCTGAGACGCCGGAAGTCATCACCTCGCAGAGCAACA
ACATGCGCGTGGAGTTCAAGTCCGACAACACCGTCTCCAAGCGCGGCTTCAGGGCCCACTTCTTCTCA
GATAAGGACGAGTGTGCCAAGGACAACGGCGGGTGTCAGCATGAGTGCGTCAACACCTTCGGGAGCTA
CCTGTGCAGGTGCAGAAACGGCTACTGGCTCCACGAGAATGGGCATGACTGCAAAGAGGCTGGCTGTG
CACACAAGATCAGCAGTGTGGAGGGGACCCTGGCGAGCCCCAACTGGCCTGACAAATACCCCAGCCGG
AGGGAGTGTACCTGGAACATCTCTTCGACTGCAGGCCACAGAGTAAAACTCAGTGCGGGGTCAGGCTG
AAGGCTGAAGTGCAGACCAAAGAGCTCTATTCCCACGCCCAGTTTGGGGACAACAACTACCCGAGCGA
GGCCCGCTGTGACTGGGTGATCGTGGCAGAGGACGGCTACGGCGTGGAGCTGACATTCCGGACCTTTG
AGGTTGAGGAGGAGGCCGACTGCGGCTACGACTACATGGAAGCCTACGACGGCTACGACAGCTCAGCG
CCCAGGCTCGGCCGCTTCTGTGGCTCTGGGCCATTAGAAGAAATCTACTCTGCAGGTGATTCCCTGAT
GATTCGATTCCGCACAGATGACACCATCAACAAGAAAGGCTTTCATGCCCGATACACCAGCACCAAGT
TCCAGGATGCCCTGCACATGAAGAAATAGTGCTGATGTTCTTGAAAGACAGAAACTGAGA

NOV8e, SNP13377383 of CG50235-04, Protein Sequence	SEQ ID NO: 94	821 aa	MW at 91432.3kD
	SNP Pos: 649		SNP Change: Gln to Arg

MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAVFWGDIALD
EDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAGKDGRENTTLLHSPGTL
HAAAKTFSPRVRRATTSRTERIWPGGVIPYVIGGNFTGSQRAIFKQAMRHWEKHTCVTFIERTDEESF
IVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIVAHELGHVVGFWHEHTRPDRDQHVTIIRENIQP
GQEYNFLKMEAGEVSSLGETYDFDSIMHYARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIA
QARKLYKCPACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIVLNFTSMDLFKSRLCWYDY
VEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSNILGKGFFAAYEATCGGDMNKDAGQIQ
SPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFEIERHDSCAYDYLEVRDGPTEESALIGHFCGYEKP
EDVKSSSNRLWMKFVSDGSINKAGFAANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAAD
KKMCEVACGGFITKLNGTITSPGWPKEYPTNKNCVWRVVAPAQYRISLQFEVFELEGNDVCKYDFVEV
RSGLSPDAKLHGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECV
NTFGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTAGHRVKL
SAGSG

NOV8f, SNP13377384 of CG50235-04, DNA Sequence	SEQ ID NO: 95	3120 bp
	ORF Start: ATG at 256	ORF Stop: TGA at 2719
	SNP Pos: 2434	SNP Change: G to A

TAACCTGTCTCACGACGGTCTAAACCCAGGCAGCCTCGGCCGCCGGGCAAGTAGCTCCGAGCGGCTGC
TTCCCGGTTGCCTCGAAGAAGACAGGGGGCGCCGCGCTCCGCTTGCTCCGCGCCTGAGCCATGCCCAG
CAGCCCTGTGTAACCACCGAGTCCCGGCCGGAGCCGACCGACCCAGTGTGCGCCGTCTTTCGGCCGAG
CTGAGCTTTCGTGCACGCAACTCCCTCTGCCCCAGCCGGCCCCGCGCCACCATGCCCCGGGCGACTGC
ACTTGGGGCCCTGGTGTCACTGCTGCTGCTGCTGCCGCTGCCTCGCGGCGCCGGGGGACTCGGGGAGC
GCCCGGACGCCACCGCAGACTACTCAGAGCTGGACGGCGAGGAGGGCACGGAGCAGCAGCTGGAGCAT
TACCACGACCCTTGCAAAGCCGCTGTCTTTTGGGGAGACATTGCCTTAGATGAAGATGACTTGAAGCT
GTTTCACATTGACAAAGCCAGAGACTGGACCAAGCAGACAGTGGGGGCAACAGGACACAGCACAGGTG
GGCTTGAAGAGCAGGCATCTGAGAGCAGCCCAGACACCACAGCCATGGACACTGGCACCAAGGAAGCT
GGAAAGGATGGCCGGGAGAATACCACACTCCTGCACAGCCCTGGGACCTTGCATGCCGCAGCCAAGAC
CTTCTCTCCCCGGGTCCGAAGAGCCACAACCTCAAGGACAGAGAGGATATGGCCTGGAGGAGTCATCC
CCTACGTCATTGGAGGGAACTTCACTGGGAGCCAGAGGGCCATTTTTAAGCAGGCCATGAGACACTGG
GAGAAGCACACCTGTGTGACCTTCATAGAAAGGACGGATGAGGAAAGCTTTATTGTATTCAGTTACAG
AACCTGTGGCTGTTGCTCCTATGTTGGGCGCCGAGGAGGAGGCCCACAGGCCATATCCATTGGGAAGA
ACTGTGACAAGTTTGGCATTGTGGCTCACGAGCTGGGCCATGTGGTTGGGTTTTGGCATGAACACACC
CGGCCAGACAGAGACCAACATGTCACCATCATCAGGGAAAACATCCAGCCAGGTCAGGAGTATAATTT
CTTAAAAATGGAAGCTGGGGAAGTGAGCTCTCTGGGAGAGACATACGACTTTGACAGCATCATGCACT

ACGCCCGGAACACCTTCTCAAGAGGAGTTTTCTTAGACACCATCCTTCCCCGTCAAGATGACAATGGC
GTCAGGCCAACCATTGGCCAGCGCGTGCGGCTCAGTCAGGGAGACATAGCTCAAGCCCGGAAGCTGTA
CAAATGCCCAGCGTGTGGGGAGACCCTGCAGGACACAACGGGAAACTTTTCTGCACCTGGTTTCCCAA
ATGGGTACCCATCTTACTCCCACTGCGTCTGGAGGATCTCGGTCACCCCAGGGGAAAAGATCGTATTA
AACTTCACATCCATGGATTTGTTTAAAAGCCGACTGTGCTGGTATGATTACGTGGAGGTCCGGGATGG
TTACTGGAGAAAAGCCCCCCTTTTGGGCAGGTTTTGTGGCGATAAGATCCCGGAGCCCCTCGTCTCCA
CGGACAGCCGGCTCTGGGTGGAGTTCCGCAGCAGCAGCAACATCTTGGGCAAGGGCTTCTTTGCAGCG
TACGAAGCTACCTGCGGGGGAGACATGAACAAAGATGCCGGTCAGATTCAATCTCCCAACTATCCGGA
TGACTACAGACCTTCCAAGGAATGTGTCTGGAGGATTACGGTTTCGGAGGGGTTTCACGTGGGACTTA
CCTTCCAAGCTTTTGAGATTGAAAGGCACGACAGCTGTGCATATGACTACCTGGAAGTCCGGGATGGC
CCCACGGAAGAGAGTGCCCTGATCGGCCACTTTTGTGGCTATGAGAAGCCGGAGGATGTGAAATCGAG
CTCCAACAGACTGTGGATGAAGTTTGTGTCCGATGGCTCTATCAATAAAGCGGGCTTTGCAGCCAATT
TTTTCAAGGAGGTGGATGAGTGTTCCTGGCCAGATCACGGCGGGTGCGAACATCGCTGTGTGAACACG
CTGGGCAGCTACAAGTGTGCCTGTGACCCTGGCTACGAGCTGGCCGCCGATAAGAAGATGTGTGAAGT
GGCCTGTGGCGGTTTCATTACCAAGCTGAATGGAACCATCACCAGCCCTGGGTGGCCGAAGGAGTATC
CCACAAACAAAAACTGTGTCTGGCAGGTGGTGGCCCCCGCTCAGTACCGGATCTCCCTTCAGTTTGAA
GTGTTTGAACTGGAAGGCAATGACGTCTGTAAGTACGACTTTGTAGAGGTGCGCAGCGGCCTGTCCCC
CGACGCCAAGCTGCACGGCAGGTTCTGCGGCTCTGAGACGCCGGAAGTCATCACCTCGCAGAGCAACA
ACATGCGCGTGGAGTTCAAGTCCGACAACACCGTCTCCAAGCGCGGCTTCAGGACCCACTTCTTCTCA
GATAAGGACGAGTGTGCCAAGGACAACGGCGGGTGTCAGCATGAGTGCGTCAACACCTTCGGGAGCTA
CCTGTGCAGGTGCAGAAACGGCTACTGGCTCCACGAGAATGGGCATGACTGCAAAGAGGCTGGCTGTG
CACACAAGATCAGCAGTGTGGAGGGGACCCTGGCGAGCCCCAACTGGCCTGACAAATACCCCAGCCGG
AGGGAGTGTACCTGGAACATCTCTTCGACTGCAGGCCACAGAGTAAAACTCAGTGCGGGGTCAGGCTG
AAGGCTGAAGTGCAGACCAAAGAGCTCTATTCCCACGCCCAGTTTGGGGACAACAACTACCCGAGCGA
GGCCCGCTGTGACTGGGTGATCGTGGCAGAGGACGGCTACGGCGTGGAGCTGACATTCCGGACCTTTG
AGGTTGAGGAGGAGGCCGACTGCGGCTACGACTACATGGAAGCCTACGACGGCTACGACAGCTCAGCG
CCCAGGCTCGGCCGCTTCTGTGGCTCTGGGCCATTAGAAGAAATCTACTCTGCAGGTGATTCCCTGAT
GATTCGATTCCGCACAGATGACACCATCAACAAGAAAGGCTTTCATGCCCGATACACCAGCACCAAGT
TCCAGGATGCCCTGCACATGAAGAAATAGTGCTGATGTTCTTGAAAGACAGAAACTGAGA

NOV8f, SNP13377384 of CG50235-04, Protein Sequence	SEQ ID NO: 96	821 aa	MW at 91434.2kD
	SNP Pos: 727		SNP Change: Ala to Thr

MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAVFWGDIALD
EDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAGKDGRENTTLLHSPGTL
HAAAKTFSPRVRRATTSRTERIWPGGVIPYVIGGNFTGSQRAIFKQAMRHWEKHTCVTFIERTDEESF
IVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIVAHELGHVVGFWHEHTRPDRDQHVTIIRENIQP
GQEYNFLKMEAGEVSSLGETYDFDSIMHYARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIA
QARKLYKCPACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIVLNFTSMDLFKSRLCWYDY
VEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSNILGKGFFAAYEATCGGDMNKDAGQIQ
SPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFEIERHDSCAYDYLEVRDGPTEESALIGHFCGYEKP
EDVKSSSNRLWMKFVSDGSINKAGFAANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAAD
KKMCEVACGGFITKLNGTITSPGWPKEYPTNKNCVWQVVAPAQYRISLQFEVFELEGNDVCKYDFVEV
RSGLSPDAKLHGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRTHFFSDKDECAKDNGGCQHECV
NTFGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTAGHRVKL
SAGSG

NOV8g, SNP13377385 of CG50235-04, DNA Sequence	SEQ ID NO: 97	3120 bp
	ORF Start: ATG at 256	ORF Stop: TGA at 2719
	SNP Pos: 2751	SNP Change: T to C

TAACCTGTCTCACGACGGTCTAAACCCAGGCAGCCTCGGCCGCCGGGCAAGTAGCTCCGAGCGGCTGC
TTCCCGGTTGCCTCGAAGAAGACAGGGGGCGCCGCGCTCCGCTTGCTCCGCGCCTGAGCCATGCCCAG
CAGCCCTGTGTAACCACCGAGTCCCGGCCGGAGCCGACCGACCCAGTGTGCGCCGTCTTTCGGCCGAG
CTGAGCTTTCGTGCACGCAACTCCCTCTGCCCCAGCCGGCCCCGCGCCACCATGCCCCGGGCGACTGC
ACTTGGGGCCCTGGTGTCACTGCTGCTGCTGCTGCCGCTGCCTCGCGGCGCCGGGGGACTCGGGGAGC
GCCCGGACGCCACCGCAGACTACTCAGAGCTGGACGGCGAGGAGGGCACGGAGCAGCAGCTGGAGCAT
TACCACGACCCTTGCAAAGCCGCTGTCTTTTGGGGAGACATTGCCTTAGATGAAGATGACTTGAAGCT
GTTTCACATTGACAAAGCCAGAGACTGGACCAAGCAGACAGTGGGGGCAACAGGACACAGCACAGGTG
GGCTTGAAGAGCAGGCATCTGAGAGCAGCCCAGACACCACAGCCATGGACACTGGCACCAAGGAAGCT

GGAAAGGATGGCCGGGAGAATACCACACTCCTGCACAGCCCTGGGACCTTGCATGCCGCAGCCAAGAC
CTTCTCTCCCCGGGTCCGAAGAGCCACAACCTCAAGGACAGAGAGGATATGGCCTGGAGGAGTCATCC
CCTACGTCATTGGAGGGAACTTCACTGGGAGCCAGAGGGCCATTTTTAAGCAGGCCATGAGACACTGG
GAGAAGCACACCTGTGTGACCTTCATAGAAAGGACGGATGAGGAAAGCTTTATTGTATTCAGTTACAG
AACCTGTGGCTGTTGCTCCTATGTTGGGCGCCGAGGAGGAGGCCCACAGGCCATATCCATTGGGAAGA
ACTGTGACAAGTTTGGCATTGTGGCTCACGAGCTGGGCCATGTGGTTGGGTTTTGGCATGAACACACC
CGGCCAGACAGAGACCAACATGTCACCATCATCAGGGAAAACATCCAGCCAGGTCAGGAGTATAATTT
CTTAAAAATGGAAGCTGGGGAAGTGAGCTCTCTGGGAGAGACATACGACTTTGACAGCATCATGCACT
ACGCCCGGAACACCTTCTCAAGAGGAGTTTTCTTAGACACCATCCTTCCCCGTCAAGATGACAATGGC
GTCAGGCCAACCATTGGCCAGCGCGTGCGGCTCAGTCAGGGAGACATAGCTCAAGCCCGGAAGCTGTA
CAAATGCCCAGCGTGTGGGGAGACCCTGCAGGACACAACGGGAAACTTTTCTGCACCTGGTTTCCCAA
ATGGGTACCCATCTTACTCCCACTGCGTCTGGAGGATCTCGGTCACCCCAGGGGAAAAGATCGTATTA
AACTTCACATCCATGGATTTGTTTAAAAGCCGACTGTGCTGGTATGATTACGTGGAGGTCCGGGATGG
TTACTGGAGAAAAGCCCCCCTTTTGGGCAGGTTTTGTGGCGATAAGATCCCGGAGCCCCTCGTCTCCA
CGGACAGCCGGCTCTGGGTGGAGTTCCGCAGCAGCAGCAACATCTTGGGCAAGGGCTTCTTTGCAGCG
TACGAAGCTACCTGCGGGGGAGACATGAACAAAGATGCCGGTCAGATTCAATCTCCCAACTATCCGGA
TGACTACAGACCTTCCAAGGAATGTGTCTGGAGGATTACGGTTTCGGAGGGGTTTCACGTGGGACTTA
CCTTCCAAGCTTTTGAGATTGAAAGGCACGACAGCTGTGCATATGACTACCTGGAAGTCCGGGATGGC
CCCACGGAAGAGAGTGCCCTGATCGGCCACTTTTGTGGCTATGAGAAGCCGGAGGATGTGAAATCGAG
CTCCAACAGACTGTGGATGAAGTTTGTGTCCGATGGCTCTATCAATAAAGCGGGCTTTGCAGCCAATT
TTTTCAAGGAGGTGGATGAGTGTTCCTGGCCAGATCACGGCGGGTGCGAACATCGCTGTGTGAACACG
CTGGGCAGCTACAAGTGTGCCTGTGACCCTGGCTACGAGCTGGCCGCCGATAAGAAGATGTGTGAAGT
GGCCTGTGGCGGTTTCATTACCAAGCTGAATGGAACCATCACCAGCCCTGGGTGGCCGAAGGAGTATC
CCACAAACAAAAACTGTGTCTGGCAGGTGGTGGCCCCCGCTCAGTACCGGATCTCCCTTCAGTTTGAA
GTGTTTGAACTGGAAGGCAATGACGTCTGTAAGTACGACTTTGTAGAGGTGCGCAGCGGCCTGTCCCC
CGACGCCAAGCTGCACGGCAGGTTCTGCGGCTCTGAGACGCCGGAAGTCATCACCTCGCAGAGCAACA
ACATGCGCGTGGAGTTCAAGTCCGACAACACCGTCTCCAAGCGCGGCTTCAGGGCCCACTTCTTCTCA
GATAAGGACGAGTGTGCCAAGGACAACGGCGGGTGTCAGCATGAGTGCGTCAACACCTTCGGGAGCTA
CCTGTGCAGGTGCAGAAACGGCTACTGGCTCCACGAGAATGGGCATGACTGCAAAGAGGCTGGCTGTG
CACACAAGATCAGCAGTGTGGAGGGGACCCTGGCGAGCCCCAACTGGCCTGACAAATACCCCAGCCGG
AGGGAGTGTACCTGGAACATCTCTTCGACTGCAGGCCACAGAGTAAAACTCAGTGCGGGGTCAGGC**TG**
AAGGCTGAAGTGCAGACCAAAGAGCTCTATCCCCACGCCCAGTTTGGGGACAACAACTACCCGAGCGA
GGCCCGCTGTGACTGGGTGATCGTGGCAGAGGACGGCTACGGCGTGGAGCTGACATTCCGGACCTTTG
AGGTTGAGGAGGAGGCCGACTGCGGCTACGACTACATGGAAGCCTACGACGGCTACGACAGCTCAGCG
CCCAGGCTCGGCCGCTTCTGTGGCTCTGGGCCATTAGAAGAAATCTACTCTGCAGGTGATTCCCTGAT
GATTCGATTCCGCACAGATGACACCATCAACAAGAAAGGCTTTCATGCCCGATACACCAGCACCAAGT
TCCAGGATGCCCTGCACATGAAGAAATAGTGCTGATGTTCTTGAAAGACAGAAACTGAGA

NOV8g, SNP13377385 of CG50235-04, Protein Sequence	SEQ ID NO: 98	821 aa	MW at 91404.2kD
			SNP Change: no change

MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAVFWGDIALD
EDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAGKDGRENTTLLHSPGTL
HAAAKTFSPRVRRATTSRTERIWPGGVIPYVIGGNFTGSQRAIFKQAMRHWEKHTCVTFIERTDEESF
IVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIVAHELGHVVGFWHEHTRPDRDQHVTIIRENIQP
GQEYNFLKMEAGEVSSLGETYDFDSIMHYARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIA
QARKLYKCPACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIVLNFTSMDLFKSRLCWYDY
VEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSNILGKGFFAAYEATCGGDMNKDAGQIQ
SPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFEIERHDSCAYDYLEVRDGPTEESALIGHFCGYEKP
EDVKSSSNRLWMKFVSDGSINKAGFAANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAAD
KKMCEVACGGFITKLNGTITSPGWPKEYPTNKNCVWQVVAPAQYRISLQFEVFELEGNDVCKYDFVEV
RSGLSPDAKLHGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECV
NTFGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTAGHRVKL
SAGSG

NOV8h, SNP13377386 of CG50235-04, DNA Sequence	SEQ ID NO: 99	3120 bp
	ORF Start: ATG at 256	ORF Stop: TGA at 2719
	SNP Pos: 2794	SNP Change: G to A

TAACCTGTCTCACGACGGTCTAAACCCAGGCAGCCTCGGCCGCCGGGCAAGTAGCTCCGAGCGGCTGC

TTCCCGGTTGCCTCGAAGAAGACAGGGGGCGCCGCGCTCCGCTTGCTCCGCGCCTGAGCCATGCCCAG
CAGCCCTGTGTAACCACCGAGTCCCGGCCGGAGCCGACCGACCCAGTGTGCGCCGTCTTTCGGCCGAG
CTGAGCTTTCGTGCACGCAACTCCCTCTGCCCCAGCCGGCCCCGCGCCACC**ATG**CCCCGGGCGACTGC
ACTTGGGGCCCTGGTGTCACTGCTGCTGCTGCTGCCGCTGCCTCGCGGCGCCGGGGGACTCGGGGAGC
GCCCGGACGCCACCGCAGACTACTCAGAGCTGGACGGCGAGGAGGGCACGGAGCAGCAGCTGGAGCAT
TACCACGACCCTTGCAAAGCCGCTGTCTTTTGGGGAGACATTGCCTTAGATGAAGATGACTTGAAGCT
GTTTCACATTGACAAAGCCAGAGACTGGACCAAGCAGACAGTGGGGGCAACAGGACACAGCACAGGTG
GGCTTGAAGAGCAGGCATCTGAGAGCAGCCCAGACACCACAGCCATGGACACTGGCACCAAGGAAGCT
GGAAAGGATGGCCGGGAGAATACCACACTCCTGCACAGCCCTGGGACCTTGCATGCCGCAGCCAAGAC
CTTCTCTCCCCGGGTCCGAAGAGCCACAACCTCAAGGACAGAGAGGATATGGCCTGGAGGAGTCATCC
CCTACGTCATTGGAGGGAACTTCACTGGGAGCCAGAGGGCCATTTTTAAGCAGGCCATGAGACACTGG
GAGAAGCACACCTGTGTGACCTTCATAGAAAGGACGGATGAGGAAAGCTTTATTGTATTCAGTTACAG
AACCTGTGGCTGTTGCTCCTATGTTGGGCGCCGAGGAGGAGGCCCACAGGCCATATCCATTGGGAAGA
ACTGTGACAAGTTTGGCATTGTGGCTCACGAGCTGGGCCATGTGGTTGGGTTTTGGCATGAACACACC
CGGCCAGACAGAGACCAACATGTCACCATCATCAGGGAAAACATCCAGCCAGGTCAGGAGTATAATTT
CTTAAAAATGGAAGCTGGGGAAGTGAGCTCTCTGGGAGAGACATACGACTTTGACAGCATCATGCACT
ACGCCCGGAACACCTTCTCAAGAGGAGTTTTCTTAGACACCATCCTTCCCCGTCAAGATGACAATGGC
GTCAGGCCAACCATTGGCCAGCGCGTGCGGCTCAGTCAGGGAGACATAGCTCAAGCCCGGAAGCTGTA
CAAATGCCCAGCGTGTGGGGAGACCCTGCAGGACACAACGGGAAACTTTTCTGCACCTGGTTTCCCAA
ATGGGTACCCATCTTACTCCCACTGCGTCTGGAGGATCTCGGTCACCCCAGGGGAAAAGATCGTATTA
AACTTCACATCCATGGATTTGTTTAAAAGCCGACTGTGCTGGTATGATTACGTGGAGGTCCGGGATGG
TTACTGGAGAAAAGCCCCCCTTTTGGGCAGGTTTTGTGGCGATAAGATCCCGGAGCCCCTCGTCTCCA
CGGACAGCCGGCTCTGGGTGGAGTTCCGCAGCAGCAGCAACATCTTGGGCAAGGGCTTCTTTGCAGCG
TACGAAGCTACCTGCGGGGGAGACATGAACAAAGATGCCGGTCAGATTCAATCTCCCAACTATCCGGA
TGACTACAGACCTTCCAAGGAATGTGTCTGGAGGATTACGGTTTCGGAGGGGTTTCACGTGGGACTTA
CCTTCCAAGCTTTTGAGATTGAAAGGCACGACAGCTGTGCATATGACTACCTGGAAGTCCGGGATGGC
CCCACGGAAGAGAGTGCCCTGATCGGCCACTTTTGTGGCTATGAGAAGCCGGAGGATGTGAAATCGAG
CTCCAACAGACTGTGGATGAAGTTTGTGTCCGATGGCTCTATCAATAAAGCGGGCTTTGCAGCCAATT
TTTTCAAGGAGGTGGATGAGTGTTCCTGGCCAGATCACGGCGGGTGCGAACATCGCTGTGTGAACACG
CTGGGCAGCTACAAGTGTGCCTGTGACCCTGGCTACGAGCTGGCCGCCGATAAGAAGATGTGTGAAGT
GGCCTGTGGCGGTTTCATTACCAAGCTGAATGGAACCATCACCAGCCCTGGGTGGCCGAAGGAGTATC
CCACAAACAAAAACTGTGTCTGGCAGGTGGTGGCCCCCGCTCAGTACCGGATCTCCCTTCAGTTTGAA
GTGTTTGAACTGGAAGGCAATGACGTCTGTAAGTACGACTTTGTAGAGGTGCGCAGCGGCCTGTCCCC
CGACGCCAAGCTGCACGGCAGGTTCTGCGGCTCTGAGACGCCGGAAGTCATCACCTCGCAGAGCAACA
ACATGCGCGTGGAGTTCAAGTCCGACAACACCGTCTCCAAGCGCGGCTTCAGGGCCCACTTCTTCTCA
GATAAGGACGAGTGTGCCAAGGACAACGGCGGGTGTCAGCATGAGTGCGTCAACACCTTCGGGAGCTA
CCTGTGCAGGTGCAGAAACGGCTACTGGCTCCACGAGAATGGGCATGACTGCAAAGAGGCTGGCTGTG
CACACAAGATCAGCAGTGTGGAGGGGACCCTGGCGAGCCCCAACTGGCCTGACAAATACCCCAGCCGG
AGGGAGTGTACCTGGAACATCTCTTCGACTGCAGGCCACAGAGTAAAACTCAGTGCGGGGTCAGGC**TG**
AAGGCTGAAGTGCAGACCAAAGAGCTCTATTCCCACGCCCAGTTTGGGGACAACAACTACCCGAGCGA
GGCCCACTGTGACTGGGTGATCGTGGCAGAGGACGGCTACGGCGTGGAGCTGACATTCCGGACCTTTG
AGGTTGAGGAGGAGGCCGACTGCGGCTACGACTACATGGAAGCCTACGACGGCTACGACAGCTCAGCG
CCCAGGCTCGGCCGCTTCTGTGGCTCTGGGCCATTAGAAGAAATCTACTCTGCAGGTGATTCCCTGAT
GATTCGATTCCGCACAGATGACACCATCAACAAGAAAGGCTTTCATGCCCGATACACCAGCACCAAGT
TCCAGGATGCCCTGCACATGAAGAAATAGTGCTGATGTTCTTGAAAGACAGAAACTGAGA

NOV8h, SNP13377386 of CG50235-04, Protein Sequence	SEQ ID NO: 100	821 aa	MW at 91404.2kD
			SNP Change: no change

MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAVFWGDIALD
EDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAGKDGRENTTLLHSPGTL
HAAAKTFSPRVRRATTSRTERIWPGGVIPYVIGGNFTGSQRAIFKQAMRHWEKHTCVTFIERTDEESF
IVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIVAHELGHVVGFWHEHTRPDRDQHVTIIRENIQP
GQEYNFLKMEAGEVSSLGETYDFDSIMHYARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIA
QARKLYKCPACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIVLNFTSMDLFKSRLCWYDY
VEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSNILGKGFFAAYEATCGGDMNKDAGQIQ
SPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFEIERHDSCAYDYLEVRDGPTEESALIGHFCGYEKP
EDVKSSSNRLWMKFVSDGSINKAGFAANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAAD
KKMCEVACGGFITKLNGTITSPGWPKEYPTNKNCVWQVVAPAQYRISLQFEVFELEGNDVCKYDFVEV
RSGLSPDAKLHGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECV

NTFGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTAGHRVKL SAGSG

NOV8i, SNP13377387 of CG50235-04, DNA Sequence	SEQ ID NO: 101	3120 bp
	ORF Start: ATG at 256	ORF Stop: TGA at 2719
	SNP Pos: 2983	SNP Change: A to G

TAACCTGTCTCACGACGGTCTAAACCCAGGCAGCCTCGGCCGCCGGGCAAGTAGCTCCGAGCGGCTGC
TTCCCGGTTGCCTCGAAGAAGACAGGGGGCGCCGCGCTCCGCTTGCTCCGCGCCTGAGCCATGCCCAG
CAGCCCTGTGTAACCACCGAGTCCCGGCCGGAGCCGACCGACCCAGTGTGCGCCGTCTTTCGGCCGAG
CTGAGCTTTCGTGCACGCAACTCCCTCTGCCCCAGCCGGCCCCGCGCCACCATGCCCCGGGCGACTGC
ACTTGGGGCCCTGGTGTCACTGCTGCTGCTGCTGCCGCTGCCTCGCGGCGCCGGGGGACTCGGGGAGC
GCCCGGACGCCACCGCAGACTACTCAGAGCTGGACGGCGAGGAGGGCACGGAGCAGCAGCTGGAGCAT
TACCACGACCCTTGCAAAGCCGCTGTCTTTTGGGGAGACATTGCCTTAGATGAAGATGACTTGAAGCT
GTTTCACATTGACAAAGCCAGAGACTGGACCAAGCAGACAGTGGGGGCAACAGGACACAGCACAGGTG
GGCTTGAAGAGCAGGCATCTGAGAGCAGCCCAGACACCACAGCCATGGACACTGGCACCAAGGAAGCT
GGAAAGGATGGCCGGGAGAATACCACACTCCTGCACAGCCCTGGGACCTTGCATGCCGCAGCCAAGAC
CTTCTCTCCCCGGGTCCGAAGAGCCACAACCTCAAGGACAGAGAGGATATGGCCTGGAGGAGTCATCC
CCTACGTCATTGGAGGGAACTTCACTGGGAGCCAGAGGGCCATTTTTAAGCAGGCCATGAGACACTGG
GAGAAGCACACCTGTGTGACCTTCATAGAAAGGACGGATGAGGAAAGCTTTATTGTATTCAGTTACAG
AACCTGTGGCTGTTGCTCCTATGTTGGGCGCCGAGGAGGAGGCCCACAGGCCATATCCATTGGGAAGA
ACTGTGACAAGTTTGGCATTGTGGCTCACGAGCTGGGCCATGTGGTTGGGTTTTGGCATGAACACACC
CGGCCAGACAGAGACCAACATGTCACCATCATCAGGGAAAACATCCAGCCAGGTCAGGAGTATAATTT
CTTAAAAATGGAAGCTGGGGAAGTGAGCTCTCTGGGAGAGACATACGACTTTGACAGCATCATGCACT
ACGCCCGGAACACCTTCTCAAGAGGAGTTTTCTTAGACACCATCCTTCCCCGTCAAGATGACAATGGC
GTCAGGCCAACCATTGGCCAGCGCGTGCGGCTCAGTCAGGGAGACATAGCTCAAGCCCGGAAGCTGTA
CAAATGCCCAGCGTGTGGGGAGACCCTGCAGGACACAACGGGAAACTTTTCTGCACCTGGTTTCCCAA
ATGGGTACCCATCTTACTCCCACTGCGTCTGGAGGATCTCGGTCACCCCAGGGGAAAAGATCGTATTA
AACTTCACATCCATGGATTTGTTTAAAAGCCGACTGTGCTGGTATGATTACGTGGAGGTCCGGGATGG
TTACTGGAGAAAAGCCCCCCTTTTGGGCAGGTTTTGTGGCGATAAGATCCCGGAGCCCCTCGTCTCCA
CGGACAGCCGGCTCTGGGTGGAGTTCCGCAGCAGCAGCAACATCTTGGGCAAGGGCTTCTTTGCAGCG
TACGAAGCTACCTGCGGGGGAGACATGAACAAAGATGCCGGTCAGATTCAATCTCCCAACTATCCGGA
TGACTACAGACCTTCCAAGGAATGTGTCTGGAGGATTACGGTTTCGGAGGGGTTTCACGTGGGACTTA
CCTTCCAAGCTTTTGAGATTGAAAGGCACGACAGCTGTGCATATGACTACCTGGAAGTCCGGGATGGC
CCCACGGAAGAGAGTGCCCTGATCGGCCACTTTTGTGGCTATGAGAAGCCGGAGGATGTGAAATCGAG
CTCCAACAGACTGTGGATGAAGTTTGTGTCCGATGGCTCTATCAATAAAGCGGGCTTTGCAGCCAATT
TTTTCAAGGAGGTGGATGAGTGTTCCTGGCCAGATCACGGCGGGTGCGAACATCGCTGTGTGAACACG
CTGGGCAGCTACAAGTGTGCCTGTGACCCTGGCTACGAGCTGGCCGCCGATAAGAAGATGTGTGAAGT
GGCCTGTGGCGGTTTCATTACCAAGCTGAATGGAACCATCACCAGCCCTGGGTGGCCGAAGGAGTATC
CCACAAACAAAAACTGTGTCTGGCAGGTGGTGGCCCCCGCTCAGTACCGGATCTCCCTTCAGTTTGAA
GTGTTTGAACTGGAAGGCAATGACGTCTGTAAGTACGACTTTGTAGAGGTGCGCAGCGGCCTGTCCCC
CGACGCCAAGCTGCACGGCAGGTTCTGCGGCTCTGAGACGCCGGAAGTCATCACCTCGCAGAGCAACA
ACATGCGCGTGGAGTTCAAGTCCGACAACACCGTCTCCAAGCGCGGCTTCAGGGCCCACTTCTTCTCA
GATAAGGACGAGTGTGCCAAGGACAACGGCGGGTGTCAGCATGAGTGCGTCAACACCTTCGGGAGCTA
CCTGTGCAGGTGCAGAAACGGCTACTGGCTCCACGAGAATGGGCATGACTGCAAAGAGGCTGGCTGTG
CACACAAGATCAGCAGTGTGGAGGGGACCCTGGCGAGCCCCAACTGGCCTGACAAATACCCCAGCCGG
AGGGAGTGTACCTGGAACATCTCTTCGACTGCAGGCCACAGAGTAAAACTCAGTGCGGGGTCAGGCTG
AAGGCTGAAGTGCAGACCAAAGAGCTCTATTCCCACGCCCAGTTTGGGGACAACAACTACCCGAGCGA
GGCCCGCTGTGACTGGGTGATCGTGGCAGAGGACGGCTACGGCGTGGAGCTGACATTCCGGACCTTTG
AGGTTGAGGAGGAGGCCGACTGCGGCTACGACTACATGGAAGCCTACGACGGCTACGACAGCTCAGCG
CCCAGGCTCGGCCGCTTCTGTGGCTCTGGGCCATTAGAAGAAATCTACTCTGCAGGTGGTTCCCTGAT
GATTCGATTCCGCACAGATGACACCATCAACAAGAAAGGCTTTCATGCCCGATACACCAGCACCAAGT
TCCAGGATGCCCTGCACATGAAGAAATAGTGCTGATGTTCTTGAAAGACAGAAACTGAGA

NOV8i, SNP13377387 of CG50235-04, Protein Sequence	SEQ ID NO: 102	821 aa	MW at 91404.2kD
			SNP Change: no change

MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAVFWGDIALD
EDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAGKDGRENTTLLHSPGTL
HAAAKTFSPRVRRATTSRTERIWPGGVIPYVIGGNFTGSQRAIFKQAMRHWEKHTCVTFIERTDEESF

IVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIVAHELGHVVGFWHEHTRPDRDQHVTIIRENIQP
GQEYNFLKMEAGEVSSLGETYDFDSIMHYARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIA
QARKLYKCPACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIVLNFTSMDLFKSRLCWYDY
VEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSNILGKGFFAAYEATCGGDMNKDAGQIQ
SPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFEIERHDSCAYDYLEVRDGPTEESALIGHFCGYEKP
EDVKSSSNRLWMKFVSDGSINKAGFAANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAAD
KKMCEVACGGFITKLNGTITSPGWPKEYPTNKNCVWQVVAPAQYRISLQFEVFELEGNDVCKYDFVEV
RSGLSPDAKLHGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECV
NTFGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTAGHRVKL
SAGSG

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 8B.

Table 8B. Comparison of the NOV8 protein sequences.

```
NOV8a   MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAV
NOV8b   MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAV
NOV8c   MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAV
NOV8d   MPRATALGALVSLLLLLPLPRGAGGLGERPDATADYSELDGEEGTEQQLEHYHDPCKAAV

NOV8a   FWGDIALDEDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAG
NOV8b   FWGDIALDEDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAG
NOV8c   FWGDIALDEDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAG
NOV8d   FWGDIALDEDDLKLFHIDKARDWTKQTVGATGHSTGGLEEQASESSPDTTAMDTGTKEAG

NOV8a   KDGRENTTLLHSPGTLHAAAKTFSPRVRRATTSRTERIWPGGVIPYVIGGNFTGSQRAIF
NOV8b   KDGRENTTLLHSPGTLHAAAKTFSPRVRRATTSRTERIWPGGVIPYVIGGNFTGSQRAIF
NOV8c   KDGRENTTLLHSPGTLHAAAKTFSPRVRRATTSRTERIWPGGVIPYVIGGNFTGSQRAIF
NOV8d   K----------------------------------------------------GSQRAIF

NOV8a   KQAMRHWEKHTCVTFIERTDEESFIVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIV
NOV8b   KQAMRHWEKHTCVTFIERTDEESFIVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIV
NOV8c   KQAMRHWEKHTCVTFIERTDEESFIVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIV
NOV8d   KQAMRHWEKHTCVTFIERTDEESFIVFSYRTCGCCSYVGRRGGGPQAISIGKNCDKFGIV

NOV8a   AHELGHVVGFWHEHTRPDRDQHVTIIRENIQPGQEYNFLKMEAGEVSSLGETYDFDSIMH
NOV8b   AHELGHVVGFWHEHTRPDRDQHVTIIRENIQPGQEYNFLKMEAGEVSSLGETYDFDSIMH
NOV8c   AHELGHVVGFWHEHTRPDRDQHVTIIRENIQPGQEYNFLKMEAGEVSSLGETYDFDSIMH
NOV8d   AHELGHVVGFWHEHTRPDRDQHVTIIRENIQPGQEYNFLKMEAGEVSSLGETYDFDSIMH

NOV8a   YARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIAQARKLYKCP-----------
NOV8b   YARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIAQARKLYKCP-----------
NOV8c   YARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIAQARKLYKCP-----------
NOV8d   YARNTFSRGVFLDTILPRQDDNGVRPTIGQRVRLSQGDIAQARKLYKCPGPTCAFVSQKT

NOV8a   ------------------ACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIV
NOV8b   ------------------ACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIV
NOV8c   ------------------ACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIV
NOV8d   SICLLHFSPTCSEGFGWQRACGETLQDTTGNFSAPGFPNGYPSYSHCVWRISVTPGEKIV

NOV8a   LNFTSMDLFKSRLCWYDYVEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSN
NOV8b   LNFTSMDLFKSRLCWYDYVEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSN
NOV8c   LNFTSMDLFKSRLCWYDYVEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSN
```

```
NOV8d   LNFTSMDLFKSRLCWYDYVEVRDGYWRKAPLLGRFCGDKIPEPLVSTDSRLWVEFRSSSN

NOV8a   ILGKGFFAAYEATCGGDMNKDAGQIQSPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFE
NOV8b   ILGKGFFAAYEATCGGDMNKDAGQIQSPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFE
NOV8c   ILGKGFFAAYEATCGGDMNKDAGQIQSPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFE
NOV8d   ILGKGFFAAYEATCGGDMNKDAGQIQSPNYPDDYRPSKECVWRITVSEGFHVGLTFQAFE

NOV8a   IERHDSCAYDYLEVRDGPTEESALIGHFCGYEKPEDVKSSSNRLWMKFVSDGSINKAGFA
NOV8b   IERHDSCAYDYLEVRDGPTEESALIGHFCGYEKPEDVKSSSNRLWMKFVSDGSINKAGFA
NOV8c   IERHDSCAYDYLEVRDGPTEESALIGHFCGYEKPEDVKSSSNRLWMKFVSDGSINKAGFA
NOV8d   IERHDSCAYDYLEVRDGPTEESALIGHFCGYEKPEDVKSSSNRLWMKFVSDGSINKAGFA

NOV8a   ANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAADKKMCEVACGGFITKLNGT
NOV8b   ANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAADKKMCEVACGGFITKLNGT
NOV8c   ANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAADKKMCEVACGGFITKLNGT
NOV8d   ANFFKEVDECSWPDHGGCEHRCVNTLGSYKCACDPGYELAADKKMCEVACGGFITKLNGT

NOV8a   ITSPGWPKEYPTNKNCVWQVVAPAQYRISLQFEVFELEGNDVCKYDFVEVRSGLSPDAKL
NOV8b   ITSPGWPKEYPTNKNCVWQVVAPTQYRISLQFEVFELEGNDVCKYDFVEVRSGLSPDAKL
NOV8c   ITSPGWPKEYPTNKNCVWQVVAPAQYRISLQFEVFELEGNDVCKYDFVEVRSGLSPDAKL
NOV8d   ITSPGWPKEYPTNKNCVWQVVAPTQYRISLQFEVFELEGNDVCKYDFVEVRSGLSPDAKL

NOV8a   HGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECVNT
NOV8b   HGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECVNT
NOV8c   HGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECVNT
NOV8d   HGRFCGSETPEVITSQSNNMRVEFKSDNTVSKRGFRAHFFSDKDECAKDNGGCQHECVNT

NOV8a   FGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTA
NOV8b   FGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTA
NOV8c   FGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTA
NOV8d   FGSYLCRCRNGYWLHENGHDCKEAGCAHKISSVEGTLASPNWPDKYPSRRECTWNISSTA

NOV8a   GHRVKLSAGSG----------------------------------------------------
NOV8b   GHRVKLTFNEFEIEQHQECAYDHLEMYDGPDSLAPILGRFCGSKKPDPTVASGSKCGGRL
NOV8c   GHRVKLTFNEFEIEQHQECAYDHLEMYDGPDSLAPILGRFCGSKKPDPTVASGSSMFLRF
NOV8d   GHRVKLTFNEFEIEQHQECAYDHLEMYDGPDSLAPILGRFCGSKKPDPTVASGSKCGGRL

NOV8a   ----------------------------------------------------------
NOV8b   KAEVQTKELYSHAQFGDNNYPSEARCDWVIVAEDGYGVELTFRTFEVEEEADCGYDYMEA
NOV8c   YSDASVQRKGFQAVHSTECGGRLKAEVQTKELYSHAQFGDNNYPSEARCDWVIVAEDGYG
NOV8d   KAEVQTKELYSHAQFGDNNYPSEARCDWVIVAEDGYGVELTFRTFEVEEEADCGYDYMEA

NOV8a   ----------------------------------------------------------
NOV8b   YDGYDSSAPRLGRFCGSGPLEEIYSAGDSLMIRFRTDDTINKKGFHARYTSTKFQDGLHM
NOV8c   VELTFRTFEVEEEADCGYDYMEAYDGYDSSAPRLGRFCGSGPLEEIYSAGDSLMIRFRTD
NOV8d   YDGYDSSAPRLGRFCGSGPLEEIYSAGDSLMIRFRTDDTINKKGFHARYTSTKFQDALHM

NOV8a   -------------------------
NOV8b   KK-----------------------
NOV8c   DTINKKGFHARYTSTKFQDALHMKK
NOV8d   KK-----------------------

NOV8a  (SEQ ID NO:  86)
NOV8b  (SEQ ID NO:  88)
```

NOV8c (SEQ ID NO: 90) NOV8d (SEQ ID NO: 92)

Further analysis of the NOV8a protein yielded the following properties shown in Table 8C.

Table 8C. Protein Sequence Properties NOV8a

SignalP analysis:	Cleavage site between residues 26 and 27
PSORT II analysis:	

```
PSG:  a new signal peptide prediction method
      N-region:  length 3;  pos.chg 1;  neg.chg 0
      H-region:  length 17;  peak value  10.45
      PSG score:   6.05

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   4.44
      possible cleavage site: between 25 and 26

>>> Seems to have a cleavable signal peptide (1 to 25)

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 26
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  6.31 (at 203)
      ALOM score:   6.31  (number of TMSs: 0)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 12
      Charge difference: -3.0   C(-1.0) - N( 2.0)
      N >= C: N-terminal side will be inside

MITDISC: discrimination of mitochondrial targeting seq
      R content:       2       Hyd Moment(75):  7.80
      Hyd Moment(95):  7.61    G content:       5
      D/E content:     1       S/T content:     2
      Score: -4.54

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 31  PRG|AG

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: PRVRRAT (5) at  145
      bipartite: none
      content of basic residues:  11.2%
      NLS Score: -0.04

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
```

```
      XXRR-like motif in the N-terminus: PRAT

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 76.7

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

-------------------------

Final Results (k = 9/23):

       33.3 %: extracellular, including cell wall
       22.2 %: Golgi
       22.2 %: vacuolar
       11.1 %: mitochondrial
       11.1 %: endoplasmic reticulum

>> prediction for CG50235-04 is exc (k=9)
```

A search of the NOV8a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 8D.

Table 8D. Geneseq Results for NOV8a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV8a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAY32240	Human tolloid-like protein mT11-2 - Homo sapiens, 1015 aa. [WO9951730-A2, 14-OCT-1999]	1..817 1..817	816/817 (99%) 817/817 (99%)	0.0
ABG79187	Human tolloid-like 2-like protein #1 - Homo sapiens, 992 aa. [WO200264791-A2, 22-AUG-2002]	1..817 1..817	815/817 (99%) 816/817 (99%)	0.0
ABG79188	Human tolloid-like 2-like protein #2 - Homo sapiens, 970 aa. [WO200264791-A2, 22-AUG-2002]	1..817 1..795	763/847 (90%) 764/847 (90%)	0.0
AAY32241	Mouse tolloid-like protein mT11-2 - Mus musculus, 1012 aa. [WO9951730-A2, 14-OCT-1999]	1..817 1..814	728/817 (89%) 759/817 (92%)	0.0
AAW40224	Human tolloid-like (Tll) protein - Homo sapiens, 1013 aa. [WO9745528-A2, 04-DEC-1997]	35..817 34..815	584/783 (74%) 666/783 (84%)	0.0

In a BLAST search of public sequence databases, the NOV8a protein was found to have homology to the proteins shown in the BLASTP data in Table 8E.

Table 8E. Public BLASTP Results for NOV8a				
Protein Accession Number	**Protein/Organism/Length**	**NOV8a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9Y6L7	Tolloid-like 2 protein - Homo sapiens (Human), 1015 aa.	1..817 1..817	816/817 (99%) 817/817 (99%)	0.0
Q9WVM6	Tolloid-like-2 protein - Mus musculus (Mouse), 1012 aa.	1..817 1..814	728/817 (89%) 759/817 (92%)	0.0
Q9UQ00	Hypothetical protein KIAA0932 - Homo sapiens (Human), 926 aa (fragment).	90..817 1..728	726/728 (99%) 727/728 (99%)	0.0
O57382	Xolloid - Xenopus laevis (African clawed frog), 1019 aa.	52..816 52..819	596/768 (77%) 680/768 (87%)	0.0
Q8JI28	Xolloid-like metalloprotease - Xenopus laevis (African clawed frog), 1007 aa.	35..817 29..810	597/783 (76%) 679/783 (86%)	0.0

PFam analysis indicates that the NOV8a protein contains the domains shown in the Table 8F.

Table 8F. Domain Analysis of NOV8a			
Pfam Domain	**NOV8a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Astacin	157..350	100/202 (50%) 172/202 (85%)	9.5e-114
CUB	351..460	50/116 (43%) 99/116 (85%)	2.4e-52
CUB	464..573	62/116 (53%) 98/116 (84%)	3.1e-60
EGF	580..616	17/47 (36%) 29/47 (62%)	1.2e-05
CUB	620..729	57/116 (49%) 101/116 (87%)	4.7e-58
EGF	736..771	18/47 (38%) 31/47 (66%)	7.4e-07

Example 9.

The NOV9 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 9A.

Table 9A. NOV9 Sequence Analysis		
NOV9a, CG50249-01 DNA Sequence	SEQ ID NO: 103	1953 bp
	ORF Start: ATG at 16	ORF Stop: TAA at 1930

GTCTGAGTCACAGAG**ATG**GGCAAGATCGAGAACAACGAGAGGGTGATCCTCAATGTCGGGGGCACCCG
GCACGAAACCTACCGCAGCACCCTCAAGACCCTGCCTGGAACACGCCTGGCCCTTCTTGCCTCCTCCG
AGCCCCCAGGCGACTGCTTGACCACGGCGGGCGACAAGCTGCAGCCGTCGCCGCCTCCACTGTCGCCG
CCGCCGAGAGCGCCCCCGCTGTCCCCCGGGCCAGGCGGCTGCTTCGAGGGCGGCGCGGGCAACTGCAG
TTCCCGCGGCGGCAGGGCCAGCGACCATCCCGGTGGCGGCCGCGAGTTCTTCTTCGACCGGCACCCGG
GCGTCTTCGCCTATGTGCTCAATTACTACCGCACCGGCAAGCTGCACTGCCCCGCAGACGTGTGCGGG
CCGCTCTTCGAGGAGGAGCTGGCCTTCTGGGGCATCGACGAGACCGACGTGGAGCCCTGCTGCTGGAT
GACCTACCGGCAGCACCGCGACGCCGAGGAGGCGCTGGACATCTTCGAGACCCCCGACCTCATTGGCG
GCGACCCCGGCGACGACGAGGACCTGGCGGCCAAGAGGCTGGGCATCGAGGACGCGGCGGGGCTCGGG
GGCCCGGACGGCAAATCTGGCCGCTGGAGGAGGCTGCAGCCCCGCATGTGGGCCCTCTTCGAAGACCC
CTACTCGTCCAGAGCCGCCAGGTTTATTGCTTTTGCTTCTTTATTCTTCATCCTGGTTTCAATTACAA
CTTTTTGCCTGGAAACACATGAAGCTTTCAATATTGTTAAAAACAAGACAGAACCAGTCATCAATGGC
ACAAGTGTTGTTCTACAGTATGAAATTGAAACGGATCCTGCCTTGACGTATGTAGAAGGAGTGTGTGT
GGTGTGGTTTACTTTTGAATTTTTAGTCCGTATTGTTTTTTCACCCAATAAACTTGAATTCATCAAAA
ATCTCTTGAATATCATTGACTTTGTGGCCATCCTACCTTTCTACTTAGAGGTGGGACTCAGTGGGCTG
TCATCCAAAGCTGCTAAAGATGTGCTTGGCTTCCTCAGGGTGGTAAGGTTTGTGAGGATCCTGAGAAT
TTTCAAGCTCACCCGCCATTTTGTAGGTCTGAGGGTGCTTGGACATACTCTTCGAGCTAGTACTAATG
AATTTTTGCTGCTGATAATTTTCCTGGCTCTAGGAGTTTTGATATTTGCTACCATGATCTACTATGCC
GAGAGAGTGGGAGCTCAACCTAACGACCCTTCAGCTAGTGAGCACACACAGTTCAAAAACATTCCCAT
TGGGTTCTGGTGGGCTGTAGTGACCATGACTACCCTGGGTTATGGGGATATGTACCCCCAAACATGGT

CAGGCATGCTGGTGGGAGCCCTGTGTGCTCTGGCTGGAGTGCTGACAATAGCCATGCCAGTGCCTGTC
ATTGTCAATAATTTTGGAATGTACTACTCCTTGGCAATGGCAAAGCAGAAACTTCCAAGGAAAAGAAA
GAAGCACATCCCTCCTGCTCCTCAGGCAAGCTCACCTACTTTTTGCAAGACAGAATTAAATATGGCCT
GCAATAGTACACAGAGTGACACATGTCTGGGCAAAGACAATCGACTTCTGGAACATAACAGATCAGTG
TTATCAGGTGACGACAGTACAGGAAGTGAGCCGCCACTATCACCCCCAGAAAGGCTCCCCATCAGACG
CTCTAGTACCAGAGACAAAAACAGAAGAGGGGAAACATGTTTCCTACTGACGACAGGTGATTACACGT
GTGCTTCTGATGGAGGGATCAGGAAAGGTTATGAAAAATCCCGAAGCTTAAACAACATAGCGGGCTTG
GCAGGCAATGCTCTGAGGCTCTCTCCAGTAACATCACCCTACAACTCTCCTTGTCCTCTGAGGCGCTC
TCGATCTCCCATCCCATCTATCTTGTAAACCAAACAACCAAACTGCATC

NOV9a, CG50249-01 Protein Sequence	SEQ ID NO: 104	638 aa	MW at 70224.7kD

MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGDCLTTAGDKLQPSPPPLSPPPRAP
PLSPGPGGCFEGGAGNCSSRGGRASDHPGGGREFFFDRHPGVFAYVLNYYRTGKLHCPADVCGPLFEE
ELAFWGIDETDVEPCCWMTYRQHRDAEEALDIFETPDLIGGDPGDDEDLAAKRLGIEDAAGLGGPDGK
SGRWRRLQPRMWALFEDPYSSRAARFIAFASLFFILVSITTFCLETHEAFNIVKNKTEPVINGTSVVL
QYEIETDPALTYVEGVCVVWFTFEFLVRIVFSPNKLEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAA
KDVLGFLRVVRFVRILRIFKLTRHFVGLRVLGHTLRASTNEFLLLIIFLALGVLIFATMIYYAERVGA
QPNDPSASEHTQFKNIPIGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIAMPVPVIVNNF
GMYYSLAMAKQKLPRKRKKHIPPAPQASSPTFCKTELNMACNSTQSDTCLGKDNRLLEHNRSVLSGDD
STGSEPPLSPPERLPIRRSSTRDKNRRGETCFLLTTGDYTCASDGGIRKGYEKSRSLNNIAGLAGNAL
RLSPVTSPYNSPCPLRRSRSPIPSIL

NOV9b, 207885588 DNA Sequence	SEQ ID NO: 105	1815 bp
	ORF Start: at 1	ORF Stop: end of sequence

AGATCTCCCACCATGGGCAAGATCGAGAACAACGAGAGGGTGATCCTCAATGTCGGGGGCACCCGGCA
CGAAACCTACCGCAGCACCCTCAAGACCCTGCCTGGAACACGCCTGGCCCTTCTTGCCTCCTCCGAGC
CCCCAGGCAACTGCAGTTCCCGCGGCGGCAGGGCCAGCGACCATCCCGGTGGCGGCCGCGAGTTCTTC
TTCGACCGGCACCCGGGCGTCTTCGCCTATGTGCTCAATTACTACCGCACCGGCAAGCTGCACTGCCC
CGCAGACGTGTGCGGGCCGCTCTTCGAGGAGGAGCTGGCCTTCTGGGGCATCGACGAGACCGACGTGG
AGCCCTGCTGCTGGATGACCTACCGGCAGCACCGCGACGCCGAGGAGGCGCTGGACATCTTCGAGACC
CCCGACCTCATTGGCGGCGACCCCGGCGACGACGAGGACCTGGCGGCCAAGAGGCTGGGCATCGAGGA
CGCGGCGGGGCTCGGGGGCCCCGACGGCAAATCTGGCCGCTGGAGGAGGCTGCAGCCCCGCATGTGGG
CCCTCTTCGAAGACCCCTACTCGTCCAGAGCCGCCAGGTTTATTGCTTTTGCTTCTTTATTCTTCATC
CTGGTTTCAATTACAACTTTTTGCCTGGAAACACATGAAGCTTTCAATATTGTTAAAAACAAGACAGA
ACCAGTCATCAATGGCACAAGTGTTGTTCTACAGTATGAAATTGAAACGGATCCTGCCTTGACGTATG
TAGAAGGAGTGTGTGTGGTGTGGTTTACTTTTGAATTTTTAGTCCGTATTGTTTTTTCACCCAATAAA
CTTGAATTCATCAAAAATCTCTTGAATATCATTGACTTTGTGGCCATCCTACCTTTCTACTTAGAGGT
GGGACTCAGTGGGCTGTCATCCAAAGCTGCTAAAGATGTGCTTGGCTTCCTCAGGGTGGTAAGGTTTG
TGAGGATCCTGAGAATTTTCAAGCTCACCCGCCATTTTGTAGGTCTGAGGGTGCTTGGACATACTCTT
CGAGCTAGTACTAATGAATTTTTGCTGCTGATAATTTTCCTGGCTCTAGGAGTTTTGATATTTGCTAC
CATGATCTACTATGCCGAGAGAGTGGGAGCTCAACCTAACGACCCTTCAGCTAGTGAGCACACACAGT
TCAAAAACATTCCCATTGGGTTCTGGTGGGCTGTAGTGACCATGACTACCCTGGGTTATGAGGATACG
TACCCCCAAACATGGTCAGGCATGCTGGTGGGAGCCCTGTGTGCTCTGGCTGGAGTGCTGACAATAGC
CATGCCAGTGCCTGTCATTGTCAATAATTTTGGAATGTACTACTCCTTGGCAATGGCAAAGCAGAAAC
TTCCAAGGAAAAGAAAGAAGCACATCCCTCCTGCTCCTCAGGCAAGCTCACCTACTTTTTGCAAGACA
GAATTAAATATGGCCTGCAATAGTACACAGAGTGACACATGTCTGGGCAAAGACAATCGACTTCTGGA
ACATAACAGATCAGTGTTATCAGGTGACGACAGTACAGGAAGTGAGCCGCCACTATCACCCCCAGAAA
GGCTCCCCATCAGACGCTCTAGTACCAGAGACAAAAACAGAAGAGGGGAAACATGTTTCCTACTGACG
ACAGGTGATTACACGTGTGCTTCTGATGGAGGGATCAGGAAAGGATATGAAAAATCCCGAAGCTTAAA
CAACATAGCGGGCTTGGCAGGCAATGCTCTGAGGCTCTCTCCAGTAACATCACCCTACAACTCTCCTT
GTCCTCTGAGGCGCTCTCGATCTCCCATCCCATCTATCTTGCTCGAG

NOV9b, 207885588 Protein Sequence	SEQ ID NO: 106	605 aa	MW at 67228.3kD

RSPTMGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGNCSSRGGRASDHPGGGREFF
FDRHPGVFAYVLNYYRTGKLHCPADVCGPLFEEELAFWGIDETDVEPCCWMTYRQHRDAEEALDIFET
PDLIGGDPGDDEDLAAKRLGIEDAAGLGGPDGKSGRWRRLQPRMWALFEDPYSSRAARFIAFASLFFI
LVSITTFCLETHEAFNIVKNKTEPVINGTSVVLQYEIETDPALTYVEGVCVVWFTFEFLVRIVFSPNK

LEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAAKDVLGFLRVVRFVRILRIFKLTRHFVGLRVLGHTL
RASTNEFLLLIIFLALGVLIFATMIYYAERVGAQPNDPSASEHTQFKNIPIGFWWAVVTMTTLGYEDT
YPQTWSGMLVGALCALAGVLTIAMPVPVIVNNFGMYYSLAMAKQKLPRKRKKHIPPAPQASSPTFCKT
ELNMACNSTQSDTCLGKDNRLLEHNRSVLSGDDSTGSEPPLSPPERLPIRRSSTRDKNRRGETCFLLT
TGDYTCASDGGIRKGYEKSRSLNNIAGLAGNALRLSPVTSPYNSPCPLRRSRSPIPSILLE

NOV9c, CG50249-02 DNA Sequence	SEQ ID NO: 107	607 bp
	ORF Start: ATG at 13	ORF Stop: at 604

AGATTTCCCACC**ATG**GGCAAGATCGAGAACAACGAGAGGGTGATCCTCAATGTCGGGGGCACCCGGCA
CGAAACCTACCGCAGCACCCTCAAGACCCTGCCTGGAACACGCCTGGCCCTTCTTGCCTCCTCCGAGC
CCCCAGGCGACTGCTTGACCACAGCGGGCAACTGCAGTTCCCGCGGCGGCAGGGCCAGCGACCATCCC
GGTGGCGGCCGCGAGTTCTTCTTCGACCGGCATCCGGGCGTCTTCGCCTATGTGCTCAATTACTACCG
CACCGGCAAGCTGCACTGTCCCGCAGACGTGTGCGGGCCGCTCTTCGAGGAGGAGCTGGCCTTCTGGG
GCATCGACGAGACCGACGTGGAGCCCTGCTGCTGGATGACCTACCGGCAGCACCGCGACGCCGAGGAG
GCGCTGGACATCTTCGAGACCCCCGACCTCATTGGCGGCGACCCCGGCGACGACGAGGACCTGGCGGC
CAAGAGGCTGGGCATCGAGGACGCGGCGGGGCTCGGGGGCCCCGACGGCAAATCTGGCCGCTGGAGGA
GGCTGCAGCCCCGCATGTGGGCCCTCTTCGAAGACCCCTACTCGTCCAGAGCCGCCAGGCTCG

NOV9c, CG50249-02 Protein Sequence	SEQ ID NO: 108	197 aa	MW at 21779.0kD

MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGDCLTTAGNCSSRGGRASDHPGGGR
EFFFDRHPGVFAYVLNYYRTGKLHCPADVCGPLFEEELAFWGIDETDVEPCCWMTYRQHRDAEEALDI
FETPDLIGGDPGDDEDLAAKRLGIEDAAGLGGPDGKSGRWRRLQPRMWALFEDPYSSRAAR

NOV9d, CG50249-03 DNA Sequence	SEQ ID NO: 109	1815 bp
	ORF Start: ATG at 13	ORF Stop: at 1810

AGATCTCCCACC**ATG**GGCAAGATCGAGAACAACGAGAGGGTGATCCTCAATGTCGGGGGCACCCGGCA
CGAAACCTACCGCAGCACCCTCAAGACCCTGCCTGGAACACGCCTGGCCCTTCTTGCCTCCTCCGAGC
CCCCAGGCAACTGCAGTTCCCGCGGCGGCAGGGCCAGCGACCATCCCGGTGGCGGCCGCGAGTTCTTC
TTCGACCGGCACCCGGGCGTCTTCGCCTATGTGCTCAATTACTACCGCACCGGCAAGCTGCACTGCCC
CGCAGACGTGTGCGGGCCGCTCTTCGAGGAGGAGCTGGCCTTCTGGGGCATCGACGAGACCGACGTGG
AGCCCTGCTGCTGGATGACCTACCGGCAGCACCGCGACGCCGAGGAGGCGCTGGACATCTTCGAGACC
CCCGACCTCATTGGCGGCGACCCCGGCGACGACGAGGACCTGGCGGCCAAGAGGCTGGGCATCGAGGA
CGCGGCGGGGCTCGGGGGCCCCGACGGCAAATCTGGCCGCTGGAGGAGGCTGCAGCCCCGCATGTGGG
CCCTCTTCGAAGACCCCTACTCGTCCAGAGCCGCCAGGTTTATTGCTTTTGCTTCTTTATTCTTCATC
CTGGTTTCAATTACAACTTTTTGCCTGGAAACACATGAAGCTTTCAATATTGTTAAAAACAAGACAGA
ACCAGTCATCAATGGCACAAGTGTTGTTCTACAGTATGAAATTGAAACGGATCCTGCCTTGACGTATG
TAGAAGGAGTGTGTGTGGTGTGGTTTACTTTTGAATTTTTAGTCCGTATTGTTTTTTCACCCAATAAA
CTTGAATTCATCAAAAATCTCTTGAATATCATTGACTTTGTGGCCATCCTACCTTTCTACTTAGAGGT
GGGACTCAGTGGGCTGTCATCCAAAGCTGCTAAAGATGTGCTTGGCTTCCTCAGGGTGGTAAGGTTTG
TGAGGATCCTGAGAATTTTCAAGCTCACCCGCCATTTTGTAGGTCTGAGGGTGCTTGGACATACTCTT
CGAGCTAGTACTAATGAATTTTTGCTGCTGATAATTTTCCTGGCTCTAGGAGTTTTGATATTTGCTAC
CATGATCTACTATGCCGAGAGAGTGGGAGCTCAACCTAACGACCCTTCAGCTAGTGAGCACACACAGT
TCAAAAACATTCCCATTGGGTTCTGGTGGGCTGTAGTGACCATGACTACCCTGGGTTATGAGGATACG
TACCCCCAAACATGGTCAGGCATGCTGGTGGGAGCCCTGTGTGCTCTGGCTGGAGTGCTGACAATAGC
CATGCCAGTGCCTGTCATTGTCAATAATTTTGGAATGTACTACTCCTTGGCAATGGCAAAGCAGAAAC
TTCCAAGGAAAAGAAAGAAGCACATCCCTCCTGCTCCTCAGGCAAGCTCACCTACTTTTTGCAAGACA
GAATTAAATATGGCCTGCAATAGTACACAGAGTGACACATGTCTGGGCAAAGACAATCGACTTCTGGA
ACATAACAGATCAGTGTTATCAGGTGACGACAGTACAGGAAGTGAGCCGCCACTATCACCCCCAGAAA
GGCTCCCCATCAGACGCTCTAGTACCAGAGACAAAAACAGAAGAGGGGAAACATGTTTCCTACTGACG
ACAGGTGATTACACGTGTGCTTCTGATGGAGGGATCAGGAAAGGATATGAAAAATCCCGAAGCTTAAA
CAACATAGCGGGCTTGGCAGGCAATGCTCTGAGGCTCTCTCCAGTAACATCACCCTACAACTCTCCTT
GTCCTCTGAGGCGCTCTCGATCTCCCATCCCATCTATCTTGCTCGAG

NOV9d, CG50249-03 Protein Sequence	SEQ ID NO: 110	599 aa	MW at 66544.6kD

MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGNCSSRGGRASDHPGGGREFFFDRH
PGVFAYVLNYYRTGKLHCPADVCGPLFEEELAFWGIDETDVEPCCWMTYRQHRDAEEALDIFETPDLI

GGDPGDDEDLAAKRLGIEDAAGLGGPDGKSGRWRRLQPRMWALFEDPYSSRAARFIAFASLFFILVSI
TTFCLETHEAFNIVKNKTEPVINGTSVVLQYEIETDPALTYVEGVCVVWFTFEFLVRIVFSPNKLEFI
KNLLNIIDFVAILPFYLEVGLSGLSSKAAKDVLGFLRVVRFVRILRIFKLTRHFVGLRVLGHTLRAST
NEFLLLIIFLALGVLIFATMIYYAERVGAQPNDPSASEHTQFKNIPIGFWWAVVTMTTLGYEDTYPQT
WSGMLVGALCALAGVLTIAMPVPVIVNNFGMYYSLAMAKQKLPRKRKKHIPPAPQASSPTFCKTELNM
ACNSTQSDTCLGKDNRLLEHNRSVLSGDDSTGSEPPLSPPERLPIRRSSTRDKNRRGETCFLLTTGDY
TCASDGGIRKGYEKSRSLNNIAGLAGNALRLSPVTSPYNSPCPLRRSRSPIPSIL

NOV9e, CG50249-04 DNA Sequence	SEQ ID NO: 111	3028 bp
	ORF Start: ATG at 22	ORF Stop: TAA at 1861

AGTCATGTCTGAGTCACAGAG**ATG**GGCAAGATCGAGAACAACGAGAGGGTGATCCTCAATGTCGGGGG
CACCCGGCACGAAACCTACCGCAGCACCCTCAAGACCCTGCCTGGAACACGCCTGGCCCTTCTTGCCT
CCTCCGAGCCCCCAGGCGACTGCTTGACCACGGCGGGCGACAAGCTGCAGCCGTCGCCGCCTCCACTG
TCGCCGCCGCCGAGAGCGCCCCCGCTGTCCCCCGGGCCAGGCGGCTGCTTCGAGGGCGGCGCGGGCAA
CTGCAGTTCCCGCGGCGGCAGGGCCAGCGACCATCCCGGTGGCGGCCGCGAGTTCTTCTTCGACCGGC
ACCCGGGCGTCTTCGCCTATGTGCTCAATTACTACCGCACCGGCAAGCTGCACTGCCCCGCAGACGTG
TGCGGGCCGCTCTTCGAGGAGGAGCTGGCCTTCTGGGGCATCGACGAGACCGACGTGGAGCCCTGCTG
CTGGATGACCTACCGGCAGCACCGCGACGCCGAGGAGGCGCTGGACATCTTCGAGACCCCCGACCTCA
TTGGCGGCGACCCCGGCGACGACGAGGACCTGGCGGCCAAGAGGCTGGGCATCGAGGACGCGGCGGGG
CTCGGGGGCCCCGACGGCAAATCTGGCCGCTGGAGGAGGCTGCAGCCCCGCATGTGGGCCCTCTTCGA
AGACCCCTACTCGTCCAGAGCCGCCAGGTTTATTGCTTTTGCTTCTTTATTCTTCATCCTGGTTTCAA
TTACAACTTTTTGCCTGGAAACACATGAAGCTTTCAATATTGTTAAAAACAAGACAGAACCAGTCATC
AATGGCACAAGTGTTGTTCTACAGTATGAAATTGAAACGGATCCTGCCTTGACGTATGTAGAAGGAGT
GTGTGTGGTGTGGTTTACTTTTGAATTTTTAGTCCGTATTGTTTTTTCACCCAATAAACTTGAATTCA
TCAAAAATCTCTTGAATATCATTGACTTTGTGGCCATCCTACCTTTCTACTTAGAGGTGGGACTCAGT
GGGCTGTCATCCAAAGCTGCTAAAGATGTGCTTGGCTTCCTCAGGGTGGTAAGGTTTGTGAGGATCCT
GAGAATTTTCAAGCTCACCCGCCATTTTGTAGGTCTGAGGGTGCTTGGACATACTCTTCGAGCTAGTA
CTAATGAATTTTTGCTGCTGATAATTTTCCTGGCTCTAGGAGTTTTGATATTTGCTACCATGATCTAC
TATGCCGAGAGAGTGGGAGCTCAACCTAACGACCCTTCAGCTAGTGAGCACACACAGTTCAAAAACAT
TCCCATTGGGTTCTGGTGGGCTGTAGTGACCATGACTACCCTGGGTTATGGGGATATGTACCCCCAAA
CATGGTCAGGCATGCTGGTGGGAGCCCTGTGTGCTCTGGCTGGAGTGCTGACAATAGCCATGCCAGTG
CCTGTCATTGTCAATAATTTTGGAATGTACTACTCCTTGGCAATGGCAAAGCAGAAACTTCCAAGGAA
AAGAAAGAAGCACATCCCTCCTGCTCCTCAGGCAAGCTCACCTACTTTTTGCAAGACAGAATTAAATA
TGGCCTGCAATAGTACACAGAGTGACACATGTCTGGGCAAAGACAATCGACTTCTGGAACATAACAGA
TCAGTGTTATCAGGTGACGACAGTACAGGAAGTGAGCCGCCACTATCACCCCCAGAAAGGCTCCCCAT
CAGACGCTCTAGTACCAGAGACAAAAACAGAAGAGGGGAAACATGTTTCCTACTGACGACAGGTGATT
ACACGTGTGCTTCTGATGGAGGGATCAGGAAAGATAACTGCAAAGAGGTTGTCATTACTGGTTACACG
CAAGCCGAGGCCAGATCTCTTACT**TAA**TGACTTGGGGGAAGGCACAAAACATGAGAGAAAGTGTTGTA
CAGAATTTATCATGGATTATTGACTGCTGAGAAAGGGACAGTGGAATTTAGCCATACAAAGGACTATA
CTGGAAACAGACTTCTGCTGCTGAATGTGCCCTGATGTGACCAGGTTGCACTTGGAAGAGATCCTCCG
CGTCTTCATGAGGCACTTAAAGCTTATAAAAGAACTGCGGCTGGAACTCATCTGGTGCTCCCCATGAG
AGTGCTCTGCTTGTAGACTGGCCAGTGTCCATGAAACAACTGTAAATACCAACATGTGTGCATGGGTC
AACAGTCTTGGCCATTTCTCATCAAAAGAAGCCAAATTCATGATCAACATCTCTGAAGTTTCAAGTAA
GGCCCACACTTCTTTGAATTACTCTTCATGGGCCCACATTAGGTTGTGCTGTGAATTACTTAAGGCAG
TGATACTGATGTAGTATAGTTTTGTCTTAATTTCCCTTATTTCTACTTCTTTGGTTGAATCTATGAAC
TTGATTGTATAATTTTCTTATAAATTACTGATGTAATCAGCTTGTCAATTATGTTGTGAAATTGTTAG
TATTCATTTATCAAAAATGACCTATGTTTAGTCACATATTTGTTTAGTTCTGGGAAATTGTTATAGCT
TAAATGGAACTCACCAACATTATTCATAGTTTAAGTCTTTTATCATTATTACCTCAATTATAAATATT
ACAAAAACATAATTCTGGCAATGAGAGTATTTTTTTATTCAATGATCAAGGAGCAATGTCAGTATATA
GTAGAATATCAATTAAATTATATCCTAAAATGTATATTTTGCATAAAAGAGATATTCTTTAATCAATT
ACTTTTTTGTGAGTTTTGTGGCGAATGAAGCTTGTACGTGTCTTTAAAACTGTTGTAGATGAAACTGT
ATAAGATTTTTACATCTTGCTTAATCAATATTTTCAGAGTCTATTAGTTCCCCTGGGATTCTGAATAT
AACATATAGCCTATTATAAATCCCTGTATCGTGGACCTTTTGTGAACATTTCAAGGCGCATGCACAAC
CTTGATGATAACCAGTGGAAATGTAACTAACTGAAATGAAGAATAAAAGGCAAATGAGCTGGGGATAA
ACTTGAATGTTATCTGATTAAATTACTCAAATTATT

NOV9e, CG50249-04 Protein Sequence	SEQ ID NO: 112	613 aa	MW at 67598.7kD

MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGDCLTTAGDKLQPSPPPLSPPPRAP

PLSPGPGGCFEGGAGNCSSRGGRASDHPGGGREFFFDRHPGVFAYVLNYYRTGKLHCPADVCGPLFEE
ELAFWGIDETDVEPCCWMTYRQHRDAEEALDIFETPDLIGGDPGDDEDLAAKRLGIEDAAGLGGPDGK
SGRWRRLQPRMWALFEDPYSSRAARFIAFASLFFILVSITTFCLETHEAFNIVKNKTEPVINGTSVVL
QYEIETDPALTYVEGVCVVWFTFEFLVRIVFSPNKLEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAA
KDVLGFLRVVRFVRILRIFKLTRHFVGLRVLGHTLRASTNEFLLLIIFLALGVLIFATMIYYAERVGA
QPNDPSASEHTQFKNIPIGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIAMPVPVIVNNF
GMYYSLAMAKQKLPRKRKKHIPPAPQASSPTFCKTELNMACNSTQSDTCLGKDNRLLEHNRSVLSGDD
STGSEPPLSPPERLPIRRSSTRDKNRRGETCFLLTTGDYTCASDGGIRKDNCKEVVITGYTQAEARSL
T

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 9B.

Table 9B. Comparison of the NOV9 protein sequences.

```
NOV9a   ----MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGDCLTTAGDKLQP
NOV9b   RSPTMGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPG------------
NOV9c   ----MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGDCLTTA------
NOV9d   ----MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPG------------
NOV9e   ----MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGDCLTTAGDKLQP

NOV9a   SPPPLSPPPRAPPLSPGPGGCFEGGAGNCSSRGGRASDHPGGGREFFFDRHPGVFAYVLN
NOV9b   ----------------------------NCSSRGGRASDHPGGGREFFFDRHPGVFAYVLN
NOV9c   ----------------------------GNCSSRGGRASDHPGGGREFFFDRHPGVFAYVLN
NOV9d   ----------------------------NCSSRGGRASDHPGGGREFFFDRHPGVFAYVLN
NOV9e   SPPPLSPPPRAPPLSPGPGGCFEGGAGNCSSRGGRASDHPGGGREFFFDRHPGVFAYVLN

NOV9a   YYRTGKLHCPADVCGPLFEEELAFWGIDETDVEPCCWMTYRQHRDAEEALDIFETPDLIG
NOV9b   YYRTGKLHCPADVCGPLFEEELAFWGIDETDVEPCCWMTYRQHRDAEEALDIFETPDLIG
NOV9c   YYRTGKLHCPADVCGPLFEEELAFWGIDETDVEPCCWMTYRQHRDAEEALDIFETPDLIG
NOV9d   YYRTGKLHCPADVCGPLFEEELAFWGIDETDVEPCCWMTYRQHRDAEEALDIFETPDLIG
NOV9e   YYRTGKLHCPADVCGPLFEEELAFWGIDETDVEPCCWMTYRQHRDAEEALDIFETPDLIG

NOV9a   GDPGDDEDLAAKRLGIEDAAGLGGPDGKSGRWRRLQPRMWALFEDPYSSRAARFIAFASL
NOV9b   GDPGDDEDLAAKRLGIEDAAGLGGPDGKSGRWRRLQPRMWALFEDPYSSRAARFIAFASL
NOV9c   GDPGDDEDLAAKRLGIEDAAGLGGPDGKSGRWRRLQPRMWALFEDPYSSRAAR-------
NOV9d   GDPGDDEDLAAKRLGIEDAAGLGGPDGKSGRWRRLQPRMWALFEDPYSSRAARFIAFASL
NOV9e   GDPGDDEDLAAKRLGIEDAAGLGGPDGKSGRWRRLQPRMWALFEDPYSSRAARFIAFASL

NOV9a   FFILVSITTFCLETHEAFNIVKNKTEPVINGTSVVLQYEIETDPALTYVEGVCVVWFTFE
NOV9b   FFILVSITTFCLETHEAFNIVKNKTEPVINGTSVVLQYEIETDPALTYVEGVCVVWFTFE
NOV9c   ------------------------------------------------------------
NOV9d   FFILVSITTFCLETHEAFNIVKNKTEPVINGTSVVLQYEIETDPALTYVEGVCVVWFTFE
NOV9e   FFILVSITTFCLETHEAFNIVKNKTEPVINGTSVVLQYEIETDPALTYVEGVCVVWFTFE

NOV9a   FLVRIVFSPNKLEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAAKDVLGFLRVVRFVRIL
NOV9b   FLVRIVFSPNKLEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAAKDVLGFLRVVRFVRIL
NOV9c   ------------------------------------------------------------
NOV9d   FLVRIVFSPNKLEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAAKDVLGFLRVVRFVRIL
NOV9e   FLVRIVFSPNKLEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAAKDVLGFLRVVRFVRIL

NOV9a   RIFKLTRHFVGLRVLGHTLRASTNEFLLLIIFLALGVLIFATMIYYAERVGAQPNDPSAS
NOV9b   RIFKLTRHFVGLRVLGHTLRASTNEFLLLIIFLALGVLIFATMIYYAERVGAQPNDPSAS
NOV9c   ------------------------------------------------------------
```

```
NOV9d   RIFKLTRHFVGLRVLGHTLRASTNEFLLLIIFLALGVLIFATMIYYAERVGAQPNDPSAS
NOV9e   RIFKLTRHFVGLRVLGHTLRASTNEFLLLIIFLALGVLIFATMIYYAERVGAQPNDPSAS

NOV9a   EHTQFKNIPIGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIAMPVPVIVNNF
NOV9b   EHTQFKNIPIGFWWAVVTMTTLGYEDTYPQTWSGMLVGALCALAGVLTIAMPVPVIVNNF
NOV9c   ------------------------------------------------------------
NOV9d   EHTQFKNIPIGFWWAVVTMTTLGYEDTYPQTWSGMLVGALCALAGVLTIAMPVPVIVNNF
NOV9e   EHTQFKNIPIGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIAMPVPVIVNNF

NOV9a   GMYYSLAMAKQKLPRKRKKHIPPAPQASSPTFCKTELNMACNSTQSDTCLGKDNRLLEHN
NOV9b   GMYYSLAMAKQKLPRKRKKHIPPAPQASSPTFCKTELNMACNSTQSDTCLGKDNRLLEHN
NOV9c   ------------------------------------------------------------
NOV9d   GMYYSLAMAKQKLPRKRKKHIPPAPQASSPTFCKTELNMACNSTQSDTCLGKDNRLLEHN
NOV9e   GMYYSLAMAKQKLPRKRKKHIPPAPQASSPTFCKTELNMACNSTQSDTCLGKDNRLLEHN

NOV9a   RSVLSGDDSTGSEPPLSPPERLPIRRSSTRDKNRRGETCFLLTTGDYTCASDGGIRKGYE
NOV9b   RSVLSGDDSTGSEPPLSPPERLPIRRSSTRDKNRRGETCFLLTTGDYTCASDGGIRKGYE
NOV9c   ------------------------------------------------------------
NOV9d   RSVLSGDDSTGSEPPLSPPERLPIRRSSTRDKNRRGETCFLLTTGDYTCASDGGIRKGYE
NOV9e   RSVLSGDDSTGSEPPLSPPERLPIRRSSTRDKNRRGETCFLLTTGDYTCASDGGIRKDNC

NOV9a   KSRSLNNIAGLAGNALRLSPVTSPYNSPCPLRRSRSPIPSIL--
NOV9b   KSRSLNNIAGLAGNALRLSPVTSPYNSPCPLRRSRSPIPSILLE
NOV9c   ----------------------------------------
NOV9d   KSRSLNNIAGLAGNALRLSPVTSPYNSPCPLRRSRSPIPSIL--
NOV9e   KEVVITGYTQAEARSLT--------------------------

NOV9a  (SEQ ID NO:  104)
NOV9b  (SEQ ID NO:  106)
NOV9c  (SEQ ID NO:  108)
NOV9d  (SEQ ID NO:  110)
NOV9e  (SEQ ID NO:  112)
```

Further analysis of the NOV9a protein yielded the following properties shown in Table 9C.

Table 9C. Protein Sequence Properties NOV9a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 9;  pos.chg 2;  neg.chg 2
      H-region:  length 8;  peak value   4.97
      PSG score:   0.57

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.54
      possible cleavage site: between 46 and 47

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood = -6.90   Transmembrane  230 - 246
      INTEGRAL    Likelihood = -3.24   Transmembrane  287 - 303
      INTEGRAL    Likelihood = -2.23   Transmembrane  314 - 330
      INTEGRAL    Likelihood = -0.16   Transmembrane  343 - 359
      INTEGRAL    Likelihood =-13.00   Transmembrane  382 - 398
      INTEGRAL    Likelihood = -7.01   Transmembrane  451 - 467
      PERIPHERAL  Likelihood =  3.61 (at 424)
      ALOM score: -13.00  (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 237
      Charge difference: -0.5   C(-0.5) - N( 0.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  5.56
      Hyd Moment(95):  3.52    G content:       1
      D/E content:     2       S/T content:     0
      Score: -7.90

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: PRKR (4) at  490
      pat4: RKRK (5) at  491
      pat4: KRKK (5) at  492
      pat4: RKKH (3) at  493
      pat7: PRKRKKH (5) at  490
      pat7: PLRRSRS (4) at  626
      bipartite: none
      content of basic residues:  10.7%
```

```
      NLS Score:  1.37

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 89

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       22.2 %: mitochondrial
       11.1 %: nuclear

>> prediction for CG50249-01 is end (k=9)
```

A search of the NOV9a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 9D.

Table 9D. Geneseq Results for NOV9a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV9a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAO14201	Human transporter and ion channel TRICH-18 - Homo sapiens, 638 aa. [WO200204520-A2, 17-JAN-2002]	1..638 1..638	638/638 (100%) 638/638 (100%)	0.0
ABP52157	Human 53763 transporter protein SEQ ID NO:11 - Homo sapiens, 638 aa. [WO200255701-A2, 18-JUL-2002]	1..638 1..638	638/638 (100%) 638/638 (100%)	0.0
ABG70285	Human novel polypeptide #1 - Homo sapiens, 638 aa. [WO200257452-A2, 25-JUL-2002]	1..638 1..638	638/638 (100%) 638/638 (100%)	0.0
ABB78396	Longer splice variant of a human voltage-gated potassium channel - Homo sapiens, 638 aa. [GB2372503-A, 28-AUG-2002]	1..638 1..638	638/638 (100%) 638/638 (100%)	0.0
ABP52167	Rat voltage gated potassium channel protein KV3.2 SEQ ID NO:33 - Rattus norvegicus, 638 aa. [WO200255701-A2, 18-JUL-2002]	1..638 1..638	623/638 (97%) 625/638 (97%)	0.0

In a BLAST search of public sequence databases, the NOV9a protein was found to have homology to the proteins shown in the BLASTP data in Table 9E.

Table 9E. Public BLASTP Results for NOV9a				
Protein Accession Number	**Protein/Organism/Length**	**NOV9a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96PR1	Voltage gated potassium channel Kv3.2b (Potassium voltage-gated potassium channel subfamily C member 2) - Homo sapiens (Human), 638 aa.	1..638 1..638	638/638 (100%) 638/638 (100%)	0.0
P22462	Potassium voltage-gated channel subfamily C member 2 (Potassium channel Kv3.2) (KSHIIIA) - Rattus norvegicus (Rat), 638 aa.	1..638 1..638	623/638 (97%) 625/638 (97%)	0.0
AAO89503	Kv3.2d voltage-gated potassium channel - Homo sapiens (Human), 629 aa.	1..593 1..593	593/593 (100%) 593/593 (100%)	0.0
Q96PR0	Voltage gated potassium channel Kv3.2a - Homo sapiens (Human), 613 aa.	1..593 1..593	593/593 (100%) 593/593 (100%)	0.0
A39402	potassium channel protein IIIA form 1, shaker-type - rat, 613 aa.	1..593 1..593	578/593 (97%) 580/593 (97%)	0.0

PFam analysis indicates that the NOV9a protein contains the domains shown in the Table 9F.

Table 9F. Domain Analysis of NOV9a			
Pfam Domain	**NOV9a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
K_tetra	9..156	50/161 (31%) 121/161 (75%)	3.2e-42
ion_trans	281..472	53/233 (23%) 155/233 (67%)	6.3e-43

Example 10.

The NOV10 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 10A.

Table 10A. NOV10 Sequence Analysis			
NOV10a, CG50307-03 DNA Sequence	SEQ ID NO: 113		1162 bp
	ORF Start: ATG at 97		ORF Stop: TGA at 1087
ATCTACGGAGTCCCTTTGGCCACATAAGATTGGCCTTAAGAGAAGGACGGAGCCACATACTGCTGACG GCCCAGAACTGGCAGAGAGAAGGTTGCC**ATG**GCTGCTGTTGACAGTTTCTACCTCTTGTACTGGGAAA TCGCCAGGTCTTGCAATTGCTATATGGAAGCTCTAGCTTTGGTTGGAGCCTGGTATACGGCCAGAAAA AGCATCACTGTCATCTGTGACTTTTACAGCCTGATCAGGCTGCATTTTATCCCCCGCCTGGGGAGCAG AGCAGACTTGATCAAGCAGTATGGAAGATGGGCCGTTGTCAGCGGTGCAACAGATGGGATTGGAAAAG CCTACGCTGAAGAGTTAGCAAGCCGAGGTCTCAATATAATCCTGATTAGTCGGAACGAGGAGAAGTTG CAGGTTGTTGCTAAAGACATAGCCGACACGTACAAAGTGGAAACTGATATTATAGTTGCGGACTTCAG CAGCGGTCGTGAGATCTACCTTCCAATTCGAGAAGCCCTGAAGGACAAAGACGTTGGCATCTTGGTAA ATAACGTGGGTGTGTTTTATCCCTACCCGCAGTATTTCACTCAGCTGTCCGAGGACAAGCTCTGGGAC ATCATAAATGTGAACATTGCCGCCGCTAGTTTGATGGTCCATGTTGTGTTACCGGGAATGGTGGAGAG AAAGAAAGGTGCCATCGTCACGATCTCTTCTGGCTCCTGCTGCAAACCCACTCCTCAGCTGGCTGCAT TTTCTGCTTCTAAGGCTTATTTAGACCACTTCAGCAGAGCCTTGCAATATGAATATGCCTCTAAAGGA ATCTTTGTACAGAGTCTAATCCCTTTCTATGTAGCCACCAGCATGACAGCACCCAGCAACTTTCTGCA CAGGTGCTCGTGGTTGGTGCCTTCGCCAAAAGTCTATGCACATCATGCTGTTTCTACTCTTGGGATTT CCAAAAGGACCACAGGATATTGGTCCCATTCTATTCAGTTTCTTTTTGCACAGTATATGCCTGAATGG CTCTGGGTGTGGGGAGCAAATATTCTCAACCGTTCACTACGTAAGGAAGCCTTATCCTGCACAGCC**TG** AGTCTGGATGGCCACTTGAGAAGTTTTGCCAACTCCTGGGAACCTCGATATTCTGACATTTGGAAAAA CACATT			
NOV10a, CG50307-03 Protein Sequence	SEQ ID NO: 114	330 aa	MW at 37031.4kD
MAAVDSFYLLYWEIARSCNCYMEALALVGAWYTARKSITVICDFYSLIRLHFIPRLGSRADLIKQYGR WAVVSGATDGIGKAYAEELASRGLNIILISRNEEKLQVVAKDIADTYKVETDIIVADFSSGREIYLPI REALKDKDVGILVNNVGVFYPYPQYFTQLSEDKLWDIINVNIAAASLMVHVVLPGMVERKKGAIVTIS SGSCCKPTPQLAAFSASKAYLDHFSRALQYEYASKGIFVQSLIPFYVATSMTAPSNFLHRCSWLVPSP KVYAHHAVSTLGISKRTTGYWSHSIQFLFAQYMPEWLWVWGANILNRSLRKEALSCTA			
NOV10b, 275624102 DNA Sequence	SEQ ID NO: 115		1012 bp
	ORF Start: at 2		ORF Stop: end of sequence
CACCGGATCCACCATGGCTGCTGTTGACAGTTTCTACCTCTTGTACAGGGAAATCGCCAGGTCTTGCA ATTGCTATATGGAAGCTCTAGCTTTGGTTGGAGCCTGGTATACGGCCAGAAAAAGCATCACTGTCATC TGTGACTTTTACAGCCTGATCAGGCTGCATTTTATCCCCCGCCTGGGGAGCAGAGCAGACTTGATCAA			

GCAGTATGGAAGATGGGCCGTTGTCAGCGGTGCAACAGATGGGATTGGAAAAGCCTACGCTGAAGAGT
TAGCAAGCCGAGGTCTCAATATAATCCTGATTAGTCGGAACGAGGAGAAGTTGCAGGTTGTTGCTAAA
GACATAGCCGACACGTACAAAGTGGAAACTGATATTATAGTTGCGGACTTCAGCAGCGGTCGTGAGAT
CTACCTTCCAATTCGAGAAGCCCTGAAGGACAAAGACGTTGGCATCTTGGTAAATAACGTGGGTGTGT
TTTATCCCTACCCGCAGTATTTCACTCAGCTGTCCGAGGACAAGCTCTGGGACATCATAAATGTGAAC
ATTGCCGCCGCTAGTTTGATGGTCCATGTTGTGTTACCGGGAATGGTGGAGAGAAAGAAAGGTGCCAT
CGTCACGATCTCTTCTGGCTCCTGCTGCAAACCCACTCCTCAGCTGGCTGCATTTTCTGCTTCTAAGG
CTTATTTAGACCACTTCAGCAGAGCCTTGCAATATGAATATGCCTCTAAAGGAATCTTTGTACAGAGT
CTAATCCCTTTCTATGTAGCCACCAGCATGACAGCACCCAGCAACTTTCTGCACAGGTGCTCGTGGTT
GGTGCCTTCGCCAAAAGTCTATGCACATCATGCTGTTTCTACTCTTGGGATTTCCAAAAGGACCACAG
GATATTGGTCCCATTCTATTCAGTTTCTTTTTGCACAGTATATGCCTGAATGGCTCTGGGTGTGGGGA
GCAAATATTCTCAACCGTTCACTACGTAAGGAAGCCTTATCCTGCACAGCCCTCGAGGGC

NOV10b, 275624102 Protein Sequence	SEQ ID NO: 116	337 aa	MW at 37647.0kD

TGSTMAAVDSFYLLYREIARSCNCYMEALALVGAWYTARKSITVICDFYSLIRLHFIPRLGSRADLIK
QYGRWAVVSGATDGIGKAYAEELASRGLNIILISRNEEKLQVVAKDIADTYKVETDIIVADFSSGREI
YLPIREALKDKDVGILVNNVGVFYPYPQYFTQLSEDKLWDIINVNIAAASLMVHVVLPGMVERKKGAI
VTISSGSCCKPTPQLAAFSASKAYLDHFSRALQYEYASKGIFVQSLIPFYVATSMTAPSNFLHRCSWL
VPSPKVYAHHAVSTLGISKRTTGYWSHSIQFLFAQYMPEWLWVWGANILNRSLRKEALSCTALEG

NOV10c, CG50307-01 DNA Sequence	SEQ ID NO: 117	1831 bp
	ORF Start: ATG at 183	ORF Stop: TGA at 1173

ACCGGTTTGGAAGACTTTGCCGGCCTGCAGGACACATGATGACATTGGACCCACCCTCCCCAGCTCGG
AGTCTTTAACTCAGTCACATCTACGGAGTCCCTTTGGCCACATAAGATTGGCCTTAAGAGAAGGACGG
AGCCACATACTGCTGACGGCCCAGAACTGGCAGAGAGAAGGTTGCCATGGCTGCTGTTGACAGTTTCT
ACCTCTTGTACAGGGAAATCGCCAGGTCTTGCAATTGCTATATGGAAGCTCTAGCTTTGGTTGGAGCC
TGGTATACGGCCAGAAAAAGCATCACTGTCATCTGTGACTTTTACAGCCTGATCAGGCTGCATTTTAT
CCCCCGCCTGGGGAGCAGAGCAGACTTGATCAAGCAGTATGGAAGATGGGCCGTTGTCAGCGGTGCAA
CAGATGGGATTGGAAAAGCCTACGCTGAAGAGTTAGCAAGCCGAGGTCTCAATATAATCCTGATTAGT
CGGAACGAGGAGAAGTTGCAGGTTGTTGCTAAAGACATAGCCGACACGTACAAAGTGGAAACTGATAT
TATAGTTGCGGACTTCAGCAGCGGTCGTGAGATCTACCTTCCAATTCGAGAAGCCCTGAAGGACAAAG
ACGTTGGCATCTTGGTAAATAACGTGGGTGTGTTTTATCCCTACCCGCAGTATTTCACTCAGCTGTCC
GAGGACAAGCTCTGGGACATCATAAATGTGAACATTGCCGCCGCTAGTTTGATGGTCCATGTTGTGTT
ACCGGGAATGGTGGAGAGAAAGAAAGGTGCCATCGTCACGATCTCTTCTGGCTCCTGCTGCAAACCCA
CTCCTCAGCTGGCTGCATTTTCTGCTTCTAAGGCTTATTTAGACCACTTCAGCAGAGCCTTGCAATAT
GAATATGCCTCTAAAGGAATCTTTGTACAGAGTCTAATNCCTTTCTATGTAGCCACCAGCATGACAGC
ACCCAGCAACTTTCTGCACAGGTGCTCGTGGTTGGTGCCTTCGCCAAAAGTCTATGCACATCATGCTG
TTTCTACTCTTGGGATTTCCAAAAGGACCACAGGATATTGGTCCCATTCTATTCAGTTTCTTTTTGCA
CAGTATATGCCTGAATGGCTCTGGGTGTGGGGAGCAAATATTCTCAACCGTTCACTACGTAAGGAAGC
CTTATCCTGCACAGCCTGAGTCTGGATGGCCACTTGAGAAGTTTTGCCAACTCCTGGGAACCTCGATA
TTCTGACATTTGGAAAAACACATTTAATTTATCTCCTGTGTTTCATTGCTGATTATTCAGCATACTGT
TGATTCGTCATTTGCAAAACACACATAATACCGTCAGAGTGCTGTGAAAAACCTTAAGGGTGTGTGGA
TGGCACAGGATCAATAATGCCTGAGGCTGATTGACGACATCTACATTTCAGTGCTTTTTCCCTAAGCT
GTTTGAAAGTTACGCTTTTCTGTTGTTCTAGAGCCACAGCAGTCTAATATTGAAATATAATATGATTG
TCAGGTCTTATAATTTCAGATGTTGTTTTTTAAGGGAAATTGACCATTTCACTAGAGGAGTTGTGCTG
GTTTTTACATGTGCATCAAGGAAAGACTACTGGAAAAGTATTTATTTTGGTAACTAAGATTGCTGGCT
ACTATTAGGGACACACTCCGGGCTGTTTGGTATAGCTCTACCTGGTTTGACTATCTGTCATGGAAATG
CTGCCTTCCACTGGTTTTTCCTTTGAGACGGGGTGTGTGCCTGGGTTGTGGGGCCCTTGGGCCCCTTT
TTTTGGTGCCCCTTCTTCCACCCACTTTCGGCCCGCGGGCCCCCTGGCGCTCTGGGTTTCCC

NOV10c, CG50307-01 Protein Sequence	SEQ ID NO: 118	330 aa	MW at 36888.2kD

MAAVDSFYLLYREIARSCNCYMEALALVGAWYTARKSITVICDFYSLIRLHFIPRLGSRADLIKQYGR
WAVVSGATDGIGKAYAEELASRGLNIILISRNEEKLQVVAKDIADTYKVETDIIVADFSSGREIYLPI
REALKDKDVGILVNNVGVFYPYPQYFTQLSEDKLWDIINVNIAAASLMVHVVLPGMVERKKGAIVTIS
SGSCCKPTPQLAAFSASKAYLDHFSRALQYEYASKGIFVQSLXPFYVATSMTAPSNFLHRCSWLVPSP
KVYAHHAVSTLGISKRTTGYWSHSIQFLFAQYMPEWLWVWGANILNRSLRKEALSCTA

NOV10d, CG50307-02 DNA Sequence	SEQ ID NO: 119	1152 bp
	ORF Start: ATG at 97	ORF Stop: TGA at 1087

ATCTACGGAGTCCCTTTGGCCACATAAGATTGGCCTTAAGAGAAGGACGGAGCCACATACTGCTGACG
GCCCAGAACTGGCAGAGAGAAGGTTGCCATGGCTGCTGTTGACAGTTTCTACCTCTTGTACAGGGAAA
TCGCCAGGTCTTGCAATTGCTATATGGAAGCTCTAGCTTTGGTTGGAGCCTGGTATACGGCCAGAAAA
AGCATCACTGTCATCTGTGACTTTTACAGCCTGATCAGGCTGCATTTTATCCCCCGCCTGGGGAGCAG
AGCAGACTTGATCAAGCAGTATGGAAGATGGGCCGTTGTCAGCGGTGCAACAGATGGGATTGGAAAAG
CCTACGCTGAAGAGTTAGCAAGCCGAGGTCTCAATATAATCCTGATTAGTCGGAACGAGGAGAAGTTG
CAGGTTGTTGCTAAAGACATAGCCGACACGTACAAAGTGGAAACTGATATTATAGTTGCGGACTTCAG
CAGCGGTCGTGAGATCTACCTTCCAATTCGAGAAGCCCTGAAGGACAAAGACGTTGGCATCTTGGTAA
ATAACGTGGGTGTGTTTTATCCCTACCCGCAGTATTTCACTCAGCTGTCCGAGGACAAGCTCTGGGAC
ATCATAAATGTGAACATTGCCGCCGCTAGTTTGATGGTCCATGTTGTGTTACCGGGAATGGTGGAGAG
AAAGAAAGGTGCCATCGTCACGATCTCTTCTGGCTCCTGCTGCAAACCCACTCCTCAGCTGGCTGCAT
TTTCTGCTTCTAAGGCTTATTTAGACCACTTCAGCAGAGCCTTGCAATATGAATATGCCTCTAAAGGA
ATCTTTGTACAGAGTCTAATCCCTTTCTATGTAGCCACCAGCATGACAGCACCCAGCAACTTTCTGCA
CAGGTGCTCGTGGTTGGTGCCTTCGCCAAAAGTCTATGCACATCATGCTGTTTCTACTCTTGGGATTT
CCAAAAGGACCACAGGATATTGGTCCCATTCTATTCAGTTTCTTTTTGCACAGTATATGCCTGAATGG
CTCTGGGTGTGGGGAGCAAATATTCTCAACCGTTCACTACGTAAGGAAGCCTTATGCTGCACAGCCTG
AGTCTGGATGGCCACTTGAGAAGTTTTGCCAACTCCTGGGAACCTCGATATTCTGACATTTGGA

NOV10d, CG50307-02 Protein Sequence	SEQ ID NO: 120	330 aa	MW at 37017.4kD

MAAVDSFYLLYREIARSCNCYMEALALVGAWYTARKSITVICDFYSLIRLHFIPRLGSRADLIKQYGR
WAVVSGATDGIGKAYAEELASRGLNIILISRNEEKLQVVAKDIADTYKVETDIIVADFSSGREIYLPI
REALKDKDVGILVNNVGVFYPYPQYFTQLSEDKLWDIINVNIAAASLMVHVVLPGMVERKKGAIVTIS
SGSCCKPTPQLAAFSASKAYLDHFSRALQYEYASKGIFVQSLIPFYVATSMTAPSNFLHRCSWLVPSP
KVYAHHAVSTLGISKRTTGYWSHSIQFLFAQYMPEWLWVWGANILNRSLRKEALCCTA

NOV10e, SNP13375811 of CG50307-03, DNA Sequence	SEQ ID NO: 121	1162 bp
	ORF Start: ATG at 97	ORF Stop: TGA at 1087
	SNP Pos: 1076	SNP Change: C to G

ATCTACGGAGTCCCTTTGGCCACATAAGATTGGCCTTAAGAGAAGGACGGAGCCACATACTGCTGACG
GCCCAGAACTGGCAGAGAGAAGGTTGCCATGGCTGCTGTTGACAGTTTCTACCTCTTGTACTGGGAAA
TCGCCAGGTCTTGCAATTGCTATATGGAAGCTCTAGCTTTGGTTGGAGCCTGGTATACGGCCAGAAAA
AGCATCACTGTCATCTGTGACTTTTACAGCCTGATCAGGCTGCATTTTATCCCCCGCCTGGGGAGCAG
AGCAGACTTGATCAAGCAGTATGGAAGATGGGCCGTTGTCAGCGGTGCAACAGATGGGATTGGAAAAG
CCTACGCTGAAGAGTTAGCAAGCCGAGGTCTCAATATAATCCTGATTAGTCGGAACGAGGAGAAGTTG
CAGGTTGTTGCTAAAGACATAGCCGACACGTACAAAGTGGAAACTGATATTATAGTTGCGGACTTCAG
CAGCGGTCGTGAGATCTACCTTCCAATTCGAGAAGCCCTGAAGGACAAAGACGTTGGCATCTTGGTAA
ATAACGTGGGTGTGTTTTATCCCTACCCGCAGTATTTCACTCAGCTGTCCGAGGACAAGCTCTGGGAC
ATCATAAATGTGAACATTGCCGCCGCTAGTTTGATGGTCCATGTTGTGTTACCGGGAATGGTGGAGAG
AAAGAAAGGTGCCATCGTCACGATCTCTTCTGGCTCCTGCTGCAAACCCACTCCTCAGCTGGCTGCAT
TTTCTGCTTCTAAGGCTTATTTAGACCACTTCAGCAGAGCCTTGCAATATGAATATGCCTCTAAAGGA
ATCTTTGTACAGAGTCTAATCCCTTTCTATGTAGCCACCAGCATGACAGCACCCAGCAACTTTCTGCA
CAGGTGCTCGTGGTTGGTGCCTTCGCCAAAAGTCTATGCACATCATGCTGTTTCTACTCTTGGGATTT
CCAAAAGGACCACAGGATATTGGTCCCATTCTATTCAGTTTCTTTTTGCACAGTATATGCCTGAATGG
CTCTGGGTGTGGGGAGCAAATATTCTCAACCGTTCACTACGTAAGGAAGCCTTATGCTGCACAGCCTG
AGTCTGGATGGCCACTTGAGAAGTTTTGCCAACTCCTGGGAACCTCGATATTCTGACATTTGGAAAAA
CACATT

NOV10e, SNP13375811 of CG50307-03, Protein Sequence	SEQ ID NO: 122	330 aa	MW at 37047.4kD
	SNP Pos: 327		SNP Change: Ser to Cys

MAAVDSFYLLYWEIARSCNCYMEALALVGAWYTARKSITVICDFYSLIRLHFIPRLGSRADLIKQYGR
WAVVSGATDGIGKAYAEELASRGLNIILISRNEEKLQVVAKDIADTYKVETDIIVADFSSGREIYLPI
REALKDKDVGILVNNVGVFYPYPQYFTQLSEDKLWDIINVNIAAASLMVHVVLPGMVERKKGAIVTIS
SGSCCKPTPQLAAFSASKAYLDHFSRALQYEYASKGIFVQSLIPFYVATSMTAPSNFLHRCSWLVPSP
KVYAHHAVSTLGISKRTTGYWSHSIQFLFAQYMPEWLWVWGANILNRSLRKEALCCTA

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 10B.

Table 10B. Comparison of the NOV10 protein sequences.

```
NOV10a  ----MAAVDSFYLLYWEIARSCNCYMEALALVGAWYTARKSITVICDFYSLIRLHFIPRL
NOV10b  TGSTMAAVDSFYLLYREIARSCNCYMEALALVGAWYTARKSITVICDFYSLIRLHFIPRL
NOV10c  ----MAAVDSFYLLYREIARSCNCYMEALALVGAWYTARKSITVICDFYSLIRLHFIPRL
NOV10d  ----MAAVDSFYLLYREIARSCNCYMEALALVGAWYTARKSITVICDFYSLIRLHFIPRL

NOV10a  GSRADLIKQYGRWAVVSGATDGIGKAYAEELASRGLNIILISRNEEKLQVVAKDIADTYK
NOV10b  GSRADLIKQYGRWAVVSGATDGIGKAYAEELASRGLNIILISRNEEKLQVVAKDIADTYK
NOV10c  GSRADLIKQYGRWAVVSGATDGIGKAYAEELASRGLNIILISRNEEKLQVVAKDIADTYK
NOV10d  GSRADLIKQYGRWAVVSGATDGIGKAYAEELASRGLNIILISRNEEKLQVVAKDIADTYK

NOV10a  VETDIIVADFSSGREIYLPIREALKDKDVGILVNNVGVFYPYPQYFTQLSEDKLWDIINV
NOV10b  VETDIIVADFSSGREIYLPIREALKDKDVGILVNNVGVFYPYPQYFTQLSEDKLWDIINV
NOV10c  VETDIIVADFSSGREIYLPIREALKDKDVGILVNNVGVFYPYPQYFTQLSEDKLWDIINV
NOV10d  VETDIIVADFSSGREIYLPIREALKDKDVGILVNNVGVFYPYPQYFTQLSEDKLWDIINV

NOV10a  NIAAASLMVHVVLPGMVERKKGAIVTISSGSCCKPTPQLAAFSASKAYLDHFSRALQYEY
NOV10b  NIAAASLMVHVVLPGMVERKKGAIVTISSGSCCKPTPQLAAFSASKAYLDHFSRALQYEY
NOV10c  NIAAASLMVHVVLPGMVERKKGAIVTISSGSCCKPTPQLAAFSASKAYLDHFSRALQYEY
NOV10d  NIAAASLMVHVVLPGMVERKKGAIVTISSGSCCKPTPQLAAFSASKAYLDHFSRALQYEY

NOV10a  ASKGIFVQSLIPFYVATSMTAPSNFLHRCSWLVPSPKVYAHHAVSTLGISKRTTGYWSHS
NOV10b  ASKGIFVQSLIPFYVATSMTAPSNFLHRCSWLVPSPKVYAHHAVSTLGISKRTTGYWSHS
NOV10c  ASKGIFVQSLXPFYVATSMTAPSNFLHRCSWLVPSPKVYAHHAVSTLGISKRTTGYWSHS
NOV10d  ASKGIFVQSLIPFYVATSMTAPSNFLHRCSWLVPSPKVYAHHAVSTLGISKRTTGYWSHS

NOV10a  IQFLFAQYMPEWLWVWGANILNRSLRKEALSCTA---
NOV10b  IQFLFAQYMPEWLWVWGANILNRSLRKEALSCTALEG
NOV10c  IQFLFAQYMPEWLWVWGANILNRSLRKEALSCTA---
NOV10d  IQFLFAQYMPEWLWVWGANILNRSLRKEALCCTA---

NOV10a  (SEQ ID NO:  114)
NOV10b  (SEQ ID NO:  116)
NOV10c  (SEQ ID NO:  118)
NOV10d  (SEQ ID NO:  120)
```

Further analysis of the NOV10a protein yielded the following properties shown in Table 10C.

Table 10C. Protein Sequence Properties NOV10a	
SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	

```
PSG:  a new signal peptide prediction method
      N-region:  length 5;  pos.chg 0;  neg.chg 1
      H-region:  length 7;  peak value   0.00
```

```
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -6.54
      possible cleavage site: between 58 and 59

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -3.88   Transmembrane  173 - 189
      PERIPHERAL  Likelihood =  2.01 (at 37)
      ALOM score:  -3.88  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 180
      Charge difference:  4.5   C( 2.5) - N(-2.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 1b  (cytoplasmic tail 173 to 330)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  6.52
      Hyd Moment(95):  5.94    G content:       0
      D/E content:     2       S/T content:     1
      Score: -6.71

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  10.0%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none
```

```
NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: too long tail

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       34.8 %: nuclear
       21.7 %: mitochondrial
       21.7 %: cytoplasmic
        8.7 %: vesicles of secretory system
        4.3 %: vacuolar
        4.3 %: peroxisomal
        4.3 %: endoplasmic reticulum

>> prediction for CG50307-03 is nuc (k=23)
```

A search of the NOV10a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 10D.

Table 10D. Geneseq Results for NOV10a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV10a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAE21050	Human drug metabolising enzyme (DME-8) protein - Homo sapiens, 330 aa. [WO200212467-A2, 14-FEB-2002]	1..330 1..330	329/330 (99%) 329/330 (99%)	0.0
AAE21521	Human dehydrogenase DHDR-5 protein - Homo sapiens, 330 aa. [WO200216562-A2, 28-FEB-2002]	1..330 1..330	329/330 (99%) 329/330 (99%)	0.0
ABJ10602	Human novel protein NOV10a SEQ ID NO: 32 - Homo sapiens, 330 aa. [WO200259315-A2, 01-AUG-2002]	1..330 1..330	329/330 (99%) 329/330 (99%)	0.0
AAM41389	Human polypeptide SEQ ID NO 6320 - Homo sapiens, 340 aa. [WO200153312-A1, 26-JUL-2001]	1..330 11..340	329/330 (99%) 329/330 (99%)	0.0
AAM39603	Human polypeptide SEQ ID NO 2748 - Homo sapiens, 330 aa. [WO200153312-A1, 26-JUL-2001]	1..330 1..330	329/330 (99%) 329/330 (99%)	0.0

In a BLAST search of public sequence databases, the NOV10a protein was found to have homology to the proteins shown in the BLASTP data in Table 10E.

Table 10E. Public BLASTP Results for NOV10a				
Protein Accession Number	**Protein/Organism/Length**	**NOV10a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
CAD28991	Sequence 1 from Patent WO0216562 - Homo sapiens (Human), 330 aa.	1..330 1..330	329/330 (99%) 329/330 (99%)	0.0
Q8NC98	Hypothetical protein FLJ90397 - Homo sapiens (Human), 330 aa.	1..330 1..330	328/330 (99%) 328/330 (99%)	0.0
Q9BY22	Steroid dehydrogenase-like protein - Homo sapiens (Human), 309 aa.	22..330 1..309	309/309 (100%) 309/309 (100%)	e-179
Q8WVE5	Steroid dehydrogenase-like - Homo sapiens (Human), 309 aa.	22..330 1..309	308/309 (99%) 308/309 (99%)	e-178
Q8BTX9	Steroid dehydrogenase-like protein homolog - Mus musculus (Mouse), 330 aa.	1..330 1..330	289/330 (87%) 310/330 (93%)	e-173

PFam analysis indicates that the NOV10a protein contains the domains shown in the Table 10F.

Table 10F. Domain Analysis of NOV10a			
Pfam Domain	**NOV10a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
adh_short	66..306	73/275 (27%) 166/275 (60%)	8e-24

Example 11.

The NOV11 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 11A.

Table 11A. NOV11 Sequence Analysis

NOV11a, CG50315-01 DNA Sequence	SEQ ID NO: 123	938 bp
	ORF Start: at 1	ORF Stop: TAG at 931

AATCTCACCAGAGTAACCGAATTCATTCTCATGGGCTTTATGGACCACCCCAAATTGGAGATTCCCCT
CTTTCTGGTGTTTCTGAGTTTCTACCTAGTCACCCTTCTTGGGAATGTGGGGATGATTATGTTAATCC
AAGTAGATGTCAAACTCTACACCCCAATGTACTTCTTCCTGAGCCACCTCTCCCTGCTGGATGCCTGT
TACACCTCAGTCATCACCCCTCAGATCCTAGCCACATTGGCCACAGGCAAAACGGTCATCTCCTACGG
CCACTGTGCTGCCCAGTTCTTTTTATTCACCATCTGTGCAGGCACAGAGTGCTTTCTGCTGGCAGTGA
TGGCCTATGATCGCTATGCTGCCATTCGCAACCCACTGCTCTATACCGTGGCCATGAATCCCAGGCTC
TACTGGAGCCTGGTGGTAGGAGCCTATGTCTGTGGGGTGTCAGGAGCCATCCTGCGTACCACTTGCAC
CTTCACCCTCTCCTTCTGTAAGGACAATCAAATAAACTTCTTCTTCTGTGACCTCCCACCCCTGCTGA
AGCTTGCCTGCAGTGACACAGCAAACATCGAGATTGTCATCATCTTCTTTGGCAATTTTGTGATTTTG
GCCAATGCCTCCGTCATCCTGATTTCCTATCTGCTCATCATCAAGACCATTTTGAAAGTGAAGTCTTC
AGGTGGCAGGGCCAAGACTTTCTCCACATGTGCCTCTCACATCACTGCTGTGGCCCTTTTCTTTGGAG
CCCTTATCTTCATGTATCTGCAAAGTGGCTCAGGCAAATCTCTGGAGGAAGACAAAGTCGTGTCTGTC
TTCTATACAGTGGTCATCCCCATGCTGAACCCTCTGATCTACAGCTTAAGAAACAAAGATGTAAAAGA
CGCCTTCAGAAAGGTCGCTAGGAGACTCCAGGTGTCCCTGAGCATGTAGATTTA

NOV11a, CG50315-01 Protein Sequence	SEQ ID NO: 124	310 aa	MW at 34522.7kD

NLTRVTEFILMGFMDHPKLEIPLFLVFLSFYLVTLLGNVGMIMLIQVDVKLYTPMYFFLSHLSLLDAC
YTSVITPQILATLATGKTVISYGHCAAQFFLFTICAGTECFLLAVMAYDRYAAIRNPLLYTVAMNPRL
YWSLVVGAYVCGVSGAILRTTCTFTLSFCKDNQINFFFCDLPPLLKLACSDTANIEIVIIFFGNFVIL
ANASVILISYLLIIKTILKVKSSGGRAKTFSTCASHITAVALFFGALIFMYLQSGSGKSLEEDKVVSV
FYTVVIPMLNPLIYSLRNKDVKDAFRKVARRLQVSLSM

NOV11b, 207580272 DNA Sequence	SEQ ID NO: 125	822 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCATGATTATGTTAATCCAAGTAGATGTCAAACTCTACACCCCAATGTACTTCTTCCTGAGCCA
CCTCTCCCTGCTGGATGCCTGTTACACCTCAGTCATCACCCCTCAGATCCTAGCCACATTGGCCACAG
GCAAAACGGTCATCTCCTACGGCCACTGTGCTGCCCAGTTCTTTTTATTCACCATCTGTGCAGGCACA
GAGTGCTTTCTGCTGGCAGTGATGGCCTATGATCGCTATGCTGCCATTCGCAACCCACTGCTCTATAC
CGTGGCCATGAATCCCAGGCTCTACTGGAGCCTGGTGGTAGGAGCCTATGTCTGTGGGGTGTCAGGAG
CCATCCTGCGTACCACTTGCACCTTCACCCTCTCCTTCTGTAAGGACAATCAAATAAACTTCTTCTTC
TGTGACCTCCCACCCCTGCTGAAGCTTGCCTGCAGTGACACAGCAAACATCGAGATTGTCATCATCTT
CTTTGGCAATTTTGTGATTTTGGCCAATGCCTCCGTCATCCTGATTTCCTATCTGCTCATCATCAAGA
CCATTTTGAAAGTGAAGTCTTCAGGTGGCAGGGCCAAGACTTTCTCCACATGTGCCTCTCACATCACT
GCTGTGGCCCTTTTCTTTGGAGCCCTTATCTTCATGTATCTGCAAAGTGGCTCAGGCAAATCTCTGGA
GGAAGACAAAGTCGTGTCTGTCTTCTATACAGTGGTCATCCCCATGCTGAACCCTCTGATCTACAGCT
TAAGAAACAAAGATGTAAAAGACGCCTTCAGAAAGGTCGCTAGGAGACTCCAGGTGTTCCTGAGCATG
CTCGAG

NOV11b, 207580272 Protein Sequence	SEQ ID NO: 126	274 aa	MW at 30387.7kD

GSMIMLIQVDVKLYTPMYFFLSHLSLLDACYTSVITPQILATLATGKTVISYGHCAAQFFLFTICAGT
ECFLLAVMAYDRYAAIRNPLLYTVAMNPRLYWSLVVGAYVCGVSGAILRTTCTFTLSFCKDNQINFFF
CDLPPLLKLACSDTANIEIVIIFFGNFVILANASVILISYLLIIKTILKVKSSGGRAKTFSTCASHIT
AVALFFGALIFMYLQSGSGKSLEEDKVVSVFYTVVIPMLNPLIYSLRNKDVKDAFRKVARRLQVFLSM
LE

NOV11c, 314411778 DNA Sequence	SEQ ID NO: 127	964 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCACCATGGCCAAGAATAATCTCACCAGAGTAACCGAATTCATTCTCATGGGCTTTATGG
ACCACCCCAAATTGGAGATTCCCCTCTTTCTGGTGTTTCTGAGTTTCTACCTAGTCACCCTTCTTGGG
AATGTGGGGATGATTATGTTAATCCAAGTAGATGTCAAACTCTACACCCCAATGTACTTCTTCCTGAG
CCACCTCTCCCTGCTGGATGCCTGTTACACCTCAGTCATCACCCCTCAGATCCTAGCCACATTGGCCA

CAGGCAAAACGGTCATCTCCTACGGCCACTGTGCTGCCCAGTTCTTTTTATTCACCATCTGTGCAGGC
ACAGAGTGCTTTCTGCTGGCAGTGATGGCCTATGATCGCTATGCTGCCATTCGCAACCCACTGCTCTA
TACCGTGGCCATGAATCCCAGGCTCTGCTGGAGCCTGGTGGTAGGAGCCTATGTCTGTGGGGTGTCAG
GAGCCATCCTGCGTACCACTTGCACCTTCACCCTCTCCTTCTGTAAGGACAATCAAATAAACTTCTTC
TTCTGTGACCTCCCACCCCTGCTGAAGCTTGCCTGCAGTGACACAGCAAACATCGAGATTGTCATCAT
CTTCTTTGGCAATTTTGTGATTTTGGCCAATGCCTCCGTCATCCTGATTTCCTATCTGCTCATCATCA
AGACCATTTTGAAAGTGAAGTCTTCAGGTGGCAGGGCCAAGACTTTCTCCACATGTGCCTCTCACATC
ACTGCTGTGGCCCTTTTCTTTGGAGCCCTTATCTTCATGTATCTGCAAAGTGGCTCAGGCAAATCTCT
GGAGGAAGACAAAGTCGTGTCTGTCTTCTATACAGTGGTCATCCCCATGCTGAACCCTCTGATCTACA
GCTTAAGAAACAAAGATGTAAAAGACGCCTTCAGAAAGGTCGCTAGGAGACTCCAGGTGTCCCTGAGC
ATGCTCGAGGGC

NOV11c, 314411778 Protein Sequence	SEQ ID NO: 128	321 aa	MW at 35552.9kD

TGSTMAKNNLTRVTEFILMGFMDHPKLEIPLFLVFLSFYLVTLLGNVGMIMLIQVDVKLYTPMYFFLS
HLSLLDACYTSVITPQILATLATGKTVISYGHCAAQFFLFTICAGTECFLLAVMAYDRYAAIRNPLLY
TVAMNPRLCWSLVVGAYVCGVSGAILRTTCTFTLSFCKDNQINFFFCDLPPLLKLACSDTANIEIVII
FFGNFVILANASVILISYLLIIKTILKVKSSGGRAKTFSTCASHITAVALFFGALIFMYLQSGSGKSL
EEDKVVSVFYTVVIPMLNPLIYSLRNKDVKDAFRKVARRLQVSLSMLEG

NOV11d, CG50315-02 DNA Sequence	SEQ ID NO: 129	968 bp
	ORF Start: ATG at 17	ORF Stop: TAG at 959

AGTCCACATTGTCTCC**ATG**GCCAAGAATAATCTCACCAGAGTAACCGAATTCATTCTCATGGGCTTTA
TGGACCACCCCAAATTGGAGATTCCCCTCTTTCTGGTGTTTCTGAGTTTCTACCTAGTCACCCTTCTT
GGGAATGTGGGGATGATTATGTTAATCCAAGTAGATGTCAAACTCTACACCCCAATGTACTTCTTCCT
GAGCCACCTCTCCCTGCTGGATGCCTGTTACACCTCAGTCATCACCCCTCAGATCCTAGCCACATTGG
CCACAGGCAAAACGGTCATCTCCTACGGCCACTGTGCTGCCCAGTTCTTTTTATTCACCATCTGTGCA
GGCACAGAGTGCTTTCTGCTGGCAGTGATGGCCTATGATCGCTATGCTGCCATTCGCAACCCACTGCT
CTATACCGTGGCCATGAATCCCAGGCTCTGCTGGAGCCTGGTGGTAGGAGCCTATGTCTGTGGGGTGT
CAGGAGCCATCCTGCGTACCACTTGCACCTTCACCCTCTCCTTCTGTAAGGACAATCAAATAAACTTC
TTCTTCTGTGACCTCCCACCCCTGCTGAAGCTTGCCTGCAGTGACACAGCAAACATCGAGATTGTCAT
CATCTTCTTTGGCAATTTTGTGATTTTGGCCAATGCCTCCGTCATCCTGATTTCCTATCTGCTCATCA
TCAAGACCATTTTGAAAGTGAAGTCTTCAGGTGGCAGGGCCAAGACTTTCTCCACATGTGCCTCTCAC
ATCACTGCTGTGGCCCTTTTCTTTGGAGCCCTTATCTTCATGTATCTGCAAAGTGGCTCAGGCAAATC
TCTGGAGGAAGACAAAGTCGTGTCTGTCTTCTATACAGTGGTCATCCCCATGCTGAACCCTCTGATCT
ACAGCTTAAGAAACAAAGATGTAAAAGACGCCTTCAGAAAGGTCGCTAGGAGACTCCAGGTGTCCCTG
AGCATG**TAG**ATCTAAG

NOV11d, CG50315-02 Protein Sequence	SEQ ID NO: 130	314 aa	MW at 34907.2kD

MAKNNLTRVTEFILMGFMDHPKLEIPLFLVFLSFYLVTLLGNVGMIMLIQVDVKLYTPMYFFLSHLSL
LDACYTSVITPQILATLATGKTVISYGHCAAQFFLFTICAGTECFLLAVMAYDRYAAIRNPLLYTVAM
NPRLCWSLVVGAYVCGVSGAILRTTCTFTLSFCKDNQINFFFCDLPPLLKLACSDTANIEIVIIFFGN
FVILANASVILISYLLIIKTILKVKSSGGRAKTFSTCASHITAVALFFGALIFMYLQSGSGKSLEEDK
VVSVFYTVVIPMLNPLIYSLRNKDVKDAFRKVARRLQVSLSM

NOV11e, CG50315-03 DNA Sequence	SEQ ID NO: 131	822 bp
	ORF Start: ATG at 7	ORF Stop: at 817

GGATCC**ATG**ATTATGTTAATCCAAGTAGATGTCAAACTCTACACCCCAATGTACTTCTTCCTGAGCCA
CCTCTCCCTGCTGGATGCCTGTTACACCTCAGTCATCACCCCTCAGATCCTAGCCACATTGGCCACAG
GCAAAACGGTCATCTCCTACGGCCACTGTGCTGCCCAGTTCTTTTTATTCACCATCTGTGCAGGCACA
GAGTGCTTTCTGCTGGCAGTGATGGCCTATGATCGCTATGCTGCCATTCGCAACCCACTGCTCTATAC
CGTGGCCATGAATCCCAGGCTCTACTGGAGCCTGGTGGTAGGAGCCTATGTCTGTGGGGTGTCAGGAG
CCATCCTGCGTACCACTTGCACCTTCACCCTCTCCTTCTGTAAGGACAATCAAATAAACTTCTTCTTC
TGTGACCTCCCACCCCTGCTGAAGCTTGCCTGCAGTGACACAGCAAACATCGAGATTGTCATCATCTT
CTTTGGCAATTTTGTGATTTTGGCCAATGCCTCCGTCATCCTGATTTCCTATCTGCTCATCATCAAGA
CCATTTTGAAAGTGAAGTCTTCAGGTGGCAGGGCCAAGACTTTCTCCACATGTGCCTCTCACATCACT
GCTGTGGCCCTTTTCTTTGGAGCCCTTATCTTCATGTATCTGCAAAGTGGCTCAGGCAAATCTCTGGA

GGAAGACAAAGTCGTGTCTGTCTTCTATACAGTGGTCATCCCCATGCTGAACCCTCTGATCTACAGCT TAAGAAACAAAGATGTAAAAGACGCCTTCAGAAAGGTCGCTAGGAGACTCCAGGTGTTCCTGAGCATG CTCGAG

NOV11e, CG50315-03 Protein Sequence	SEQ ID NO: 132	270 aa	MW at 30001.3kD

MIMLIQVDVKLYTPMYFFLSHLSLLDACYTSVITPQILATLATGKTVISYGHCAAQFFLFTICAGTEC FLLAVMAYDRYAAIRNPLLYTVAMNPRLYWSLVVGAYVCGVSGAILRTTCTFTLSFCKDNQINFFFCD LPPLLKLACSDTANIEIVIIFFGNFVILANASVILISYLLIIKTILKVKSSGGRAKTFSTCASHITAV ALFFGALIFMYLQSGSGKSLEEDKVVSVFYTVVIPMLNPLIYSLRNKDVKDAFRKVARRLQVFLSM

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 11B.

Table 11B. Comparison of the NOV11 protein sequences.

```
NOV11a  --------NLTRVTEFILMGFMDHPKLEIPLFLVFLSFYLVTLLGNVGMIMLIQVDVKLY
NOV11b  ----------------------------------------------GSMIMLIQVDVKLY
NOV11c  TGSTMAKNNLTRVTEFILMGFMDHPKLEIPLFLVFLSFYLVTLLGNVGMIMLIQVDVKLY
NOV11d  ----MAKNNLTRVTEFILMGFMDHPKLEIPLFLVFLSFYLVTLLGNVGMIMLIQVDVKLY
NOV11e  ----------------------------------------------MIMLIQVDVKLY

NOV11a  TPMYFFLSHLSLLDACYTSVITPQILATLATGKTVISYGHCAAQFFLFTICAGTECFLLA
NOV11b  TPMYFFLSHLSLLDACYTSVITPQILATLATGKTVISYGHCAAQFFLFTICAGTECFLLA
NOV11c  TPMYFFLSHLSLLDACYTSVITPQILATLATGKTVISYGHCAAQFFLFTICAGTECFLLA
NOV11d  TPMYFFLSHLSLLDACYTSVITPQILATLATGKTVISYGHCAAQFFLFTICAGTECFLLA
NOV11e  TPMYFFLSHLSLLDACYTSVITPQILATLATGKTVISYGHCAAQFFLFTICAGTECFLLA

NOV11a  VMAYDRYAAIRNPLLYTVAMNPRLYWSLVVGAYVCGVSGAILRTTCTFTLSFCKDNQINF
NOV11b  VMAYDRYAAIRNPLLYTVAMNPRLYWSLVVGAYVCGVSGAILRTTCTFTLSFCKDNQINF
NOV11c  VMAYDRYAAIRNPLLYTVAMNPRLCWSLVVGAYVCGVSGAILRTTCTFTLSFCKDNQINF
NOV11d  VMAYDRYAAIRNPLLYTVAMNPRLCWSLVVGAYVCGVSGAILRTTCTFTLSFCKDNQINF
NOV11e  VMAYDRYAAIRNPLLYTVAMNPRLYWSLVVGAYVCGVSGAILRTTCTFTLSFCKDNQINF

NOV11a  FFCDLPPLLKLACSDTANIEIVIIFFGNFVILANASVILISYLLIIKTILKVKSSGGRAK
NOV11b  FFCDLPPLLKLACSDTANIEIVIIFFGNFVILANASVILISYLLIIKTILKVKSSGGRAK
NOV11c  FFCDLPPLLKLACSDTANIEIVIIFFGNFVILANASVILISYLLIIKTILKVKSSGGRAK
NOV11d  FFCDLPPLLKLACSDTANIEIVIIFFGNFVILANASVILISYLLIIKTILKVKSSGGRAK
NOV11e  FFCDLPPLLKLACSDTANIEIVIIFFGNFVILANASVILISYLLIIKTILKVKSSGGRAK

NOV11a  TFSTCASHITAVALFFGALIFMYLQSGSGKSLEEDKVVSVFYTVVIPMLNPLIYSLRNKD
NOV11b  TFSTCASHITAVALFFGALIFMYLQSGSGKSLEEDKVVSVFYTVVIPMLNPLIYSLRNKD
NOV11c  TFSTCASHITAVALFFGALIFMYLQSGSGKSLEEDKVVSVFYTVVIPMLNPLIYSLRNKD
NOV11d  TFSTCASHITAVALFFGALIFMYLQSGSGKSLEEDKVVSVFYTVVIPMLNPLIYSLRNKD
NOV11e  TFSTCASHITAVALFFGALIFMYLQSGSGKSLEEDKVVSVFYTVVIPMLNPLIYSLRNKD

NOV11a  VKDAFRKVARRLQVSLSM---
NOV11b  VKDAFRKVARRLQVFLSMLE-
NOV11c  VKDAFRKVARRLQVSLSMLEG
NOV11d  VKDAFRKVARRLQVSLSM---
NOV11e  VKDAFRKVARRLQVFLSM---

NOV11a  (SEQ ID NO:  124)
NOV11b  (SEQ ID NO:  126)
```

NOV11c (SEQ ID NO: 128) NOV11d (SEQ ID NO: 130) NOV11e (SEQ ID NO: 132)

Further analysis of the NOV11a protein yielded the following properties shown in Table 11C.

Table 11C. Protein Sequence Properties NOV11a

SignalP analysis:	Cleavage site between residues 38 and 39

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 7;  pos.chg 1;  neg.chg 1
      H-region:  length 7;  peak value   6.23
      PSG score:   1.83

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.00
      possible cleavage site: between 37 and 38

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood = -6.58   Transmembrane   21 -  37
      INTEGRAL    Likelihood = -4.09   Transmembrane   97 - 113
      INTEGRAL    Likelihood = -1.65   Transmembrane  138 - 154
      INTEGRAL    Likelihood = -8.70   Transmembrane  202 - 218
      INTEGRAL    Likelihood = -4.14   Transmembrane  240 - 256
      INTEGRAL    Likelihood = -3.13   Transmembrane  269 - 285
      PERIPHERAL  Likelihood =  0.85 (at 183)
      ALOM score:  -8.70  (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 28
      Charge difference:  2.0   C( 0.5) - N(-1.5)
      C > N: C-terminal side will be inside

>>>Caution: Inconsistent mtop result with signal peptide
>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  1.96
      Hyd Moment(95):  9.20    G content:       1
      D/E content:     2       S/T content:     2
      Score: -6.22

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 14  TRV|TE

NUCDISC: discrimination of nuclear localization signals
```

```
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.7%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: LTRV

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       44.4 %: endoplasmic reticulum
       22.2 %: vacuolar
```

```
        11.1 %: Golgi
        11.1 %: vesicles of secretory system
        11.1 %: mitochondrial

>> prediction for CG50315-01 is end (k=9)
```

A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 11D.

5

Table 11D. Geneseq Results for NOV11a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB06632	G protein-coupled receptor GPCR17b protein SEQ ID NO:74 - Homo sapiens, 310 aa. [WO200212343-A2, 14-FEB-2002]	1..310 1..310	310/310 (100%) 310/310 (100%)	e-176
ABB06631	G protein-coupled receptor GPCR17a protein SEQ ID NO:72 - Homo sapiens, 314 aa. [WO200212343-A2, 14-FEB-2002]	1..310 5..314	309/310 (99%) 309/310 (99%)	e-175
AAU95529	Human olfactory and pheromone G protein-coupled receptor #16 - Homo sapiens, 314 aa. [WO200224726-A2, 28-MAR-2002]	1..310 5..314	309/310 (99%) 309/310 (99%)	e-175
ABP95742	Human GPCR polypeptide SEQ ID NO 294 - Homo sapiens, 314 aa. [WO200216548-A2, 28-FEB-2002]	1..310 5..314	309/310 (99%) 309/310 (99%)	e-175
AAG72468	Human OR-like polypeptide query sequence, SEQ ID NO: 2149 - Homo sapiens, 314 aa. [WO200127158-A2, 19-APR-2001]	1..310 5..314	309/310 (99%) 309/310 (99%)	e-175

In a BLAST search of public sequence databases, the NOV11a protein was found to have homology to the proteins shown in the BLASTP data in Table 11E.

Table 11E. Public BLASTP Results for NOV11a				
Protein Accession Number	**Protein/Organism/Length**	**NOV11a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8NGQ6	Seven transmembrane helix receptor - Homo sapiens (Human), 314 aa.	1..310 5..314	309/310 (99%) 309/310 (99%)	e-175
CAD37586	Sequence 201 from Patent WO0224726 - Homo sapiens (Human), 314 aa.	1..310 5..314	308/310 (99%) 308/310 (99%)	e-173
Q8VG66	Olfactory receptor MOR211-1 - Mus musculus (Mouse), 316 aa.	1..307 5..311	253/307 (82%) 279/307 (90%)	e-146
Q8VG65	Olfactory receptor MOR211-2 - Mus musculus (Mouse), 314 aa.	1..310 5..314	246/310 (79%) 278/310 (89%)	e-144
Q96RA8	Olfactory receptor - Homo sapiens (Human), 216 aa (fragment).	64..279 1..216	215/216 (99%) 215/216 (99%)	e-119

PFam analysis indicates that the NOV11a protein contains the domains shown in the Table 11F.

Table 11F. Domain Analysis of NOV11a			
Pfam Domain	**NOV11a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	37..286	55/276 (20%) 174/276 (63%)	1.5e-28

Example 12.

The NOV12 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 12A.

Table 12A. NOV12 Sequence Analysis		
NOV12a, CG50341-01 DNA Sequence	SEQ ID NO: 133	997 bp
	ORF Start: ATG at 21	ORF Stop: TAA at 966
GGCGCTTATAATTTTGAACTATGACCAATCAGACACAGATGATGGAATTCTTGCTTGTGAGATTTACT		

GAGAATTGGGTGCTCCTGAGGCTGCATGCTTTGCTCTTCTCACTGATCTACCTCACGGCTGTGCTGAT
GAATTTAGTCATCATTCTCCTCATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCC
GACATTTGTCCTTCTTAGACCTGTGTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTC
GCCTCCACTGACTCCATCTCCTTCCTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGG
ATCAGAGATTGGCATCCTTACTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACT
GTGAGGCTGTCATGAGCAGAGGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCC
TTGGGACTCTTGTACACAGCTGGAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTT
CTTCTGCGATGTCCCTGCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTG
TGGCCATTGGGGTCTGTTATGCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTC
TCTGCTGTGTTAAGGATATCACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCT
CATTGTTGTCACTGTGTTTCTTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTT
CTATTCTAGACTTGCTGGTGTCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTAC
TGTCTGAAGAACAAGGACATTAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGT
AATGAAAGATGACTAAAGTTGAAGATGGGAAGTACTTTTTTTGNN

NOV12a, CG50341-01 Protein Sequence	SEQ ID NO: 134	315 aa	MW at 34934.2kD

MTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAMYFFLRHLSFLD
LCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDRYAAICCPLHCEAVMSR
GLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVPALLKLTCSKEHAIISVSVAIGVCY
AFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLV
SVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDD

NOV12b, 169475616 DNA Sequence	SEQ ID NO: 135	819 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCCTTCTT
AGACCTGTGTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGACTCCA
TCTCCTTCCTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGGATCAGAGATTGGCATC
CTTACTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTGAGGCTGTCATGAG
CAGAGGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTTGTACA
CAGCTGGAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATGTCCCT
GCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGTGGCCATTGGGGTCTG
TTATGCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTTAAGGA
TATCACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCACTGTG
TTTCTTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTTCTATTCTAGACTTGCT
GGTGTCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAACAAGG
ACATTAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATGACCTC
GAG

NOV12b, 169475616 Protein Sequence	SEQ ID NO: 136	273 aa	MW at 29857.9kD

GSMILDHRLHMAMYFFLRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGI
LTAMSYDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVP
ALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTV
FLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDDL
E

NOV12c, CG50341-02 DNA Sequence	SEQ ID NO: 137	996 bp
	ORF Start: ATG at 21	ORF Stop: TAA at 966

GGCGCTTATAATTTTGAACTATGACCAATCAGACACAGATGATGGAATTCTTGCTTGTGAGATTTACT
GAGAATTGGGTGCTCCTGAGGCTGCATGCTTTGCTCTTCTCACTGATCTACCTCACGGCTGTGCTGAT
GAATTTAGTCATCATTCTCCTCATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCC
GACATTTGTCCTTCTTAGACCTGTGTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTC
GCCTCCACTGACTCCATCTCCTTCCTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGG
ATCAGAGATTGGCATCCTTACTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACT
GTGAGGCTGTCATGAGCAGAGGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCC
TTGGGACTCTTGTACACAGCTGGAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTT
CTTCTGCGATGTCCCTGCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTG

TGGCCATTGGGGTCTGTTATGCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTC
TCTGCTGTGTTAAGGATATCACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCT
CATTGTTGTCACTGTGTTTCTTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTT
CTATTCTAGACTTGCTGGTGTCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTAC
TGTCTGAAGAACAAGGACATTAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGT
AATGAAAGATGAC**TAA**AGATTGAAGATGGGAAGTACTTTTTTTG

NOV12c, CG50341-02 Protein Sequence	SEQ ID NO: 138	315 aa	MW at 34934.2kD

MTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAMYFFLRHLSFLD
LCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDRYAAICCPLHCEAVMSR
GLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVPALLKLTCSKEHAIISVSVAIGVCY
AFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLV
SVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDD

NOV12d, CG50341-03 DNA Sequence	SEQ ID NO: 139	819 bp
	ORF Start: ATG at 7	ORF Stop: at 814

GGATCC**ATG**ATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCCTTCTT
AGACCTGTGTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGACTCCA
TCTCCTTCCTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGGATCAGAGATTGGCATC
CTTACTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTGAGGCTGTCATGAG
CAGAGGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTTGTACA
CAGCTGGAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATGTCCCT
GCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGTGGCCATTGGGGTCTG
TTATGCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTTAAGGA
TATCACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCACTGTG
TTTCTTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTTCTATTCTAGACTTGCT
GGTGTCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAACAAGG
ACATTAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATGACCTC
GAG

NOV12d, CG50341-03 Protein Sequence	SEQ ID NO: 140	269 aa	MW at 29471.5kD

MILDHRLHMAMYFFLRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILT
AMSYDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVPAL
LKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTVFL
VTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDD

NOV12e, CG50341-04 DNA Sequence	SEQ ID NO: 141	957 bp
	ORF Start: ATG at 7	ORF Stop: at 952

GGATCC**ATG**ACCAATCAGACACAGATGATGGAATTCTTGCTTGTGAGATTTACTGAGAATTGGGTGCT
CCTGAGGCTGCATGCTTTGCTCTTCTCACTGATCTACCTCACGGCTGTGCTGATGAATTTAGTCATCA
TTCTCCTCATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCCTTC
TTAGACCTGTGTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGACTC
CATCTCCTTCCTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGGATCAGAGATTGGCA
TCCTTACTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTGAGGCTGTCATG
AGCAGAGGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTTGTA
CACAGCTGGAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATGTCC
CTGCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGTGGCCATTGGGGTC
TGTTATGCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTTAAG
GATATCACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCACTG
TGTTTCTTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTTCTATTCTAGACTTG
CTGGTGTCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAACAA
GGACATTAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATGACC
TCGAG

NOV12e, CG50341-04 Protein Sequence	SEQ ID NO: 142	315 aa	MW at 34934.2kD

MTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAMYFFLRHLSFLD
LCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDRYAAICCPLHCEAVMSR
GLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVPALLKLTCSKEHAIISVSVAIGVCY
AFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLV
SVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDD

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 12B.

Table 12B. Comparison of the NOV12 protein sequences.

```
NOV12a  MTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAMYFF
NOV12b  ----------------------------------------GSMILDHRLHMAMYFF
NOV12c  MTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAMYFF
NOV12d  ----------------------------------------MILDHRLHMAMYFF
NOV12e  MTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAMYFF

NOV12a  LRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDR
NOV12b  LRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDR
NOV12c  LRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDR
NOV12d  LRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDR
NOV12e  LRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDR

NOV12a  YAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVP
NOV12b  YAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVP
NOV12c  YAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVP
NOV12d  YAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVP
NOV12e  YAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVP

NOV12a  ALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCV
NOV12b  ALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCV
NOV12c  ALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCV
NOV12d  ALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCV
NOV12e  ALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCV

NOV12a  PHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKSALS
NOV12b  PHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKSALS
NOV12c  PHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKSALS
NOV12d  PHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKSALS
NOV12e  PHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKSALS

NOV12a  KVLWNVRSSGVMKDD--
NOV12b  KVLWNVRSSGVMKDDLE
NOV12c  KVLWNVRSSGVMKDD--
NOV12d  KVLWNVRSSGVMKDD--
NOV12e  KVLWNVRSSGVMKDD--

NOV12a  (SEQ ID NO:  134)
NOV12b  (SEQ ID NO:  136)
NOV12c  (SEQ ID NO:  138)
```

NOV12d (SEQ ID NO: 140) NOV12e (SEQ ID NO: 142)

Further analysis of the NOV12a protein yielded the following properties shown in Table 12C.

Table 12C. Protein Sequence Properties NOV12a

SignalP analysis:	Cleavage site between residues 47 and 48

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 9;  pos.chg 0;  neg.chg 1
      H-region:  length 4;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.11
      possible cleavage site: between 36 and 37

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   5
      INTEGRAL    Likelihood = -9.87   Transmembrane   32 -  48
      INTEGRAL    Likelihood = -9.66   Transmembrane   90 - 106
      INTEGRAL    Likelihood = -9.71   Transmembrane  198 - 214
      INTEGRAL    Likelihood = -5.79   Transmembrane  239 - 255
      INTEGRAL    Likelihood =  0.16   Transmembrane  267 - 283
      PERIPHERAL  Likelihood =  1.11 (at 56)
      ALOM score:  -9.87  (number of TMSs: 5)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 39
      Charge difference:  1.0   C( 2.5) - N( 1.5)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  2.94
      Hyd Moment(95):  4.71    G content:       0
      D/E content:     2       S/T content:     3
      Score: -5.83

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
```

```
      content of basic residues:    6.7%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: VMKD

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:
      Leucine zipper pattern (PS00029):  *** found ***
         LHALLFSLIYLTAVLMNLVIIL at 24
 none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       44.4 %: endoplasmic reticulum
       11.1 %: mitochondrial
       11.1 %: Golgi
       11.1 %: vacuolar
```

```
       11.1 %: vesicles of secretory system
       11.1 %: cytoplasmic

>> prediction for CG50341-01 is end (k=9)
```

A search of the NOV12a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 12D.

Table 12D. Geneseq Results for NOV12a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU11100	Human novel G protein-coupled receptor, NOV7 - Homo sapiens, 315 aa. [WO200177177-A2, 18-OCT-2001]	1..315 1..315	315/315 (100%) 315/315 (100%)	e-179
AAU85251	G-coupled olfactory receptor #112 - Homo sapiens, 315 aa. [WO200198526-A2, 27-DEC-2001]	1..315 1..315	315/315 (100%) 315/315 (100%)	e-179
AAU95586	Human olfactory and pheromone G protein-coupled receptor #73 - Homo sapiens, 315 aa. [WO200224726-A2, 28-MAR-2002]	1..315 1..315	315/315 (100%) 315/315 (100%)	e-179
AAU24631	Human olfactory receptor AOLFR125 - Homo sapiens, 315 aa. [WO200168805-A2, 20-SEP-2001]	1..315 1..315	315/315 (100%) 315/315 (100%)	e-179
ABP95851	Human GPCR polypeptide SEQ ID NO 512 - Homo sapiens, 314 aa. [WO200216548-A2, 28-FEB-2002]	1..313 1..313	313/313 (100%) 313/313 (100%)	e-178

In a BLAST search of public sequence databases, the NOV12a protein was found to have homology to the proteins shown in the BLASTP data in Table 12E.

Table 12E. Public BLASTP Results for NOV12a				
Protein Accession Number	**Protein/Organism/Length**	**NOV12a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8NGZ2	Seven transmembrane helix receptor - Homo sapiens (Human), 315 aa.	1..315 1..315	315/315 (100%) 315/315 (100%)	e-179
Q96R53	Olfactory receptor - Homo sapiens (Human), 217 aa (fragment).	66..282 1..217	217/217 (100%) 217/217 (100%)	e-120
Q8NHC5	Seven transmembrane helix receptor - Homo sapiens (Human), 309 aa.	1..303 1..302	158/307 (51%) 225/307 (72%)	9e-88
CAD35484	Sequence 23 from Patent WO0208289 - Homo sapiens (Human), 314 aa.	1..303 1..303	155/303 (51%) 211/303 (69%)	4e-87
Q8VF67	Olfactory receptor MOR218-3 - Mus musculus (Mouse), 315 aa.	1..302 5..306	156/302 (51%) 214/302 (70%)	4e-87

PFam analysis indicates that the NOV12a protein contains the domains shown in the Table 12F.

Table 12F. Domain Analysis of NOV12a			
Pfam Domain	**NOV12a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	39..288	57/277 (21%) 170/277 (61%)	1.8e-14

Example 13.

The NOV13 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 13A.

Table 13A. NOV13 Sequence Analysis		
NOV13a, CG50365-01 DNA Sequence	SEQ ID NO: 143	828 bp
	ORF Start: ATG at 16	ORF Stop: TAA at 802
CCACCCCGAGGGACCATGTCGAGGCTCAGCTGGGGATACCGCGAGCACAACGGTCCTATTCACTGGAA GGAATTTTTCCCTATTGCTGATGGTGATCAGCAATCTCCAATTGAGATTAAAACCAAAGAAGTGAAAT		

ATGACTCTTCCCTCCGACCACTTAGTATCAAGTATGACCCAAGCTCAGCTAAAATCATCAGCAACAGC
GGCCATTCCTTCAATGTTGACTTTGATGACACAGAGAACAAATCAGTTCTGCGTGGTGGTCCTCTCAC
TGGAAGCTACAGGTTACGGCAGGTTCACCTTCACTGGGGGTCCGCTGATGACCACGGCTCCGAGCACA
TAGTAGATGGAGTGAGCTATGCTGCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACCCCAGC
TTTGTTGAGGCAGCTCATGAACCAGATGGACTGGCTGTCTTGGGAGTGTTTTTACAGGTGGGTGAACC
TAATTCCCAACTGCAAAAGATTACTGACACTTTGGATTCCATTAAAGAAAAGGGTAAACAAACTCGAT
TCACAAATTTTGACCTATTGTCTCTGCTTCCACCATCCTGGGACTACTGGACATATCCTGGTTCTCTT
ACAGTTCCACCTCTTCTTGAGAGTGTCACATGGATTGTTTTAAAGCAACCTATAAACATCAGCTCTCA
ACAGCTGGCCAAATTTCGCAGTCTCCTGTGCACAGCGGAGGGTGAAGCAGCAGCTTTTCTGGTGAGCA
ATCACCGCCCACCACAGCCTCTAAAGGGCCGCAAAGTGAGAGCCTCTTTCCATTAAAAATTGTCACCA
ATGAACTCCCCC

NOV13a, CG50365-01 Protein Sequence	SEQ ID NO: 144	262 aa	MW at 29428.8kD

MSRLSWGYREHNGPIHWKEFFPIADGDQQSPIEIKTKEVKYDSSLRPLSIKYDPSSAKIISNSGHSFN
VDFDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDHGSEHIVDGVSYAAELHVVHWNSDKYPSFVEAA
HEPDGLAVLGVFLQVGEPNSQLQKITDTLDSIKEKGKQTRFTNFDLLSLLPPSWDYWTYPGSLTVPPL
LESVTWIVLKQPINISSQQLAKFRSLLCTAEGEAAAFLVSNHRPPQPLKGRKVRASFH

NOV13b, 278019595 DNA Sequence	SEQ ID NO: 145	784 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCAGCTGGGGATACCGCGAGCACAACGGTCCTATTCACTGGAAGGAATTTTTCCCTATTG
CTGATGGTGATCAGCAATCTCCAATTGAGATTAAAACCAAAGAAGTGAAATATGACTCTTCCCTCCGA
CCACTTAGTATCAAGTATGACCCAAGCTCAGCTAAAATCATCAGCAACAGCGGCCATTCCTTCAATGT
TGACTTTGATGACACAGAGAACAAATCAGTTCTGCGTGGTGGTCCTCTCACTGGAAGCTACAGGTTAC
GGCAGGTTCACCTTCACTGGGGGTCCGCTGATGACCACGGCTCCGAGCACATAGTAGATGGAGTGAGC
TATGCTGCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACCCCAGCTTTGTTGAGGCAGCTCA
TGAACCAGATGGACTGGCTGTCTTGGGAGTGTTTTTACAGATTGGTGAACCTAATTCCCAACTGCAAA
AGATTACTGACACTTTGGATTCCATTAAAGAAAAGGGTAAACAAACTCGATTCACAAATTTTGACCTA
TTGTCTCTGCTTCCACCATCCTGGGACTACTGGACATATCCTGGTTCTCTTACAGTTCCACCTCTTCT
TGAGAGTGTCACATGGATTGTTTTAAAGCAACCTATAAACATCAGCTCTCAACAGCTGGCCAAATTTC
GCAGTCTCCTGTGCACAGCGGAGGGTGAAGCAGCAGCTTTTCTGGTGAGCAATCACCGCCCACCACAG
CCTCTAAAGGGCCGCAAAGTGAGAGCCCTCGAGGGC

NOV13b, 278019595 Protein Sequence	SEQ ID NO: 146	261 aa	MW at 29128.4kD

TGSSWGYREHNGPIHWKEFFPIADGDQQSPIEIKTKEVKYDSSLRPLSIKYDPSSAKIISNSGHSFNV
DFDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDHGSEHIVDGVSYAAELHVVHWNSDKYPSFVEAAH
EPDGLAVLGVFLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNFDLLSLLPPSWDYWTYPGSLTVPPLL
ESVTWIVLKQPINISSQQLAKFRSLLCTAEGEAAAFLVSNHRPPQPLKGRKVRALEG

NOV13c, CG50365-02 DNA Sequence	SEQ ID NO: 147	833 bp
	ORF Start: ATG at 21	ORF Stop: TAA at 807

ATGTCGAGGCTCAGCTGGGGATGTCGAGGCTCAGCTGGGGATACCGCGAGCACAACGGTCCTATTCAC
TGGAAGGAATTTTTCCCTATTGCTGATGGTGATCAGCAATCTCCAATTGAGATTAAAACCAAAGAAGT
GAAATATGACTCTTCCCTCCGACCACTTAGTATCAAGTATGACCCAAGCTCAGCTAAAATCATCAGCA
ACAGCGGCCATTCCTTCAATGTTGACTTTGATGACACAGAGAACAAATCAGTTCTGCGTGGTGGTCCT
CTCACTGGAAGCTACAGGTTACGGCAGGTTCACCTTCACTGGGGGTCCGCTGATGACCACGGCTCCGA
GCACATAGTAGATGGAGTGAGCTATGCTGCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACC
CCAGCTTTGTTGAGGCAGCTCATGAACCAGATGGACTGGCTGTCTTGGGAGTGTTTTTACAGATTGGT
GAACCTAATTCCCAACTGCAAAAGATTACTGACACTTTGGATTCCATTAAAGAAAAGGGTAAACAAAC
TCGATTCACAAATTTTGACCTATTGTCTCTGCTTCCACCATCCTGGGACTACTGGACATATCCTGGTT
CTCTTACAGTTCCACCTCTTCTTGAGAGTGTCACATGGATTGTTTTAAAGCAACCTATAAACATCAGC
TCTCAACAGCTGGCCAAATTTCGCAGTCTCCTGTGCACAGCGGAGGGTGAAGCAGCAGCTTTTCTGGT
GAGCAATCACCGCCCACCACAGCCTCTAAAGGGCCGCAAAGTGAGAGCCTCTTTCCATTAAAAATTGT
CACCAATGAACTCCCCC

NOV13c, CG50365-02 Protein Sequence	SEQ ID NO: 148	262 aa	MW at 29442.8kD

MSRLSWGYREHNGPIHWKEFFPIADGDQQSPIEIKTKEVKYDSSLRPLSIKYDPSSAKIISNSGHSFN
VDFDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDHGSEHIVDGVSYAAELHVVHWNSDKYPSFVEAA
HEPDGLAVLGVFLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNFDLLSLLPPSWDYWTYPGSLTVPPL
LESVTWIVLKQPINISSQQLAKFRSLLCTAEGEAAAFLVSNHRPPQPLKGRKVRASFH

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 13B.

Table 13B. Comparison of the NOV13 protein sequences.

```
NOV13a  MSRLSWGYREHNGPIHWKEFFPIADGDQQSPIEIKTKEVKYDSSLRPLSIKYDPSSAKII
NOV13b  -TGSSWGYREHNGPIHWKEFFPIADGDQQSPIEIKTKEVKYDSSLRPLSIKYDPSSAKII
NOV13c  MSRLSWGYREHNGPIHWKEFFPIADGDQQSPIEIKTKEVKYDSSLRPLSIKYDPSSAKII

NOV13a  SNSGHSFNVDFDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDHGSEHIVDGVSYAAELH
NOV13b  SNSGHSFNVDFDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDHGSEHIVDGVSYAAELH
NOV13c  SNSGHSFNVDFDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDHGSEHIVDGVSYAAELH

NOV13a  VVHWNSDKYPSFVEAAHEPDGLAVLGVFLQVGEPNSQLQKITDTLDSIKEKGKQTRFTNF
NOV13b  VVHWNSDKYPSFVEAAHEPDGLAVLGVFLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNF
NOV13c  VVHWNSDKYPSFVEAAHEPDGLAVLGVFLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNF

NOV13a  DLLSLLPPSWDYWTYPGSLTVPPLLESVTWIVLKQPINISSQQLAKFRSLLCTAEGEAAA
NOV13b  DLLSLLPPSWDYWTYPGSLTVPPLLESVTWIVLKQPINISSQQLAKFRSLLCTAEGEAAA
NOV13c  DLLSLLPPSWDYWTYPGSLTVPPLLESVTWIVLKQPINISSQQLAKFRSLLCTAEGEAAA

NOV13a  FLVSNHRPPQPLKGRKVRASFH
NOV13b  FLVSNHRPPQPLKGRKVRALEG
NOV13c  FLVSNHRPPQPLKGRKVRASFH

NOV13a  (SEQ ID NO:  144)
NOV13b  (SEQ ID NO:  146)
NOV13c  (SEQ ID NO:  148)
```

Further analysis of the NOV13a protein yielded the following properties shown in Table 13C.

Table 13C. Protein Sequence Properties NOV13a

SignalP analysis:	No Known Signal Sequence Predicted

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 10;  pos.chg 2;  neg.chg 1
      H-region:  length 7;  peak value  -7.23
      PSG score: -11.62

GvH:  von Heijne's method for signal seq. recognition
```

```
      GvH score (threshold: -2.1):  -8.81
      possible cleavage site: between 57 and 58

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  2.97 (at 197)
      ALOM score:   2.97  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       2       Hyd Moment(75):  8.62
      Hyd Moment(95):  7.75    G content:       2
      D/E content:     2       S/T content:     2
      Score: -5.03

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 13  SRL|SW

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: PLKGRKV (4) at  251
      bipartite: none
      content of basic residues:  10.3%
      NLS Score: -0.13

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: SRLS

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none
```

```
Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 55.5

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

-------------------------

Final Results (k = 9/23):

       43.5 %: cytoplasmic
       34.8 %: mitochondrial
       21.7 %: nuclear

>> prediction for CG50365-01 is cyt (k=23)
```

A search of the NOV13a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 13D.

Table 13D. Geneseq Results for NOV13a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV13a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAE25377	Human NZMS-1 protein - Homo sapiens, 247 aa. [WO200246385-A2, 13-JUN-2002]	1..262 1..247	246/262 (93%) 247/262 (93%)	e-143
AAU19418	Human diagnostic and therapeutic polypeptide (DITHP) #4 - Homo sapiens, 274 aa. [WO200162927-A2, 30-AUG-2001]	1..262 13..273	246/262 (93%) 248/262 (93%)	e-143
ABB08900	Human lyase HLYA-1 protein - Homo sapiens, 242 aa. [WO200200840-A2, 03-JAN-2002]	1..242 1..242	241/242 (99%) 242/242 (99%)	e-143
AAB63110	Human secreted protein sequence encoded by gene 27 SEQ ID NO:120 - Homo sapiens, 184 aa. [WO200061748-A1, 19-OCT-2000]	78..261 1..184	181/184 (98%) 183/184 (99%)	e-105
AAO15236	Human carbonic anhydrase I (Cln115) protein - Homo sapiens, 261 aa. [US2002042088-A1, 11-APR-2002]	1..261 1..261	157/261 (60%) 202/261 (77%)	5e-94

In a BLAST search of public sequence databases, the NOV13a protein was found to have homology to the proteins shown in the BLASTP data in Table 13E.

Table 13E. Public BLASTP Results for NOV13a				
Protein Accession Number	**Protein/Organism/Length**	**NOV13a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8N1Q1	Carbonic anhydrase XIII (EC 4.2.1.1) (Carbonate dehydratase XIII) (CA- XIII) - Homo sapiens (Human), 262 aa.	1..262 1..262	261/262 (99%) 262/262 (99%)	e-155
Q9D6N1	Carbonic anhydrase XIII (EC 4.2.1.1) (Carbonate dehydratase XIII) (CA- XIII) - Mus musculus (Mouse), 262 aa.	1..262 1..262	238/262 (90%) 250/262 (94%)	e-141
Q8HY33	Carbonic anhydrase 1 - Monodelphis domestica (Short-tailed grey opossum), 262 aa.	1..261 1..261	164/261 (62%) 210/261 (79%)	e-99
Q8JG56	Erythrocyte carbonic anhydrase - Lepisosteus osseus (Long-nosed gar), 261 aa.	5..261 4..260	165/257 (64%) 198/257 (76%)	1e-95
Q8UWA5	Carbonic anhydrase 2 (EC 4.2.1.1) (Carbonate dehydratase) - Tribolodon hakonensis, 260 aa.	6..261 5..259	164/256 (64%) 200/256 (78%)	8e-95

PFam analysis indicates that the NOV13a protein contains the domains shown in the Table 13F.

Table 13F. Domain Analysis of NOV13a			
Pfam Domain	**NOV13a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
carb_anhydrase	6..261	172/283 (61%) 221/283 (78%)	1.9e-162

Example 14.

The NOV14 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 14A.

Table 14A. NOV14 Sequence Analysis		
NOV14a, CG50367-01 DNA Sequence	SEQ ID NO: 149	2762 bp
	ORF Start: ATG at 3	ORF Stop: TGA at 2745

CTATGGGCTGGAGGCCCCGGAGAGCTCGGGGGACCCCGTTGCTGCTGCTGCTACTACTGCTGCTGCTC
TGGCCAGTGCCAGGCGCCGGGGTGCTTCAAGGACATATCCCTGGGCAGCCAGTCACCCCGCACTGGGT
CCTGGATGGACAACCCTGGCGCACCGTCAGCCTGGAGGAGCCGGTCTCGAAGCCAGACATGGGGCTGG
TGGCCCTGGAGGCTGAAGGCCAGGAGCTCCTGCTTGAGCTGGAGAAGAACCACAGGCTGCTGGCCCCA
GGATACATAGAAACCCACTACGGCCCAGATGGGCAGCCAGTGGTGCTGGCCCCCAACCACACGGATCA
TTGCCACTACCAAGGGCGAGTAAGGGGCTTCCCCGACTCCTGGGTAGTCCTCTGCACCTGCTCTGGGA
TGAGTGGCCTGATCACCCTCAGCAGGAATGCCAGCTATTATCTGCGTCCCTGGCCACCCCGGGGCTCC
AAGGACTTCTCAACCCACGAGATCTTTCGGATGGAGCAGCTGCTCACCTGGAAAGGAACCTGTGGCCA
CAGGGATCCTGGGAACAAAGCGGGCATGACCAGCCTTCCTGGTGGTCCCCAGAGCAGGGGCAGGCGAG
AAGCGCGCAGGACCCGGAAGTACCTGGAACTGTACATTGTGGCAGACCACACCCTGTTCTTGACTCGG
CACCGAAACTTGAACCACACCAAACAGCGTCTCCTGGAAGTCGCCAACTACGTGGACCAGCTTCTCAG
GACTCTGGACATTCAGGTGGCGCTGACCGGCCTGGAGGTGTGGACCGAGCGGGACCGCAGCCGCGTCA
CGCAGGACGCCAACGCCACGCTCTGGGCCTTCCTGCAGTGGCGCCGGGGGCTGTGGGCGCAGCGGCCC
CACGACTCCGCGCAGCTGCTCACGGGCCGCGCCTTCCAGGGCGCCACAGTGGGCCTGGCGCCCGTCGA
GGGCATGTGCCGCGCCGAGAGCTCGGGAGGCGTGAGCACGGACCACTCGGAGCTCCCCATCGGCGCCG
CAGCCACCATGGCCCATGAGATCGGCCACAGCCTCGGCCTCAGCCACGACCCCGACGGCTGCTGCGTG
GAGGCTGCGGCCGAGTCCGGAGGCTGCGTCATGGCTGCGGCCACCGGGCACCCGTTTCCGCGCGTGTT
CAGCGCCTGCAGCCGCCGCCAGCTGCGCGCCTTCTTCCGCAAGGGGGGCGGCGCTTGCCTCTCCAATG
CCCCGGACCCCGGACTCCCGGTGCCGCCGGCGCTCTGCGGGAACGGCTTCGTGGAAGCGGGCGAGGAG
TGTGACTGCGGCCCTGGCCAGGAGTGCCGCGACCTCTGCTGCTTTGCTCACAACTGCTCGCTGCGCCC
GGGGGCCCAGTGCGCCCACGGGGACTGCTGCGTGCGCTGCCTGCTGAAGCCGGCTGGAGCGCTGTGCC
GCCAGGCCATGGGTGACTGTGACCTCCCTGAGTTTTGCACGGGCACCTCCTCCCACTGTCCCCCAGAC
GTTTACCTACTGGACGGCTCACCCTGTGCCAGGGGCAGTGGCTACTGCTGGGATGGCGCATGTCCCAC
GCTGGAGCAGCAGTGCCAGCAGCTCTGGGGGCCTGGCTCCCACCCAGCTCCCGAGGCCTGTTTCCAGG
TGGTGAACTCTGCGGGAGATGCTCATGGAAACTGCGGCCAGGACAGCGAGGGCCACTTCCTGCCCTGT
GCAGGGAGGGATGCCCTGTGTGGGAAGCTGCAGTGCCAGGGTGGAAAGCCCAGCCTGCTCGCACCGCA
CATGGTGCCAGTGGACTCTACCGTTCACCTAGATGGCCAGGAAGTGACTTGTCGGGGAGCCTTGGCAC
TCCCCAGTGCCCAGCTGGACCTGCTTGGCCTGGGCCTGGTAGAGCCAGGCACCCAGTGTGGACCTAGA
ATGGTGTGCCAGAGCAGGCGCTGCAGGAAGAATGCCTTCCAGGAGCTTCAGCGCTGCCTGACTGCCTG
CCACAGCCACGGGGTTTGCAATAGCAACCATAACTGCCACTGTGCTCCAGGCTGGGCTCCACCCTTCT
GTGACAAGCCAGGCTTTGGTGGCAGCATGGACAGTGGCCCTGTGCAGGCTGAAAACCATGACACCTTC
CTGCTGGCCATGCTCCTCAGCGTCCTGCTGCCTCTGCTCCCAGGGGCCGGCCTGGCCTGGTGTTGCTA
CCGACTCCCAGGAGCCCATCTGCAGCGATGCAGCTGGGGCTGCAGAAGGGACCCTGCGTGCAGTGGCC
CCAAAGATGGCCCACACAGGGACCACCCCCTGGGCGGCGTTCACCCCATGGAGTTGGGCCCCACAGCC
ACTGGACAGCCCTGGCCCCTGGCCCCAGGGGCTCCTGCTGACCATATTCACAACATTTACCCTCCACC
ATTTCTCCCAGACCCTGAGAACTCTCATGAGCCCAGCAGCCACCCTGAGAAGCCTCTGCCAGCAGTCT
CGCCTGACCCCCAAGGTGGTTCCCTTGCAGCCTGGGGCCCCAGTCCTTTAGGGGACAACATATCCTCC
TCATTCTCAGCAGATCAAGTCCAGATGCCAAGATCCTGCCTCTGTGGCGAACCCTGGGGAGGCCACGT
GGGAAGGAAAGAGGGCTCTAAGAGGGGAGGCCCCAGACTGGGGGAGAGGCCTGTCTGGAGCCCAGGAT
CACCTGGCTGTGCTGCAGAACTGGAGAAGAGAAGCTCAGCAGAAAGGAGCTGGCATGGGGCCAACAGC
AGAAAAGCAGGAGGCACGCAGAAGTGACTGGGAAGCAGGAGG

NOV14a, CG50367-01 Protein Sequence	SEQ ID NO: 150	914 aa	MW at 98055.2kD

MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPVSKPDMGLV
ALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRVRGFPDSWVVLCTCSGM
SGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCGHRDPGNKAGMTSLPGGPQSRGRRE
ARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEVANYVDQLLRTLDIQVALTGLEVWTERDRSRVT
QDANATLWAFLQWRRGLWAQRPHDSAQLLTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAA
ATMAHEIGHSLGLSHDPDGCCVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNA
PDPGLPVPPALCGNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKPAGALCR
QAMGDCDLPEFCTGTSSHCPPDVYLLDGSPCARGSGYCWDGACPTLEQQCQQLWGPGSHPAPEACFQV
VNSAGDAHGNCGQDSEGHFLPCAGRDALCGKLQCQGGKPSLLAPHMVPVDSTVHLDGQEVTCRGALAL

PSAQLDLLGLGLVEPGTQCGPRMVCQSRRCRKNAFQELQRCLTACHSHGVCNSNHNCHCAPGWAPPFC
DKPGFGGSMDSGPVQAENHDTFLLAMLLSVLLPLLPGAGLAWCCYRLPGAHLQRCSWGCRRDPACSGP
KDGPHRDHPLGGVHPMELGPTATGQPWPLAPGAPADHIHNIYPPPFLPDPENSHEPSSHPEKPLPAVS
PDPQGGSLAAWGPSPLGDNISSSFSADQVQMPRSCLCGEPWGGHVGRKEGSKRGGPRLGERPVWSPGS
PGCAAELEKRSSAERSWHGANSRKAGGTQK

NOV14b, CG50367-05 DNA Sequence	SEQ ID NO: 151	3397 bp
	ORF Start: ATG at 37	ORF Stop: TGA at 2398

GCGAGCCGCTGCCTAGAGGCCGAGGAGCTCACAGCTATGGGCTGGAGGCCCCGGAGAGCTCGGGGGAC
CCCGTTGCTGCTGCTGCTACTACTGCTGCTGCTCTGGCCAGTGCCAGGCGCCGGGGTGCTTCAAGGAC
ATATCCCTGGGCAGCCAGTCACCCCGCACTGGGTCCTGGATGGACAACCCTGGCGCACCGTCAGCCTG
GAGGAGCCGGTCTCGAAGCCAGACATGGGGCTGGTGGCCCTGGAGGCTGAAGGCCAGGAGCTCCTGCT
TGAGCTGGAGAAGAACCACAGGCTGCTGGCCCCAGGATACATAGAAACCCACTACGGCCCAGATGGGC
AGCCAGTGGTGCTGGCCCCCAACCACACGGATCATTGCCACTACCAAGGGCGAGTAAGGGGCTTCCCC
GACTCCTGGGTAGTCCTCTGCACCTGCTCTGGGATGAGTGGCCTGATCACCCTCAGCAGGAATGCCAG
CTATTATCTGCGTCCCTGGCCACCCCGGGGCTCCAAGGACTTCTCAACCCACGAGATCTTTCGGATGG
AGCAGCTGCTCACCTGGAAAGGAACCTGTGGCCACAGGGATCCTGGGAACAAAGCGGGCATGACCAGC
CTTCCTGGTGGTCCCCAGAGCAGGGGCAGGCGAGAAGCGCGCAGGACCCGGAAGTACCTGGAACTGTA
CATTGTGGCAGACCACACCCTGTTCTTGACTCGGCACCGAAACTTGAACCACACCAAACAGCGTCTCC
TGGAAGTCGCCAACTACGTGGACCAGCTTCTCAGGACTCTGGACATTCAGGTGGCGCTGACCGGCCTG
GAGGTGTGGACCGAGCGGGACCGCAGCCGCGTCACGCAGGACGCCAACGCCACGCTCTGGGCCTTCCT
GCAGTGGCGCCGGGGGCTGTGGGCGCAGCGGCCCCACGACTCCGCGCAGCTGCTCACGGGCCGCGCCT
TCCAGGGCGCCACAGTGGGCCTGGCGCCCGTCGAGGGCATGTGCCGCGCCGAGAGCTCGGGAGGCGTG
AGCACGGACCACTCGGAGCTCCCCATCGGCGCCGCAGCCACCATGGCCCATGAGATCGGCCACAGCCT
CGGCCTCAGCCACGACCCCGACGGCTGCTGCGTGGAGGCTGCGGCCGAGTCCGGAGGCTGCGTCATGG
CTGCGGCCACCGGGCACCCGTTTCCGCGCGTGTTCAGCGCCTGCAGCCGCCGCCAGCTGCGCGCCTTC
TTCCGCAAGGGGGGCGGCGCTTGCCTCTCCAATGCCCCGGACCCCGGACTCCCGGTGCCGCCGGCGCT
CTGCGGGAACGGCTTCGTGGAAGCGGGCGAGGAGTGTGACTGCGGCCCTGGCCAGGAGTGCCGCGACC
TCTGCTGCTTTGCTCACAACTGCTCGCTGCGCCCGGGGGCCCAGTGCGCCCACGGGGACTGCTGCGTG
CGCTGCCTGCTGAAGCCGGCTGGAGCGCTGTGCCGCCAGGCCATGGGTGACTGTGACCTCCCTGAGTT
TTGCACGGGCACCTCCTCCCACTGTCCCCCAGACGTTTACCTACTGGACGGCTCACCCTGTGCCAGGG
GCAGTGGCTACTGCTGGGATGGCGCATGTCCCACGCTGGAGCAGCAGTGCCAGCAGCTCTGGGGGCCT
GGCTCCCACCCAGCTCCCGAGGCCTGTTTCCAGGTGGTGAACTCTGCGGGAGATGCTCATGGAAACTG
CGGCCAGGACAGCGAGGGCCACTTCCTGCCCTGTGCAGGGAGGGATGCCCTGTGTGGGAAGCTGCAGT
GCCAGGGTGGAAAGCCCAGCCTGCTCGCACCGCACATGGTGCCAGTGGACTCTACCGTTCACCTAGAT
GGCCAGGAAGTGACTTGTCGGGGAGCCTTGGCACTCCCCAGTGCCCAGCTGGACCTGCTTGGCCTGGG
CCTGGTAGAGTCAGGCACCCAGTGTGGACCTAGAATGGTTTGCAATAGCAACCATAACTGCCACTGTG
CTCCAGGCTGGGCTCCACCCTTCTGTGACAAGCCAGGCTTTGGTGGCAGCATGGACAGTGGCCCTGTG
CAGGCTGAAAACCATGACACCTTCCTGCTGGCCATGCTCCTCAGCGTCCTGCTGCCTCTGCTCCCAGG
GGCCGGCCTGGCCTGGTGTTGCTACCGACTCCCAGGAGCCCATCTGCAGCGATGCAGCTGGGGCTGCA
GAAGGGACCCTGCGTGCAGTGGCCCCAAAGATGGCCCACACAGGGACCACCCCCTGGGCGGCGTTCAC
CCCATGGAGTTGGGCCCCACAGCCACTGGACAGCCCTGGCCCCTGGACCCTGAGAACTCTCATGAGCC
CAGCAGCCACCCTGAGAAGCCTCTGCCAGCAGTCTCGCCCGACCCCCAAGCAGATCAAGTCCAGATGC
CAAGATCCTGCCTCTGGTGAGAGGTAGCTCCTAAAATGAACAGATTTAAAGACAGGTGGCCACTGACA
GCCACTCCAGGAACTTGAACTGCAGGGGCAGAGCCAGTGAATCACCGGACCTCCAGCACCTGCAGGCA
GCTTGGAAGTTTCTTCCCCGAGTGGAGCTTCGACCCACCCACTCCAGGAACCCAGAGCCACACTAGAA
GTTCCTGAGGGCTGGAGAACACTGCTGGGCACACTCTCCAGCTCAATAAACCATCAGTCCCAGAAGCA
AAGGTCACACAGCCCCTGACCTCCCTCACCAGTGGAGGCTGGGTAGTGCTGGCCATCCCAAAAGGGCT
CTGTCCTGGGAGTCTGGTGTGTCTCCTACATGCAATTTCCACGGACCCAGCTCTGTGGAGGGCATGAC
TGCTGGCCAGAAGCTAGTGGTCCTGGGGCCCTATGGTTCGACTGAGTCCACACTCCCCTGCAGCCTGG
CTGGCCTCTGCAAACAAACATAATTTTGGGGACCTTCCTTCCTGTTTCTTCCCACCCTGTCTTCTCCC
CTAGGTGGTTCCTGAGCCCCCACCCCCAATCCCAGTGCTACACCTGAGGTTCTGGAGCTCAGAATCTG
ACAGCCTCTCCCCCATTCTGTGTGTGTCGGGGGGACAGAGGGAACCATTTAAGAAAAGATACCAAAGT
AGAAGTCAAAAGAAAGACATGTTGGCTATAGGCGTGGTGGCTCATGCCTATAATCCCAGCACTTTGGG
AAGCCGGGGTAGGAGGATCACCAGAGGCCAGCAGGTCCACACCAGCCTGGGCAACACAGCAAGACACC
GCATCTACAGAAAAATTTTAAAATTAGCTGGGCGTGGTGGTGTGTACCTGTAGGCCTAGCTGCTCAGG
AGGCTGAAGCAGGAGGATCACTTGAGCCTGAGTTCAACACTGCAGTGAGCTATGGTGGCACCACTGCA
CTCCAGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAAATAAATTTTAAAAAGACATATTACACT

NOV14b, CG50367-05 Protein Sequence	SEQ ID NO: 152	787 aa	MW at 84713.5kD

MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPVSKPDMGLV
ALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRVRGFPDSWVVLCTCSGM
SGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCGHRDPGNKAGMTSLPGGPQSRGRRE
ARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEVANYVDQLLRTLDIQVALTGLEVWTERDRSRVT
QDANATLWAFLQWRRGLWAQRPHDSAQLLTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAA
ATMAHEIGHSLGLSHDPDGCCVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNA
PDPGLPVPPALCGNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKPAGALCR
QAMGDCDLPEFCTGTSSHCPPDVYLLDGSPCARGSGYCWDGACPTLEQQCQQLWGPGSHPAPEACFQV
VNSAGDAHGNCGQDSEGHFLPCAGRDALCGKLQCQGGKPSLLAPHMVPVDSTVHLDGQEVTCRGALAL
PSAQLDLLGLGLVESGTQCGPRMVCNSNHNCHCAPGWAPPFCDKPGFGGSMDSGPVQAENHDTFLLAM
LLSVLLPLLPGAGLAWCCYRLPGAHLQRCSWGCRRDPACSGPKDGPHRDHPLGGVHPMELGPTATGQP
WPLDPENSHEPSSHPEKPLPAVSPDPQADQVQMPRSCLW

NOV14c, CG50367-06 DNA Sequence	SEQ ID NO: 153	3351 bp
	ORF Start: ATG at 37	ORF Stop: TGA at 1579

GCGAGCCGCTGCCTAGAGGCCGAGGAGCTCACAGCT**ATG**GGCTGGAGGCCCCGGAGAGCTCGGGGGAC
CCCGTTGCTGCTGCTGCTACTACTGCTGCTGCTCTGGCCAGTGCCAGGCGCCGGGGTGCTTCAAGGAC
ATATCCCTGGGCAGCCAGTCACCCCGCACTGGGTCCTGGATGGACAACCCTGGCGCACCGTCAGCCTG
GAGGAGCCGGTCTCGAAGCCAGACATGGGGCTGGTGGCCCTGGAGGCTGAAGGCCAGGAGCTCCTGCT
TGAGCTGGAGAAGAACCACAGGCTGCTGGCCCCAGGATACATAGAAACCCACTACGGCCCAGATGGGC
AGCCAGTGGTGCTGGCCCCCAACCACACGGATCATTGCCACTACCAAGGGCGAGTAAGGGGCTTCCCC
GACTCCTGGGTAGTCCTCTGCACCTGCTCTGGGATGAGTGGCCTGATCACCCTCAGCAGGAATGCCAG
CTATTATCTGCGTCCCTGGCCACCCCGGGGCTCCAAGGACTTCTCAACCCACGAGATCTTTCGGATGG
AGCAGCTGCTCACCTGGAAAGGAACCTGTGGCCACAGGGATCCTGGGAACAAAGCGGGCATGACCAGC
CTTCCTGGTGGTCCCCAGAGCAGGGGCAGGCGAGAAGCGCGCAGGACCCGGAAGTACCTGGAACTGTA
CATTGTGGCAGACCACACCCTGTTCTTGACTCGGCACCGAAACTTGAACCACACCAAACAGCGTCTCC
TGGAAGTCGCCAACTACGTGGACCAGCTTCTCAGGACTCTGGACATTCAGGTGGCGCTGACCGGCCTG
GAGGTGTGGACCGAGCGGGACCGCAGCCGCGTCACGCAGGACGCCAACGCCACGCTCTGGGCCTTCCT
GCAGTGGCGCCGGGGGCTGTGGGCGCAGCGGCCCCACGACTCCGCGCAGCTGCTCACGGGCCGCGCCT
TCCAGGGCGCCACAGTGGGCCTGGCGCCCGTCGAGGGCATGTGCCGCGCCGAGAGCTCGGGAGGCGTG
AGCACGGACCACTCGGAGCTCCCCATCGGCGCCGCAGCCACCATGGCCCATGAGATCGGCCACAGCCT
CGGCCTCAGCCACGACCCCGACGGCTGCTGCGTGGAGGCTGCGGCCGAGTCCGGAGGCTGCGTCATGG
CTGCGGCCACCGGGCACCCGTTTCCGCGCGTGTTCAGCGCCTGCAGCCGCCGCCAGCTGCGCGCCTTC
TTCCGCAAGGGGGGCGGCGCTTGCCTCTCCAATGCCCCGGACCCCGGACTCCCGGTGCCGCCGGCGCT
CTGCGGGAACGGCTTCGTGGAAGCGGGCGAGGAGTGTGACTGCGGCCCTGGCCAGGAGTGCCGCGACC
TCTGCTGCTTTGCTCACAACTGCTCGCTGCGCCCGGGGGCCCAGTGCGCCCACGGGGACTGCTGCGTG
CGCTGCCTGACGTTTACCTACTGGACGGCTCACCCTGTGCCAGGGGCAGTGGCTACTGCTGGGATGGC
GCATGTCCCACGCTGGAGCAGCAGTGCCAGCAGCTCTGGGGGCCTGGCTCCCACCCAGCTCCCGAGGC
CTGTTTCCAGGTGG**TGA**ACTCTGCGGGAGATGCTCATGGAAACTGCGGCCAGGACAGCGAGGGCCACT
TCCTGCCCTGTGCAGGGAGGGATGCCCTGTGTGGGAAGCTGCAGTGCCAGGGTGGAAAGCCCAGCCTG
CTCGCACCGCACATGGTGCCAGTGGACTCTACCGTTCACCTAGATGGCCAGGAAGTGACTTGTCGGGG
AGCCTTGGCACTCCCCAGTGCCCAGCTGGACCTGCTTGGCCTGGGCCTGGTAGAGCCAGGCACCCAGT
GTGGACCTAGAATGGTGTGCCAGAGCAGGCGCTGCAGGAAGAATGCCTTCCAGGAGCTTCAGCGCTGC
CTGACTGCCTGCCACAGCCACGGGGTTTGCAATAGCAACCATAACTGCCACTGTGCTCCAGGCTGGGC
TCCACCCTTCTGTGACAAGCCAGGCTTTGGTGGCAGCATGGACAGTGGCCCTGTGCAGGCTGAAAACC
ATGACACCTTCCTGCTGGCCATGCTCCTCAGCGTCCTGCTGCCTCTGCTCCCAGGGGCCGGCCTGGCC
TGGTGTTGCTACCGACTCCCAGGAGCCCATCTGCAGCGATGCAGCTGGGGCTGCAGAAGGGACCCTGC
GTGCAGTGGCCCCAAAGATGGCCCACACAGGGACCACCCCCTGGGCGGCGTTCACCCCATGGAGTTGG
GCCCCACAGCCACTGGACAGCCCTGGCCCCTGGACCCTGAGAACTCTCATGAGCCCAGCAGCCACCCT
GAGAAGCCTCTGCCAGCAGTCTCGCCTGACCCCCAAGATCAAGTCCAGATGCCAAGATCCTGCCTCTG
GTGAGAGGTAGCTCCTAAAATGAACAGATTTAAAGACAGGTGGCCACTGACAGCCACTCCAGGAACTT
GAACTGCAGGGGCAGAGCCAGTGAATCACCGGACCTCCAGCACCTGCAGGCAGCTTGGAAGTTTCTTC
CCCGAGTGGAGCTTCGACCCACCCACTCCAGGAACCCAGAGCCACATTAGAAGTTCCTGAGGGCTGGA
GAACACTGCTGGGCACACTCTCCAGCTCAATAAACCATCAGTCCCAGAAGCAAAGGTCACACAGCCCC
TGACCTCCCTCACCAGTGGAGGCTGGGTAGTGCTGGCCATCCCAAAAGGGCTCTGTCCTGGGAGTCTG
GTGTGTCTCCTACATGCAATTTCCACGGACCCAGCTCTGTGGAGGGCATGACTGCTGGCCAGAAGCTA

GTGGTCCTGGGGCCCTATGGTTCGACTGAGTCCACACTCCCCTGGAGCCTGGCTGGCCTCTGCAAACA
AACATAATTTTGGGGACCTTCCTTCCTGTTTCTTCCCACCCTGTCTTCTCCCCTAGGTGGTTCCTGAG
CCCCCACCCCCAATCCCAGTGCTACACCTGAGGTTCTGGAGCTCAGAATCTGACAGCCTCTCCCCCAT
TCTGTGTGTGTCGGGGGGACAGAGGGAACCATTTAAGAAAAGATACCAAAGTAGAAGTCAAAAGAAAG
ACATGTTGGCTATAGGCGTGGTGGCTCATGCCTATAATCCCAGCACTTTGGGAAGCCGGGGTAGGAGG
ATCACCAGAGGCCAGCAGGTCCACACCAGCCTGGGCAACACAGCAAGACACCGCATCTACAGAAAAAT
TTTAAAATTAGCTGGGCGTGGTGGTGTGTACCTGTAGGCCTAGCTGCTCAGGAGGCTGAAGCAGGAGG
ATCACTTGAGCCTGAGTTCAACACTGCAGTGAGCTATGGTGGCACCACTGCACTCCAGCCTGGGTGAC
AGAGCAAGACCCTGTCTCT

NOV14c, CG50367-06 Protein Sequence	SEQ ID NO: 154	514 aa	MW at 55996.3kD

MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPVSKPDMGLV
ALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRVRGFPDSWVVLCTCSGM
SGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCGHRDPGNKAGMTSLPGGPQSRGRRE
ARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEVANYVDQLLRTLDIQVALTGLEVWTERDRSRVT
QDANATLWAFLQWRRGLWAQRPHDSAQLLTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAA
ATMAHEIGHSLGLSHDPDGCCVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNA
PDPGLPVPPALCGNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLTFTYWTAHP
VPGAVATAGMAHVPRWSSSASSSGGLAPTQLPRPVSRW

NOV14d, CG50367-07 DNA Sequence	SEQ ID NO: 155	3139 bp
	ORF Start: ATG at 37	ORF Stop: TGA at 1354

GCGAGCCGCTGCCTAGAGGCCGAGGAGCTCACAGCT**ATG**GGCTGGAGGCCCCGGAGAGCTCGGGGGAC
CCCGTTGCTGCTGCTGCTACTACTGCTGCTGCTCTGGCCAGTGCCAGGCGCCGGGGTGCTTCAAGGAC
ATATCCCTGGGCAGCCAGTCACCCCGCACTGGGTCCTGGATGGACAACCCTGGCGCACCGTCAGCCTG
GAGGAGCCGGTCTCGAAGCCAGACATGGGGCTGGTGGCCCTGGAGGCTGAAGGCCAGGAGCTCCTGCT
TGAGCTGGAGAAGAACCACAGGCTGCTGGCCCCAGGATACATAGAAACCCACTACGGCCCAGATGGGC
AGCCAGTGGTGCTGGCCCCCAACCACACGGATCATTGCCACTACCAAGGGCGAGTAAGGGGCTTCCCC
GACTCCTGGGTAGTCCTCTGCACCTGCTCTGGGATGAGTGGCCTGATCACCCTCAGCAGGAATGCCAG
CTATTATCTGCGTCCCTGGCCACCCCGGGGCTCCAAGGACTTCTCAACCCACGAGATCTTTCGGATGG
AGCAGCTGCTCACCTGGAAAGGAACCTGTGGCCACAGGGATCCTGGGAACAAAGCGGGCATGACCAGC
CTTCCTGGTGGTCCCCAGAGCAGGGGCAGGCGAGAAGCGCGCAGGACCCGGAAGTACCTGGAACTGTA
CATTGTGGCAGACCACACCCTGTTCTTGACTCGGCACCGAAACTTGAACCACACCAAACAGCGTCTCC
TGGAAGTCGCCAACTACGTGGACCAGCTTCTCAGGACTCTGGACATTCAGGTGGCGCTGACCGGCCTG
GAGGTGTGGACCGAGCGGGACCGCAGCCGCGTCACGCAGGACGCCAACGCCACGCTCTGGGCCTTCCT
GCAGTGGCGCCGGGGACTGTGGGCGCAGCGGCCCCACGACTCCGCGCAGCTGCTCACGACCACTCGGA
GCTCCCCATCGGCGCCGCAGCCACCATGGCCCATGAGATCGGCCACAGCCTCGGCCTCAGCCACGACC
CCGACGGCTGCTGCGTGGAGGCTGCGGCCGAGTCCGGAGGCTGCGTCATGGCTGCGGCCACCGGGCAC
CCGTTTCCGCGCGTGTTCAGCGCCTGCAGCCGCCGCCAGCTGCGCGCCTTCTTCCGCAAGGGGGGCGG
CGCTTGCCTCTCCAATGCCCCGGACCCCGGACTCCCGGTGCCGCCGGCGCTCTGCGGGAACGGCTTCG
TGGAAGCGGGCGAGGAGTGTGACTGCGGCCCTGGCCAGGAGTGCCGCGACCTCTGCTGCTTTGCTCAC
AACTGCTCGCTGCGCCCGGGGGCCCAGTGCGCCCACGGGGACTGCTGCGTGCGCTGCCTGC**TGA**AGCC
GGCTGGAGCGCTGTGCCGCCAGGCCATGGGTGACTGTGACCTCCCTGAGTTTTGCACGGGCACCTCCT
CCCACTGTCCCCCAGACGTTTACCTACTGGACGGCTCACCCTGTGCCAGGGGCAGTGGCTACTGCTGG
GATGGCGCATGTCCCACGCTGGAGCAGCAGTGCCAGCAGCTCTGGGGGCCTGGCTCCCACCCAGCTCC
CGAGGCCTGTTTCCAGGTGGTGAACTCTGCGGGAGATGCTCATGGAAACTGCGGCCAGGACAGCGAGG
GCCACTTCCTGCCCTGTGCAGGGAGGGATGCCCTGTGTGGGAAGCTGCAGTGCCAGGGTGGAAAGCCC
AGCCTGCTCGCACCGCACATGGTGCCAGTGGACTCTACCGTTCACCTAGATGGCCAGGAAGTGACTTG
TCGGGGAGCCTTGGCACTCCCCAGTGCCCAGCTGGACCTGCTTGGCCTGGGCCTGGTAGAGCCAGGCA
CCCAGTGTGGACCTAGAATGGTGTGCCAGAGCAGGCGCTGCAGGAAGAATGCCTTCCAGGAGCTTCAG
CGCTGCCTGACTGCCTGCCACAGCCACGGGGTTTGCAATAGCAACCATAACTGCCACTGTGCTCCAGG
CTGGGCTCCACCCTTCTGTGACAAGCCAGGCTTTGGTGGCAGCATGGACAGTGGCCCTGTGCAGGCTG
AAAACCATGACACCTTCCTGCTGGCCATGCTCCTCAGCGTCCTGCTGCCTCTGCTCCCAGGGGCCGGC
CTGGCCTGGTGTTGCTACCGACTCCCAGGAGCCCATCTGCAGCGATGCAGCTGGGGCTGCAGAAGGGA
CCCTGCGTGCAGTGGCCCCAAAGATGGCCCACACAGGGACCACCCCCTGGGCGGCGTTCACCCCATGG
AGTTGGGCCCCACAGCCACTGGACAGCCCTGGCCCCTGGACCCTGAGAACTCTCATGAGCCCAGCAGC
CACCCTGAGAAGCCTCTGCCAGCAGTCTCGCCTGACCCCCAAGATCAAGTCCAGATGCCAAGATCCTG
CCTCTGGTGAGAGGTAGCTCCTAAAATGAACAGATTTAAAGACAGGTGGCCACTGACAGCCACTCCAG

GAACTTGAACTGCAGGGGCAGAGCCAGTGAATCACCGGACCTCCAGCACCTGCAGGCAGCTTGGAAGT
TTCTTCCCCGAGTGGAGCTTCGACCCACCCACTCCAGGAACCCAGAGCCACATTAGAAGTTCCTGAGG
GCTGGAGAACACTGCTGGGCACACTCTCCAGCTCAATAAACCATCAGTCCCAGAAGCAAAGGTCACAC
AGCCCCTGACCTCCCTCACCAGTGGAGGCTGGGTAGTGCTGGCCATCCCAAAAGGGCTCTGTCCTGGG
AGTCTGGTGTGTCTCCTACATGCAATTTCCACGGACCCAGCTCTGTGGAGGGCATGACTGCTGGCCAG
AAGCTAGTGGTCCTGGGGCCCTATGGTTCGACTGAGTCCACACTCCCCTGGAGCCTGGCTGGCCTCTG
CAAACAAACATAATTTTGGGGACCTTCCTTCCTGTTTCTTCCCACCCTGTCTTCTCCCCTAGGTGGTT
CCTGAGCCCCCACCCCCAATCCCAGTGCTACACCTGAGGTTCTGGAGCTCAGAATCTGACAGCCTCTC
CCCCATTCTGTGTGTGTCGGGGGGACAGAGGGAACCATTTAAGAAAAGATACCAAAGTAGAAGTCAAA
AGAAAGACATGTTGGCTATAGGCGTGGTGGCTCATGCCTATAATCCCAGCACTTTGGGAAGCCGGGGT
AGGAGGATCAC

NOV14d, CG50367-07 Protein Sequence	SEQ ID NO: 156	439 aa	MW at 48163.4kD

MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPVSKPDMGLV
ALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRVRGFPDSWVVLCTCSGM
SGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCGHRDPGNKAGMTSLPGGPQSRGRRE
ARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEVANYVDQLLRTLDIQVALTGLEVWTERDRSRVT
QDANATLWAFLQWRRGLWAQRPHDSAQLLTTTRSSPSAPQPPWPMRSATASASATTPTAAAWRLRPSP
EAASWLRPPGTRFRACSAPAAAASCAPSSARGAALASPMPRTPDSRCRRRSAGTASWKRARSVTAALA
RSAATSAALLTTARCARGPSAPTGTAACAAC

NOV14e, 249356906 DNA Sequence	SEQ ID NO: 157	1278 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTGAGCTGGAGAAGAACCACAGGCTGCTGGCCCCAGGATACATAGAAACCCACTACGGCCCAGA
TGGGCAGCCAGTGGTGCTGGCCCCCAACCACACGGATCATTGCCACTACCAAGGGCGAGTAAGGGGCT
TCCCCGACTCCTGGGTAGTCCTCTGCACCTGCTCTGGGATGAGTGGCCTGATCACCCTCAGCAGGAAT
GCCAGCTATTATCTGCGTCCCTGGCCACCCCGGGGCTCCAAGGACTTCTCAACCCACGAGATCTTTCG
GATGGAGCAGCTGCTCACCTGGAAAGGAACCTGTGGCCACAGGGATCCTGGGAACAAAGCGGGCATGA
CCAGCCTTCCTGGTGGTCCCCAGAGCAGGGGCAGGCGAGAAGCGCGCAGGACCCGGAAGTACCTGGAA
CTGTACATTGTGGCAGACCACACCCTGTTCTTGACTCGGCACCGAAACTTGAACCACACCAAACAGCG
TCTCCTGGAAGTCGACAACTACGTGGACCAGCTTCTCAGGACTCTGGACATTCAGGTGGCGCTGACCG
GCCTGGAGGTGTGGACCGAGCGGGACCGCAGCCGCGTCACGCAGGACGCCAACGCCACGCTCTGGGCC
TTCCTGCAGTGGCGCCGGGGACTGTGGGCGCAGCGGCCCCACGACTCCGCGCAGCTGCTCACGGGCCG
CGCCTTCCAGGGCGCCACAGTGGGCCTGGCGCCCGTCGAGGGCATGTGCCGCGCCGAGAGCTCGGGAG
GCGTGAGCACGGACCACTCGGAGCTCCCCATCGGCGCCGCAGCCACCATGGCCCATGAGATCGGCCAC
AGCCTCGGCCTCAGCCACGACCCCGACGGCTGCTGCGTGGAGGCTGCGGCCGAGTCCGGAGGCTGCGT
CATGGCTGCGGCCACCGGGCACCCGTTTCCGCGCGTGTTCAGCGCCTGCAGCCGCCGCCAGCTGCGCG
CCTTCTTCCGCAAGGGGGGCGGCGCTTGCCTCTCCAATGCCCCGGACCCCGGACTCCCGGTGCCGCCG
GCGCTCTGCGGGAACGGCTTCGTGGAAGCGGGCGAGGAGTGTGACTGCGGCCCTGGCCAGGAGTGCCG
CGACCTCTGCTGCTTTGCTCACAACTGCTCGCTGCGCCCGGGGGCCCAGTGCGCCCACGGGGACTGCT
GCGTGCGCTGCCTGCTGAAGCCGGCTGGAGCGCTGTGCCGCCAGGCCATGGGTGACTGTGACCTCCCT
GAGTTTTGCACGGGCACCTCCTCCCACTGTCCCCCAGACGTTTACCTACTCGAG

NOV14e, 249356906 Protein Sequence	SEQ ID NO: 158	426 aa	MW at 46516.2kD

KLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRVRGFPDSWVVLCTCSGMSGLITLSRN
ASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCGHRDPGNKAGMTSLPGGPQSRGRREARRTRKYLE
LYIVADHTLFLTRHRNLNHTKQRLLEVDNYVDQLLRTLDIQVALTGLEVWTERDRSRVTQDANATLWA
FLQWRRGLWAQRPHDSAQLLTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAAATMAHEIGH
SLGLSHDPDGCCVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNAPDPGLPVPP
ALCGNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKPAGALCRQAMGDCDLP
EFCTGTSSHCPPDVYLLE

NOV14f, CG50367-02 DNA Sequence	SEQ ID NO: 159	2705 bp
	ORF Start: ATG at 3	ORF Stop: TGA at 2688

CTATGGGCTGGAGGCCCCGGAGAGCTCGGGGGACCCCGTTGCTGCTGCTGCTACTACTGCTGCTGCTC
TGGCCAGTGCCAGGCGCCGGGGTGCTTCAAGGACATATCCCTGGGCAGCCAGTCACCCCGCACTGGGT

CCTGGATGGACAACCCTGGCGCACCGTCAGCCTGGAGGAGCCGGTCTCGAAGCCAGACATGGGGCTGG
TGGCCCTGGAGGCTGAAGGCCAGGAGCTCCTGCTTGAGCTGGAGAAGAACCACAGGCTGCTGGCCCCA
GGATACATAGAAACCCACTACGGCCCAGATGGGCAGCCAGTGGTGCTGGCCCCCAACCACACGGATCA
TTGCCACTACCAAGGGCGAGTAAGGGGCTTCCCCGACTCCTGGGTAGTCCTCTGCACCTGCTCTGGGA
TGAGTGGCCTGATCACCCTCAGCAGGAATGCCAGCTATTATCTGCGTCCCTGGCCACCCCGGGGCTCC
AAGGACTTCTCAACCCACGAGATCTTTCGGATGGAGCAGCTGCTCACCTGGAAAGGAACCTGTGGCCA
CAGGGATCCTGGGAACAAAGCGGGCATGACCAGCCTTCCTGGTGGTCCCCAGAGCAGGGGCAGGCGAG
AAGCGCGCAGGACCCGGAAGTACCTGGAACTGTACATTGTGGCAGACCACACCCTGTTCTTGACTCGG
CACCGAAACTTGAACCACACCAAACAGCGTCTCCTGGAAGTCGCCAACTACGTGGACCAGCTTCTCAG
GACTCTGGACATTCAGGTGGCGCTGACCGGCCTGGAGGTGTGGACCGAGCGGGACCGCAGCCGCGTCA
CGCAGGACGCCAACGCCACGCTCTGGGCCTTCCTGCAGTGGCGCCGGGGGCTGTGGGCGCAGCGGCCC
CACGACTCCGCGCAGCTGCTCACGGGCCGCGCCTTCCAGGGCGCCACAGTGGGCCTGGCGCCCGTCGA
GGGCATGTGCCGCGCCGAGAGCTCGGGAGGCGTGAGCACGGACCACTCGGAGCTCCCCATCGGCGCCG
CAGCCACCATGGCCCATGAGATCGGCCACAGCCTCGGCCTCAGCCACGACCCCGACGGCTGCTGCGTG
GAGGCTGCGGCCGAGTCCGGAGGCTGCGTCATGGCTGCGGCCACCGGGCACCCGTTTCCGCGCGTGTT
CAGCGCCTGCAGCCGCCGCCAGCTGCGCGCCTTCTTCCGCAAGGGGGGCGGCGCTTGCCTCTCCAATG
CCCCGGACCCCGGACTCCCGGTGCCGCCGGCGCTCTGCGGGAACGGCTTCGTGGAAGCGGGCGAGGAG
TGTGACTGCGGCCCTGGCCAGGAGTGCCGCGACCTCTGCTGCTTTGCTCACAACTGCTCGCTGCGCCC
GGGGGCCCAGTGCGCCCACGGGGACTGCTGCGTGCGCTGCCTGCTGAAGCCGGCTGGAGCGCTGTGCC
GCCAGGCCATGGGTGACTGTGACCTCCCTGAGTTTTGCACGGGCACCTCCTCCCACTGTCCCCCAGAC
GTTTACCTACTGGACGGCTCACCCTGTGCCAAGGGCAGTGGCTACTGCTGGGATGGCGCATGTCCCAC
GCTGGAGCAGCAGTGCCAGCAGCTCTGGGGGCCTGGCTCCCACCCAGCTCCCGAGGCCTGTTTCCAGG
TGGTGAACTCTGCGGGAGATGCTCATGGAAACTGCGGCCAGGACAGCGAGGGCCACTTCCTGCCCTGT
GCAGGGAGGGATGCCCTGTGTGGGAAGCTGCAGTGCCAGGGTGGAAAGCCCAGCCTGCTCGCACCGCA
CATGGTGCCAGTGGACTCTACCGTTCACCTAGATGGCCAGGAAGTGACTTGTCGGGGAGCCTTGGCAC
TCCCCAGTGCCCAGCTGGACCTGCTTGGCCTGGGCCTGGTAGAGCCAGGCACCCAGTGTGGACCTAGA
ATGGTGTGCCAGAGCAGGCGCTGCAGGAAGAATGCCTTCCAGGAGCTTCAGCGCTGCCTGACTGCCTG
CCACAGCCACGGGGTTTGCAATAGCAACCATAACTGCCACTGTGCTCCAGGCTGGGCTCCACCCTTCT
GTGACAAGCCAGGCTTTGGTGGCAGCATGGACAGTGGCCCTGTGCAGGCTGAAAACCATGACACCTTC
CTGCTGGCCATGCTCCTCAGCGTCCTGCTGCCTCTGCTCCCAGGCGCCGGCCTGGCCTGGTGTTGCTA
CCGACTCCCAGGAGCCCATCTGCAGCGATGCAGCTGGGGCTGCAGAAGGGACCCTGCGTGCAGTGGCC
CCAAAGATGGCCCACACAGAGACCACCCCCTGGGCGGCGTTCACCCCATGGAGTTGGGCCCCACAGCC
ACTGGACAGCCCTGGCCCCTGGACCCTGAGAACTCTCATGAGCCCAGCAGCCACCCTGAGAAGCCTCT
GCCAGCAGTCTCGCCTGACCCCCAAGGTGGTTCCCTTGCAGCCTGGGGCCCCAGTCCTTTAGGGGACA
ACATATCCTCCTCATTCTCAGCAGATCAAGTCCAGATGCCAAGATCCTGCCTCTGTGGCGAACCCTGG
GGAGGCCACGTGGGAAGGAAAGAGGGCTCTAAGAGGGGAGGCCCCAGACTGGGGGAGAGGCCTGTCTG
GAGCCCAGGATCACCTGGCTGTGCTGCAGAACTGGAGAAGAGAAGCTCAGCAGAAAGGAGCTGGCATG
GGGCCAACAGCAGAAAAGCAGGAGGCACGCAGAAGTGACTGGGAAGCAGGAGG

NOV14f, CG50367-02 Protein Sequence	SEQ ID NO: 160	895 aa	MW at 96021.0kD

MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPVSKPDMGLV
ALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRVRGFPDSWVVLCTCSGM
SGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCGHRDPGNKAGMTSLPGGPQSRGRRE
ARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEVANYVDQLLRTLDIQVALTGLEVWTERDRSRVT
QDANATLWAFLQWRRGLWAQRPHDSAQLLTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAA
ATMAHEIGHSLGLSHDPDGCCVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNA
PDPGLPVPPALCGNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKPAGALCR
QAMGDCDLPEFCTGTSSHCPPDVYLLDGSPCAKGSGYCWDGACPTLEQQCQQLWGPGSHPAPEACFQV
VNSAGDAHGNCGQDSEGHFLPCAGRDALCGKLQCQGGKPSLLAPHMVPVDSTVHLDGQEVTCRGALAL
PSAQLDLLGLGLVEPGTQCGPRMVCQSRRCRKNAFQELQRCLTACHSHGVCNSNHNCHCAPGWAPPFC
DKPGFGGSMDSGPVQAENHDTFLLAMLLSVLLPLLPGAGLAWCCYRLPGAHLQRCSWGCRRDPACSGP
KDGPHRDHPLGGVHPMELGPTATGQPWPLDPENSHEPSSHPEKPLPAVSPDPQGGSLAAWGPSPLGDN
ISSSFSADQVQMPRSCLCGEPWGGHVGRKEGSKRGGPRLGERPVWSPGSPGCAAELEKRSSAERSWHG
ANSRKAGGTQK

NOV14g, CG50367-03 DNA Sequence	SEQ ID NO: 161	2642 bp
	ORF Start: ATG at 3	ORF Stop: TGA at 2625

CTATGGGCTGGAGGCCCCGGAGAGCTCGGGGGACCCCGTTGCTGCTGCTGCTACTACTGCTGCTGCTC

TGGCCAGTGCCAGGCGCCGGGGTGCTTCAAGGACATATCCCTGGGCAGCCAGTCACCCCGCACTGGGT
CCTGGATGGACAACCCTGGCGCACCGTCAGCCTGGAGGAGCCGGTCTCGAAGCCAGACATGGGGCTGG
TGGCCCTGGAGGCTGAAGGCCAGGAGCTCCTGCTTGAGCTGGAGAAGAACCACAGGCTGCTGGCCCCA
GGATACATAGAAACCCACTACGGCCCAGATGGGCAGCCAGTGGTGCTGGCCCCCAACCACACGGATCA
TTGCCACTACCAAGGGCGAGTAAGGGGCTTCCCCGACTCCTGGGTAGTCCTCTGCACCTGCTCTGGGA
TGAGTGGCCTGATCACCCTCAGCAGGAATGCCAGCTATTATCTGCGTCCCTGGCCACCCCGGGGCTCC
AAGGACTTCTCAACCCACGAGATCTTTCGGATGGAGCAGCTGCTCACCTGGAAAGGAACCTGTGGCCA
CAGGGATCCTGGGAACAAAGCGGGCATGACCAGCCTTCCTGGTGGTCCCCAGAGCAGGGGCAGGCGAG
AAGCGCGCAGGACCCGGAAGTACCTGGAACTGTACATTGTGGCAGACCACACCCTGTTCTTGACTCGG
CACCGAAACTTGAACCACACCAAACAGCGTCTCCTGGAAGTCGCCAACTACGTGGACCAGCTTCTCAG
GACTCTGGACATTCAGGTGGCGCTGACCGGCCTGGAGGTGTGGACCGAGCGGGACCGCAGCCGCGTCA
CGCAGGACGCCAACGCCACGCTCTGGGCCTTCCTGCAGTGGCGCCGGGGGCTGTGGGCGCAGCGGCCC
CACGACTCCGCGCAGCTGCTCACGGGCCGCGCCTTCCAGGGCGCCACAGTGGGCCTGGCGCCCGTCGA
GGGCATGTGCCGCGCCGAGAGCTCGGGAGGCGTGAGCACGGACCACTCGGAGCTCCCCATCGGCGCCG
CAGCCACCATGGCCCATGAGATCGGCCACAGCCTCGGCCTCAGCCACGACCCCGACGGCTGCTGCGTG
GAGGCTGCGGCCGAGTCCGGAGGCTGCGTCATGGCTGCGGCCACCGGGCACCCGTTTCCGCGCGTGTT
CAGCGCCTGCAGCCGCCGCCAGCTGCGCGCCTTCTTCCGCAAGGGGGGCGGCGCTTGCCTCTCCAATG
CCCCGGACCCCGGACTCCCGGTGCCGCCGGCGCTCTGCGGGAACGGCTTCGTGGAAGCGGGCGAGGAG
TGTGACTGCGGCCCTGGCCAGGAGTGCCGCGACCTCTGCTGCTTTGCTCACAACTGCTCGCTGCGCCC
GGGGGCCCAGTGCGCCCACGGGGACTGCTGCGTGCGCTGCCTGCTGAAGCCGGCTGGAGCGCTGTGCC
GCCAGGCCATGGGTGACTGTGACCTCCCTGAGTTTTGCACGGGCACCTCCTCCCACTGTCCCCCAGAC
GTTTACCTACTGGACGGCTCACCCTGTGCCAAGGGCAGTGGCTACTGCTGGGATGGCGCATGTCCCAC
GCTGGAGCAGCAGTGCCAGCAGCTCTGGGGGCCTGGCTCCCACCCAGCTCCCGAGGCCTGTTTCCAGG
TGGTGAACTCTGCGGGAGATGCTCATGGAAACTGCGGCCAGGACAGCGAGGGCCACTTCCTGCCCTGT
GCAGGGAGGGATGCCCTGTGTGGGAAGCTGCAGTGCCAGGGTGGAAAGCCCAGCCTGCTCGCACCGCA
CATGGTGCCAGTGGACTCTACCGTTCACCTAGATGGCCAGGAAGTGACTTGTCGGGGAGCCTTGGCAC
TCCCCAGTGCCCAGCTGGACCTGCTTGGCCTGGGCCTGGTAGAGCCAGGCACCCAGTGTGGACCTAGA
ATGGTGTGCCAGAGCAGGCGCTGCAGGAAGAATGCCTTCCAGGAGCTTCAGCGCTGCCTGACTGCCTG
CCACAGCCACGGGGTTTGCAATAGCAACCATAACTGCCACTGTGCTCCAGGCTGGGCTCCACCCTTCT
GTGACAAGCCAGGCTTTGGTGGCAGCATGGACAGTGGCCCTGTGCAGGCTGAAAACCATGACACCTTC
CTGCTGGCCATGCTCCTCAGCGTCCTGCTGCCTCTGCTCCCAGGCGCCGGCCTGGCCTGGTGTTGCTA
CCGACTCCCAGGAGCCCATCTGCAGCGATGCAGCTGGGGCTGCAGAAGGGACCCTGCGTGCAGTGGCC
CCAAAGATGGCCCACACAGAGACCACCCCCTGGGCGGCGTTCACCCCATGGAGTTGGGCCCCACAGCC
ACTGGACAGCCCTGGCCCCTGGACCCTGAGAACTCTCATGAGCCCAGCAGCCACCCTGAGAAGCCTCT
GCCAGCAGTCTCGCCTGACCCCCAAGCAGATCAAGTCCAGATGCCAAGATCCTGCCTCTGTGGCGAAC
CCTGGGGAGGCCACGTGGGAAGGAAAGAGGGCTCTAAGAGGGGAGGCCCCAGACTGGGGGAGAGGCCT
GTCTGGAGCCCAGGATCACCTGGCTGTGCTGCAGAACTGGAGAAGAGAAGCTCAGCAGAAAGGAGCTG
GCATGGGGCCAACAGCAGAAAAGCAGGAGGCACGCAGAAGTGACTGGGAAGCAGGAGG

NOV14g, CG50367-03 Protein Sequence	SEQ ID NO: 162	874 aa	MW at 94031.9kD

MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPVSKPDMGLV
ALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRVRGFPDSWVVLCTCSGM
SGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCGHRDPGNKAGMTSLPGGPQSRGRRE
ARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEVANYVDQLLRTLDIQVALTGLEVWTERDRSRVT
QDANATLWAFLQWRRGLWAQRPHDSAQLLTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAA
ATMAHEIGHSLGLSHDPDGCCVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNA
PDPGLPVPPALCGNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKPAGALCR
QAMGDCDLPEFCTGTSSHCPPDVYLLDGSPCAKGSGYCWDGACPTLEQQCQQLWGPGSHPAPEACFQV
VNSAGDAHGNCGQDSEGHFLPCAGRDALCGKLQCQGGKPSLLAPHMVPVDSTVHLDGQEVTCRGALAL
PSAQLDLLGLGLVEPGTQCGPRMVCQSRRCRKNAFQELQRCLTACHSHGVCNSNHNCHCAPGWAPPFC
DKPGFGGSMDSGPVQAENHDTFLLAMLLSVLLPLLPGAGLAWCCYRLPGAHLQRCSWGCRRDPACSGP
KDGPHRDHPLGGVHPMELGPTATGQPWPLDPENSHEPSSHPEKPLPAVSPDPQADQVQMPRSCLCGEP
WGGHVGRKEGSKRGGPRLGERPVWSPGSPGCAAELEKRSSAERSWHGANSRKAGGTQK

NOV14h, CG50367-04 DNA Sequence	SEQ ID NO: 163	3468 bp
	ORF Start: ATG at 37	ORF Stop: TGA at 2473

GCGAGCCGCTGCCTAGAGGCCGAGGAGCTCACAGCTATGGGCTGGAGGCCCCGGAGAGCTCGGGGGAC
CCCGTTGCTGCTGCTGCTACTACTGCTGCTGCTCTGGCCAGTGCCAGGCGCCGGGGTGCTTCAAGGAC

ATATCCCTGGGCAGCCAGTCACCCCGCACTGGGTCCTGGATGGACAACCCTGGCGCACCGTCAGCCTG
GAGGAGCCGGTCTCGAAGCCAGACATGGGGCTGGTGGCCCTGGAGGCTGAAGGCCAGGAGCTCCTGCT
TGAGCTGGAGAAGAACCACAGGCTGCTGGCCCCAGGATACATAGAAACCCACTACGGCCCAGATGGGC
AGCCAGTGGTGCTGGCCCCCAACCACACGGATCATTGCCACTACCAAGGGCGAGTAAGGGGCTTCCCC
GACTCCTGGGTAGTCCTCTGCACCTGCTCTGGGATGAGTGGCCTGATCACCCTCAGCAGGAATGCCAG
CTATTATCTGCGTCCCTGGCCACCCCGGGGCTCCAAGGACTTCTCAACCCACGAGATCTTTCGGATGG
AGCAGCTGCTCACCTGGAAAGGAACCTGTGGCCACAGGGATCCTGGGAACAAAGCGGGCATGACCAGC
CTTCCTGGTGGTCCCCAGAGCAGGGGCAGGCGAGAAGCGCGCAGGACCCGGAAGTACCTGGAACTGTA
CATTGTGGCAGACCACACCCTGTTCTTGACTCGGCACCGAAACTTGAACCACACCAAACAGCGTCTCC
TGGAAGTCGCCAACTACGTGGACCAGCTTCTCAGGACTCTGGACATTCAGGTGGCGCTGACCGGCCTG
GAGGTGTGGACCGAGCGGGACCGCAGCCGCGTCACGCAGGACGCCAACGCCACGCTCTGGGCCTTCCT
GCAGTGGCGCCGGGGGCTGTGGGCGCAGCGGCCCCACGACTCCGCGCAGCTGCTCACGGGCCGCGCCT
TCCAGGGCGCCACAGTGGGCCTGGCGCCCGTCGAGGGCATGTGCCGCGCCGAGAGCTCGGGAGGCGTG
AGCACGGACCACTCGGAGCTCCCCATCGGCGCCGCAGCCACCATGGCCCATGAGATCGGCCACAGCCT
CGGCCTCAGCCACGACCCCGACGGCTGCTGCGTGGAGGCTGCGGCCGAGTCCGGAGGCTGCGTCATGG
CTGCGGCCACCGGGCACCCGTTTCCGCGCGTGTTCAGCGCCTGCAGCCGCCGCCAGCTGCGCGCCTTC
TTCCGCAAGGGGGGCGGCGCTTGCCTCTCCAATGCCCCGGACCCCGGACTCCCGGTGCCGCCGGCGCT
CTGCGGGAACGGCTTCGTGGAAGCGGGCGAGGAGTGTGACTGCGGCCCTGGCCAGGAGTGCCGCGACC
TCTGCTGCTTTGCTCACAACTGCTCGCTGCGCCCGGGGGCCCAGTGCGCCCACGGGGACTGCTGCGTG
CGCTGCCTGCTGAAGCCGGCTGGAGCGCTGTGCCGCCAGGCCATGGGTGACTGTGACCTCCCTGAGTT
TTGCACGGGCACCTCCTCCCACTGTCCCCCAGACGTTTACCTACTGGACGGCTCACCCTGTGCCAGGG
GCAGTGGCTACTGCTGGGATGGCGCATGTCCCACGCTGGAGCAGCAGTGCCAGCAGCTCTGGGGGCCT
GGCTCCCACCCAGCTCCCGAGGCCTGTTTCCAGGTGGTGAACTCTGCGGGAGATGCTCATGGAAACTG
CGGCCAGGACAGCGAGGGCCACTTCCTGCCCTGTGCAGGGAGGGATGCCCTGTGTGGGAAGCTGCAGT
GCCAGGGTGGAAAGCCCAGCCTGCTCGCACCGCACATGGTGCCAGTGGACTCTACCGTTCACCTAGAT
GGCCAGGAAGTGACTTGTCGGGGAGCCTTGGCACTCCCCAGTGCCCAGCTGGACCTGCTTGGCCTGGG
CCTGGTAGAGCCAGGCACCCAGTGTGGACCTAGAATGGTGTGCCAGAGCAGGCGCTGCAGGAAGAATG
CCTTCCAGGAGCTTCAGCGCTGCCTGACTGCCTGCCACAGCCACGGGGTTTGCAATAGCAACCATAAC
TGCCACTGTGCTCCAGGCTGGGCTCCACCCTTCTGTGACAAGCCAGGCTTTGGTGGCAGCATGGACAG
TGGCCCTGTGCAGGCTGAAAACCATGACACCTTCCTGCTGGCCATGCTCCTCAGCGTCCTGCTGCCTC
TGCTCCCAGGGGCCGGCCTGGCCTGGTGTTGCTACCGACTCCCAGGAGCCCATCTGCAGCGATGCAGC
TGGGGCTGCAGAAGGGACCCTGCGTGCAGTGGCCCCAAAGATGGCCCACACAGGGACCACCCCCTGGG
CGGCGTTCACCCCATGGAGTTGGGCCCCACAGCCACTGGACAGCCCTGGCCCCTGGACCCTGAGAACT
CTCATGAGCCCAGCAGCCACCCTGAGAAGCCTCTGCCAGCAGTCTCGCCTGACCCCCAAGATCAAGTC
CAGATGCCAAGATCCTGCCTCTGGTGAGAGGTAGCTCCTAAAATGAACAGATTTAAAGACAGGTGGCC
ACTGACAGCCACTCCAGGAACTTGAACTGCAGGGGCAGAGCCAGTGAATCACCGGACCTCCAGCACCT
GCAGGCAGCTTGGAAGTTTCTTCCCCGAGTGGAGCTTCGACCCACCCACTCCAGGAACCCAGAGCCAC
ATTAGAAGTTCCTGAGGGCTGGAGAACACTGCTGGGCACACTCTCCAGCTCAATAAACCATCAGTCCC
AGAAGCAAAGGTCACACAGCCCCTGACCTCCCTCACCAGTGGAGGCTGGGTAGTGCTGGCCATCCCAA
AAGGGCTCTGTCCTGGGAGTCTGGTGTGTCTCCTACATGCAATTTCCACGGACCCAGCTCTGTGGAGG
GCATGACTGCTGGCCAGAAGCTAGTGGTCCTGGGGCCCTATGGTTCGACTGAGTCCACACTCCCCTGG
AGCCTGGCTGGCCTCTGCAAACAAACATAATTTTGGGGACCTTCCTTCCTGTTTCTTCCCACCCTGTC
TTCTCCCCTAGGTGGTTCCTGAGCCCCCACCCCCAATCCCAGTGCTACACCTGAGGTTCTGGAGCTCA
GAATCTGACAGCCTCTCCCCCATTCTGTGTGTGTCGGGGGGACAGAGGGAACCATTTAAGAAAAGATA
CCAAAGTAGAAGTCAAAAGAAAGACATGTTGGCTATAGGCGTGGTGGCTCATGCCTATAATCCCAGCA
CTTTGGGAAGCCGGGGTAGGAGGATCACCAGAGGCCAGCAGGTCCACACCAGCCTGGGCAACACAGCA
AGACACCGCATCTACAGAAAAATTTTAAAATTAGCTGGGCGTGGTGGTGTGTACCTGTAGGCCTAGCT
GCTCAGGAGGCTGAAGCAGGAGGATCACTTGAGCCTGAGTTCAACACTGCAGTGAGCTATGGTGGCAC
CACTGCACTCCAGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAAATAAATTTTAAAAAGACATATTA

NOV14h, CG50367-04 Protein Sequence	SEQ ID NO: 164	812 aa	MW at 87666.9kD

MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPVSKPDMGLV
ALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRVRGFPDSWVVLCTCSGM
SGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCGHRDPGNKAGMTSLPGGPQSRGRRE
ARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEVANYVDQLLRTLDIQVALTGLEVWTERDRSRVT
QDANATLWAFLQWRRGLWAQRPHDSAQLLTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAA
ATMAHEIGHSLGLSHDPDGCCVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNA
PDPGLPVPPALCGNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKPAGALCR

QAMGDCDLPEFCTGTSSHCPPDVYLLDGSPCARGSGYCWDGACPTLEQQCQQLWGPGSHPAPEACFQV
VNSAGDAHGNCGQDSEGHFLPCAGRDALCGKLQCQGGKPSLLAPHMVPVDSTVHLDGQEVTCRGALAL
PSAQLDLLGLGLVEPGTQCGPRMVCQSRRCRKNAFQELQRCLTACHSHGVCNSNHNCHCAPGWAPPFC
DKPGFGGSMDSGPVQAENHDTFLLAMLLSVLLPLLPGAGLAWCCYRLPGAHLQRCSWGCRRDPACSGP
KDGPHRDHPLGGVHPMELGPTATGQPWPLDPENSHEPSSHPEKPLPAVSPDPQDQVQMPRSCLW

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 14B.

Table 14B. Comparison of the NOV14 protein sequences.

```
NOV14a  MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPV
NOV14b  MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPV
NOV14c  MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPV
NOV14d  MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPV
NOV14e  ------------------------------------------------------------
NOV14f  MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPV
NOV14g  MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPV
NOV14h  MGWRPRRARGTPLLLLLLLLLLWPVPGAGVLQGHIPGQPVTPHWVLDGQPWRTVSLEEPV

NOV14a  SKPDMGLVALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRV
NOV14b  SKPDMGLVALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRV
NOV14c  SKPDMGLVALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRV
NOV14d  SKPDMGLVALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRV
NOV14e  ----------------KLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRV
NOV14f  SKPDMGLVALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRV
NOV14g  SKPDMGLVALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRV
NOV14h  SKPDMGLVALEAEGQELLLELEKNHRLLAPGYIETHYGPDGQPVVLAPNHTDHCHYQGRV

NOV14a  RGFPDSWVVLCTCSGMSGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCG
NOV14b  RGFPDSWVVLCTCSGMSGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCG
NOV14c  RGFPDSWVVLCTCSGMSGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCG
NOV14d  RGFPDSWVVLCTCSGMSGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCG
NOV14e  RGFPDSWVVLCTCSGMSGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCG
NOV14f  RGFPDSWVVLCTCSGMSGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCG
NOV14g  RGFPDSWVVLCTCSGMSGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCG
NOV14h  RGFPDSWVVLCTCSGMSGLITLSRNASYYLRPWPPRGSKDFSTHEIFRMEQLLTWKGTCG

NOV14a  HRDPGNKAGMTSLPGGPQSRGRREARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEV
NOV14b  HRDPGNKAGMTSLPGGPQSRGRREARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEV
NOV14c  HRDPGNKAGMTSLPGGPQSRGRREARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEV
NOV14d  HRDPGNKAGMTSLPGGPQSRGRREARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEV
NOV14e  HRDPGNKAGMTSLPGGPQSRGRREARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEV
NOV14f  HRDPGNKAGMTSLPGGPQSRGRREARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEV
NOV14g  HRDPGNKAGMTSLPGGPQSRGRREARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEV
NOV14h  HRDPGNKAGMTSLPGGPQSRGRREARRTRKYLELYIVADHTLFLTRHRNLNHTKQRLLEV

NOV14a  ANYVDQLLRTLDIQVALTGLEVWTERDRSRVTQDANATLWAFLQWRRGLWAQRPHDSAQL
NOV14b  ANYVDQLLRTLDIQVALTGLEVWTERDRSRVTQDANATLWAFLQWRRGLWAQRPHDSAQL
NOV14c  ANYVDQLLRTLDIQVALTGLEVWTERDRSRVTQDANATLWAFLQWRRGLWAQRPHDSAQL
NOV14d  ANYVDQLLRTLDIQVALTGLEVWTERDRSRVTQDANATLWAFLQWRRGLWAQRPHDSAQL
NOV14e  DNYVDQLLRTLDIQVALTGLEVWTERDRSRVTQDANATLWAFLQWRRGLWAQRPHDSAQL
NOV14f  ANYVDQLLRTLDIQVALTGLEVWTERDRSRVTQDANATLWAFLQWRRGLWAQRPHDSAQL
NOV14g  ANYVDQLLRTLDIQVALTGLEVWTERDRSRVTQDANATLWAFLQWRRGLWAQRPHDSAQL
```

```
NOV14h  ANYVDQLLRTLDIQVALTGLEVWTERDRSRVTQDANATLWAFLQWRRGLWAQRPHDSAQL

NOV14a  LTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAAATMAHEIGHSLGLSHDPDGC
NOV14b  LTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAAATMAHEIGHSLGLSHDPDGC
NOV14c  LTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAAATMAHEIGHSLGLSHDPDGC
NOV14d  LTTTRSSPSAP--QPPWPMRSATASASATTPTAAAWRLRPSPEAASWLRPPGTRFRACSA
NOV14e  LTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAAATMAHEIGHSLGLSHDPDGC
NOV14f  LTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAAATMAHEIGHSLGLSHDPDGC
NOV14g  LTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAAATMAHEIGHSLGLSHDPDGC
NOV14h  LTGRAFQGATVGLAPVEGMCRAESSGGVSTDHSELPIGAAATMAHEIGHSLGLSHDPDGC

NOV14a  CVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNAPDPGLPVPPALC
NOV14b  CVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNAPDPGLPVPPALC
NOV14c  CVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNAPDPGLPVPPALC
NOV14d  PAAAASCAPSSARGAALASPMPRTPDSRCRR------RSAGTASWKRARSVTAALARSAA
NOV14e  CVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNAPDPGLPVPPALC
NOV14f  CVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNAPDPGLPVPPALC
NOV14g  CVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNAPDPGLPVPPALC
NOV14h  CVEAAAESGGCVMAAATGHPFPRVFSACSRRQLRAFFRKGGGACLSNAPDPGLPVPPALC

NOV14a  GNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKP-------AGA
NOV14b  GNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKP-------AGA
NOV14c  GNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLTFTYWTAHPVPGA
NOV14d  TSAALLTTARCARGPSAPTGTAACAAC-------------------------------
NOV14e  GNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKP-------AGA
NOV14f  GNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKP-------AGA
NOV14g  GNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKP-------AGA
NOV14h  GNGFVEAGEECDCGPGQECRDLCCFAHNCSLRPGAQCAHGDCCVRCLLKP-------AGA

NOV14a  LCRQAMGDCDLPEFCTGTSSHCPPDVYLLDGSPCARGSGYCWDGACPTLEQQCQQLWGPG
NOV14b  LCRQAMGDCDLPEFCTGTSSHCPPDVYLLDGSPCARGSGYCWDGACPTLEQQCQQLWGPG
NOV14c  VATAGMAHVPRWSSSASSSGGLAPTQLPRPVSRW-------------------------
NOV14d  -------------------------------------------------------------
NOV14e  LCRQAMGDCDLPEFCTGTSSHCPPDVYLLE-----------------------------
NOV14f  LCRQAMGDCDLPEFCTGTSSHCPPDVYLLDGSPCAKGSGYCWDGACPTLEQQCQQLWGPG
NOV14g  LCRQAMGDCDLPEFCTGTSSHCPPDVYLLDGSPCAKGSGYCWDGACPTLEQQCQQLWGPG
NOV14h  LCRQAMGDCDLPEFCTGTSSHCPPDVYLLDGSPCARGSGYCWDGACPTLEQQCQQLWGPG

NOV14a  SHPAPEACFQVVNSAGDAHGNCGQDSEGHFLPCAGRDALCGKLQCQGGKPSLLAPHMVPV
NOV14b  SHPAPEACFQVVNSAGDAHGNCGQDSEGHFLPCAGRDALCGKLQCQGGKPSLLAPHMVPV
NOV14c  -------------------------------------------------------------
NOV14d  -------------------------------------------------------------
NOV14e  -------------------------------------------------------------
NOV14f  SHPAPEACFQVVNSAGDAHGNCGQDSEGHFLPCAGRDALCGKLQCQGGKPSLLAPHMVPV
NOV14g  SHPAPEACFQVVNSAGDAHGNCGQDSEGHFLPCAGRDALCGKLQCQGGKPSLLAPHMVPV
NOV14h  SHPAPEACFQVVNSAGDAHGNCGQDSEGHFLPCAGRDALCGKLQCQGGKPSLLAPHMVPV

NOV14a  DSTVHLDGQEVTCRGALALPSAQLDLLGLGLVEPGTQCGPRMVCQSRRCRKNAFQELQRC
NOV14b  DSTVHLDGQEVTCRGALALPSAQLDLLGLGLVESGTQCGPRMVCN-------------
NOV14c  -------------------------------------------------------------
NOV14d  -------------------------------------------------------------
NOV14e  -------------------------------------------------------------
NOV14f  DSTVHLDGQEVTCRGALALPSAQLDLLGLGLVEPGTQCGPRMVCQSRRCRKNAFQELQRC
NOV14g  DSTVHLDGQEVTCRGALALPSAQLDLLGLGLVEPGTQCGPRMVCQSRRCRKNAFQELQRC
```

```
NOV14h  DSTVHLDGQEVTCRGALALPSAQLDLLGLGLVEPGTQCGPRMVCQSRRCRKNAFQELQRC

NOV14a  LTACHSHGVCNSNHNCHCAPGWAPPFCDKPGFGGSMDSGPVQAENHDTFLLAMLLSVLLP
NOV14b  -----------SNHNCHCAPGWAPPFCDKPGFGGSMDSGPVQAENHDTFLLAMLLSVLLP
NOV14c  ------------------------------------------------------------
NOV14d  ------------------------------------------------------------
NOV14e  ------------------------------------------------------------
NOV14f  LTACHSHGVCNSNHNCHCAPGWAPPFCDKPGFGGSMDSGPVQAENHDTFLLAMLLSVLLP
NOV14g  LTACHSHGVCNSNHNCHCAPGWAPPFCDKPGFGGSMDSGPVQAENHDTFLLAMLLSVLLP
NOV14h  LTACHSHGVCNSNHNCHCAPGWAPPFCDKPGFGGSMDSGPVQAENHDTFLLAMLLSVLLP

NOV14a  LLPGAGLAWCCYRLPGAHLQRCSWGCRRDPACSGPKDGPHRDHPLGGVHPMELGPTATGQ
NOV14b  LLPGAGLAWCCYRLPGAHLQRCSWGCRRDPACSGPKDGPHRDHPLGGVHPMELGPTATGQ
NOV14c  ------------------------------------------------------------
NOV14d  ------------------------------------------------------------
NOV14e  ------------------------------------------------------------
NOV14f  LLPGAGLAWCCYRLPGAHLQRCSWGCRRDPACSGPKDGPHRDHPLGGVHPMELGPTATGQ
NOV14g  LLPGAGLAWCCYRLPGAHLQRCSWGCRRDPACSGPKDGPHRDHPLGGVHPMELGPTATGQ
NOV14h  LLPGAGLAWCCYRLPGAHLQRCSWGCRRDPACSGPKDGPHRDHPLGGVHPMELGPTATGQ

NOV14a  PWPLAPGAPADHIHNIYPPPFLPDPENSHEPSSHPEKPLPAVSPDPQGGSLAAWGPSPLG
NOV14b  PWPLDP------------------ENSHEPSSHPEKPLPAVSPDPQ------------
NOV14c  ------------------------------------------------------------
NOV14d  ------------------------------------------------------------
NOV14e  ------------------------------------------------------------
NOV14f  PWPLDP------------------ENSHEPSSHPEKPLPAVSPDPQGGSLAAWGPSPLG
NOV14g  PWPLDP------------------ENSHEPSSHPEKPLPAVSPDPQ------------
NOV14h  PWPLDP------------------ENSHEPSSHPEKPLPAVSPDPQ------------

NOV14a  DNISSSFSADQVQMPRSCLCGEPWGGHVGRKEGSKRGGPRLGERPVWSPGSPGCAAELEK
NOV14b  --------ADQVQMPRSCLW----------------------------------------
NOV14c  ------------------------------------------------------------
NOV14d  ------------------------------------------------------------
NOV14e  ------------------------------------------------------------
NOV14f  DNISSSFSADQVQMPRSCLCGEPWGGHVGRKEGSKRGGPRLGERPVWSPGSPGCAAELEK
NOV14g  --------ADQVQMPRSCLCGEPWGGHVGRKEGSKRGGPRLGERPVWSPGSPGCAAELEK
NOV14h  ---------DQVQMPRSCLW----------------------------------------

NOV14a  RSSAERSWHGANSRKAGGTQK
NOV14b  ---------------------
NOV14c  ---------------------
NOV14d  ---------------------
NOV14e  ---------------------
NOV14f  RSSAERSWHGANSRKAGGTQK
NOV14g  RSSAERSWHGANSRKAGGTQK
NOV14h  ---------------------

NOV14a  (SEQ ID NO:  150)
NOV14b  (SEQ ID NO:  152)
NOV14c  (SEQ ID NO:  154)
NOV14d  (SEQ ID NO:  156)
NOV14e  (SEQ ID NO:  158)
NOV14f  (SEQ ID NO:  160)
NOV14g  (SEQ ID NO:  162)
```

NOV14h (SEQ ID NO: 164)

Further analysis of the NOV14a protein yielded the following properties shown in Table 14C.

Table 14C. Protein Sequence Properties NOV14a

SignalP analysis:	Cleavage site between residues 30 and 31

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 9;  pos.chg 4;  neg.chg 0
      H-region:  length 37;  peak value  11.70
      PSG score:   7.30

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   5.85
      possible cleavage site: between 27 and 28

>>> Seems to have a cleavable signal peptide (1 to 27)

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 28
      Tentative number of TMS(s) for the threshold  0.5:   2
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -6.16   Transmembrane  702 - 718
      PERIPHERAL  Likelihood =  0.63 (at 126)
      ALOM score:  -6.16  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 13
      Charge difference: -5.0   C( 0.0) - N( 5.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 1a  (cytoplasmic tail 719 to 914)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       4       Hyd Moment(75):  2.69
      Hyd Moment(95):  8.42    G content:       6
      D/E content:     1       S/T content:     2
      Score: -3.80

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 19  ARG|TP

NUCDISC: discrimination of nuclear localization signals
      pat4: RPRR (4) at    4
      pat7: PRRARGT (5) at    5
      pat7: PQSRGRR (3) at  197
      bipartite: none
      content of basic residues:   8.9%
      NLS Score:  0.47
```

```
KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: GWRP

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: too long tail

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 70.6

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       44.4 %: extracellular, including cell wall
       22.2 %: Golgi
       22.2 %: endoplasmic reticulum
       11.1 %: plasma membrane

>> prediction for CG50367-01 is exc (k=9)
```

A search of the NOV14a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 14D.

Table 14D. Geneseq Results for NOV14a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABJ10597	Human novel protein NOV7a SEQ ID NO: 22 - Homo sapiens, 914 aa. [WO200259315-A2, 01-AUG-2002]	1..914 1..914	914/914 (100%) 914/914 (100%)	0.0
ABJ10598	Human novel protein NOV7b SEQ ID NO: 24 - Homo sapiens, 895 aa. [WO200259315-A2, 01-AUG-2002]	1..914 1..895	894/914 (97%) 895/914 (97%)	0.0
ABJ10599	Human novel protein NOV7c SEQ ID NO: 26 - Homo sapiens, 874 aa. [WO200259315-A2, 01-AUG-2002]	1..914 1..874	873/914 (95%) 874/914 (95%)	0.0
ABU58632	Human PRO polypeptide #233 - Homo sapiens, 813 aa. [US2003027272-A1, 06-FEB-2003]	1..852 1..812	812/852 (95%) 812/852 (95%)	0.0
ABU57163	Human PRO polypeptide #233 - Homo sapiens, 813 aa. [US2003027280-A1, 06-FEB-2003]	1..852 1..812	812/852 (95%) 812/852 (95%)	0.0

In a BLAST search of public sequence databases, the NOV14a protein was found to have homology to the proteins shown in the BLASTP data in Table 14E.

Table 14E. Public BLASTP Results for NOV14a				
Protein Accession Number	**Protein/Organism/Length**	**NOV14a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9BZ11	ADAM 33 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 33) - Homo sapiens (Human), 813 aa.	1..852 1..812	812/852 (95%) 812/852 (95%)	0.0
Q8N0W6	A disintegrin and metalloprotease domain 33 - Homo sapiens (Human), 812 aa.	1..852 1..811	811/852 (95%) 811/852 (95%)	0.0
CAC33153	Sequence 1 from Patent WO0109293 - Homo sapiens (Human), 802 aa.	1..661 1..661	661/661 (100%) 661/661 (100%)	0.0
Q8R5G5	Metalloprotease/disintegrin ADAM33 - Mus musculus (Mouse), 797 aa.	1..805 1..784	571/806 (70%) 632/806 (77%)	0.0
Q923W9	ADAM 33 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 33) - Mus musculus (Mouse), 797 aa.	1..805 1..784	570/806 (70%) 631/806 (77%)	0.0

PFam analysis indicates that the NOV14a protein contains the domains shown in the Table 14F.

Table 14F. Domain Analysis of NOV14a			
Pfam Domain	**NOV14a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Pep_M12B_propep	80..198	44/122 (36%) 93/122 (76%)	1.3e-31
Reprolysin	210..409	84/206 (41%) 163/206 (79%)	3e-88
disintegrin	426..501	38/76 (50%) 52/76 (68%)	2.5e-23
EGF	653..680	9/47 (19%) 20/47 (43%)	0.43

Example 15.

The NOV15 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 15A.

5

Table 15A. NOV15 Sequence Analysis

NOV15a, CG50718-02 DNA Sequence	SEQ ID NO: 165	6881 bp
	ORF Start: ATG at 16	ORF Stop: TAA at 6859

AAAAATGTAATAAAGATGGATTTTCTTATCATTTTTCTTTTACTTTTTATTGGGACTTCAGAGACACA
GGTAGATGTTTCCAATGTCGTTCCTGGTACTAGGTACGATATAACCATCTCTTCAATTTCTACAACAT
ACACCTCACCTGTTACTAGAATAGGGGCTTCTAATGAACCAGGGCCTCCAGTCTTCCTAGCCGGGGAA
AGAGTCGGATCTGCTGGGATTCTTCTGTCTTGGAATACACCACCTAATCCAAATGGAAGGATTATATC
TTACATTGTCAAATATAAGGAAGTTTGTCCGTGGATGCAAACAGTATATACACAAGTCAGATCAAAGC
CAGACAGTCTGGAAGTTCTTCTTACTAATCTTAATCCTGGAACAACATATGAAATTAAGGTAGCTGCT
GAAAACAGTGCTGGCATTGGAGTGTTTAGTGATCCATTTCTCTTCCAAACTGCAGAAAGTGCTCCAGG
AAAAGTGGTGGATTTCACAGGTGAGGCTGTCCCGTTCAGCAGTAAGCTGATGTGGTATACCTCGGCAA
CCAAAAAAAAAATTACCAGCTTCAAGATTAGTGTCAAGCATAACAGAAGTGGGATAGTAGTGAAAGAA
GTGTCAATCAGAGTGGAGTGCATTTTAAGTGCTTCCCTTCCTTTGCACTGCAACGAGAATAGTGAATC
TTTTTTATGGAGTACAGCCAGCCCTTCTCCAACCCTTGGTAGAGTTACACCTCCATCGCGTACCACAC
ATTCATCAAGCACGTTGACACAGAATGAGATCAGCTCTGTGAAAGAGCCTATCAGTTTTGTAGTGACA
CACTTGAGACCTTATACAACATATCTTTTTGAAGTTTCAGCTGCTACAACTGAAGCAGGTTATATTGA
TAGTACGATTGTCAGAACACCAGAATCAGTGCCTGAAGGACCACCACAAAACTGCGTAACAGGCAACA
TCACAGGAAAGTCCTTTTCAATTTTATGGGACCCACCAACTATAGTAACAGGGAAATTTAGTTATAGA
GTTGAATTATATGGACCATCAGCAGGTCGCATTTTGGATAACAGCACAAAAGACCTCAAGTTTGCATT
CACTAACCTAACACCATTTACAATGTATGATGTCTATATTGCGGCTGAAACCAGTGCAGGGACTGGGC
CCAAGTCAAATATTTCAGTATTCACTCCACCAGATGTTCCAGGGGCAGTGTTTGATTTACAACTTGCA
GAGGTAGAATCCACGCAAGTAAGAATTACTTGGAAGAAACCACGACAACCAAATGGAATTATTAACCA
ATACCGAGTGAAAGTGCTAGTTCCAGAGACAGGAATAATTTTGGAAAATACTTTGCTCACTGGAAATA
ATGAGATAAATGACCCCATGGCTCCAGAAATTGTGAACATAGTACAGCCAATGGTAGGATTATATGAG
GGTTCAGCAGAGATGTCGTCTGACCTTCACTCACTTGCTACATTTATATATAACAGCCATCCAGATAA
AAACTTTCCTGCAAGGAATAGAGCTGAAGACCAGACTTCACCAGTTGTAACTACAAGGAATCAGTATA
TTACTGACATTGCAGCTGAACAGCTGACTTATGTTCTTATCAGATTAAGGAGATTTTGGGCTGAGACA
ATGGGGTTTTCTAGATATACAATCATGTCATCTGCAAGCAGGGACAATTTGACTTCCCCAGGCCCTTT
GTCAGCCCAAAATTTCAGAGTTACACATGTTACCATAACAGAAGTATTTTTACACTGGGATCCTCCAG
ATCCTGTATTTTTTCATCATTACCTTATCACTATTTTGGATGTTGAAAACCAATCCAAGAGTATTATT
TTAAGGACATTAAACAGTTTGTCTCTTGTCCTTATAGGGTTAAAGAAATACACAAAATACAAAATGAG
AGTGGCAGCCTCAACCCACGTTGGAGAAAGTTCTTTGTCTGAAGAAAATGACATCTTTGTGAGAACTT
CAGAAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTAATTGATGTTACCGCAGATGAAATAAGG
TTGAAGTGGTCACCACCCGAAAAGCCCAATGGGATCATTATTGCTTATGAAGTGCTATATAAAAATAT
AGATACTTTATATATGAAGAACACATCAACAACAGACATAATATTAAGGAACTTAAGACCTCACACCC
TCTATAACATTTCTGTAAGGTCTTACACCAGATTTGGTCATGGCAATCAGGTATCTTCTTTACTCTCT
GTAAGGACTTCGGAGTCAGTGCCTGATAGTGCACCAGAAAATATCACTTACAAAAATATTTCTTCTGG
AGAGATTGAGCTATCATTCCTTCCCCCAAGTAGTCCCAATGGAATCATACAAAAATATACAATTTATC
TCAAGAGAAGTAATGGAAATGAGGAAAGAACTATAAATACAACCTCTTTAACCCAAAACATTAAAGGT
CTGAAGAAATATACCCAATATATCATTGAGGTGTCTGCTAGTACACTCAAAGGTGAAGGAGTTCGGAG
TGCTCCCATAAGTATACTGACGGAGGAAGATGCTCCTGATTCTCCCCCTCAAGACTTCTCTGTAAAAC
AGTTGTCTGGTGTCACGGTGAAGTTGTCATGGCAACCACCCCTGGAGCCAAATGGAATTATCCTTTAT
TACACAGTTTATGTCTGGAGATCATCATTAAAAACTATTAATGTCACTGAAACATCATTGGAGTTATC
AGATTTGGATTATAATGTTGAATACAGTGCTTATGTAACAGCTAGCACCAGATTTGGTGATGGGAAAA
CAAGAAGCAATATCATTAGCTTTCAAACACCAGAGGGACCAAGCGATCCTCCCAAAGATGTTTATTAT
GCAAACCTCAGTTCTTCATCAATAATTCTTTTCTGGACACCTCCTTCAAAACCTAATGGGATTATACA
ATATTACTCTGTTTATTACAGAAATACTTCAGGTACTTTTATGCAGAATTTTACACTCCATGAAGTAA
CCAATGACTTTGACAATATGACTGTATCCACAATTATAGATAAACTGACAATATTCAGCTACTATACA
TTTTGGTTAACAGCAAGTACTTCAGTTGGAAATGGGAATAAAAGCAGTGACATCATTGAAGTATACAC
AGATCAAGACGTACCTGAAGGGTTTGTTGGAAACCTGACTTACGAATCCATTTCGTCAACTGCAATAA

ATGTAAGCTGGGTCCCACCGGCTCAACCAAACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAG
CAGACTCCTCGCCATGTGAGACCACCTCTTGTTACATATGAGAGAAGCATATATTTTGATAATCTGGA

AAAATACACTGATTATATATTAAAAATTACTCCATCAACAGAAAAGGGATTCTCTGATACCTATACTG
CCCAGCTATACATCAAGACTGAAGAAGATATCCCAGAAACTTCACCAATAATCAACACTTTTAAAAAC
CTTTCCTCTACCTCAGTTCTCTTATCATGGGATCCCCCAGTAAAGCCAAATGGTGCAATAATAAGTTA
TGATTTAACTTTACAAGGACCAAATGAAAATTATTCTTTCATTACTTCTGATAATTACATAATATTGG
AAGAGCTTTCACCATTTACATTATATAGCTTTTTTGCTGCCGCAAGAACTAGAAAAGGACTTGGTCCT
TCCAGTATTCTTTTCTTTTACACAGATGAGTCAGTGCCGTTAGCACCTCCACAAAATTTGACTTTAAT
CAACTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCCTCTTCCAGGTGGTATTGTTAAAG
TATATAGTTTTAAAATTCATGAACATGAAACTGACACTATATATTATAAGAATATATCAGGATTTAAA
ACTGAAGCCAAACTTGTTGGACTGGAACCAGTCAGCACCTACTCTATCCGTGTATCTGCGTTCACCAA
AGTTGGAAATGGCAATCAATTTAGTAATGTAGTAAAATTCACAACCCAAGAATCAGTTCCAGATGTCG
TGCAGAATATGCAGTGCATGGCAACTAGCTGGCAGTCAGTTTTAGTGAAATGGGATCCACCCAAAAAG
GCAAATGGAATAATAACGCAGTATATGGTAACAGTTGAAAGGAATTCTACAAAAGTTTCTCCCCAAGA
TCACATGTACACTTTCATAAAGCTTCTTGCCAATACCTCATATGTCTTTAAAGTAAGAGCTTCAACCT
CAGCTGGTGAAGGTGATGAAAGCACATGCCATGTCAGCACACTACCTGAAACAGTTCCCAGTGTTCCC
ACAAATATTGCTTTTTCTGATGTTCAGTCAACTAGTGCAACATTGACATGGATAAGACCTGACACTAT
CCTTGGCTACTTTCAAAATTACAAAATTACCACTCAACTTCGTGCTCAAAAATGCAAAGAATGGGAAT
CCGAAGAATGTGTTGAATATCAAAAAATTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTA
TATGGATTAAAGAAATTTAGATGGTATAGATTCCAAGTGGCTGCCAGCACCAATGCTGGCTATGGCAA
TGCTTCAAACTGGATTTCTACAAAAACTCTGCCTGGCCCTCCAGATGGTCCTCCTGAAAATGTTCATG
TAGTAGCAACATCACCTTTTAGCATCAGCATAAGCTGGAGTGAACCTGCTGTCATTACTGGACCAACA
TGTTATCTGATTGATGTCAAATCGGTAGATAATGATGAATTTAATATATCCTTCATCAAGTCAAATGA
AGAAAATAAAACCATAGAAATTAAAGATTTAGAAATATTCACAAGGTATTCTGTAGTGATCACTGCAT
TTACTGGGAACATTAGTGCTGCATATGTAGAAGGGAAGTCAAGTGCTGAAATGATTGTTACTACTTTA
GAATCAGCCCCAAAGGACCCACCTAACAACATGACATTTCAGAAGATACCAGATGAAGTTACAAAATT
TCAATTAACGTTCCTTCCTCCTTCTCAACCTAATGGAAATATCCAAGTATATCAAGCTCTGGTTTACC
GAGAAGATGATCCTACTGCTGTCCAGATTCACAACCTCAGTATTATACAGAAAACCAACACATTCGTC
ATTGCAATGCTAGAAGGACTAAAAGGTGGACATACATACAATATCAGTGTTTACGCAGTCAATAGTGC
TGGTGCAGGTCCAAAGGTTCCGATGAGAATAACCATGGATATCAAAGCTCCAGCACGACCAAAAACCA
AACCAACCCCTATTTATGATGCCACAGGAAAACTGCTTGTGACTTCAACAACAATTACAATCAGAATG
CCAATATGTTACTACAGTGATGATCATGGACCAATAAAAAATGTACAAGTGCTTGTGACAGAAACAGG
AGCTCAGCATGATGGAAATGTAACAAAGTGGTATGATGCATATTTTAATAAAGCAAGGCCATATTTTA
CAAATGAAGGCTTTCCTAACCCTCCATGTACAGAAGGAAAGACAAAGTTTAGTGGCAATGAAGAAATC
TACATCATAGGTGCTGATAATGCATGCATGATTCCTGGCAATGAAGACAAAATTTGCAATGGACCACT
GAAACCAAAAAAGCAATACTTATTTAAATTTAGAGCTACAAATATTATGGGACAATTTACTGACTCTG
ATTATTCTGACCCTGTTAAGACTTTAGGCGAAGGACTTTCAGAAAGAACCGTAGAGATCATTCTTTCC
GTCACTTTGTGTATCCTTTCAATAATTCTCCTTGGAACAGCTATTTTTGCATTTGCAAGAATTCGACA
GAAGCAGAAAGAAGGTGGCACATACTCTCCTCAGGATGCAGAAATTATTGACACTAAATTGAAGCTGG
ATCAGCTCATCACAGTGGCAGACCTGGAACTGAAGGACGAGAGATTAACGCGGCCAATAAGCAAGAAA
TCCTTCCTGCAACATGTTGAAGAGCTTTGCACAAACAACAACCTAAAGTTTCAAGAAGAATTTTCGGA
ATTACCAAAATTTCTTCAGGATCTTTCTTCAACTGATGCTGATCTGCCTTGGAATAGAGCAAAAAACC
GCTTCCCAAACATAAAACCATATAATAATAACAGAGTAAAGCTGATAGCTGACGCTAGTGTTCCAGGT
TCGGATTATATTAATGCCAGCTATATTTCTGGTTATTTATGTCCAAATGAATTTATTGCTACTCAAGG
TCCACTACCAGGAACAGTTGGAGATTTTTGGAGAATGGTGTGGGAAACCAGAGCAAAAACATTAGTAA
TGCTAACACAGTGTTTTGAAAAAGGACGGATCAGATGCCATCAGTATTGGCCAGAGGACAACAAGCCA
GTTACTGTCTTTGGAGATATAGTGATTACAAAGCTAATGGAGGATGTTCAAATAGATTGGACTATCAG
GGATCTGAAAATTGAAAGGCATGGGGATTGCATGACTGTTCGACAGTGTAACTTTACTGCCTGGCCAG
AGCATGGGGTTCCTGAGAACAGCGCCCCTCTAATTCACTTTGTGAAGTTGGTTCGAGCAAGCAGGGCA
CATGACACCACACCTATGATTGTTCACTGCAGTGCTGGAGTTGGAAGAACTGGAGTTTTTATTGCTCT
GGACCATTTAACACAACATATAAATGACCATGATTTTGTGGATATATATGGACTAGTAGCTGAACTGA

GAAGTGAAAGAATGTGCATGGTGCAGAATCTGGCACAGTATATCTTTTTACACCAGTGCATTCTGGAT
CTCTTATCAAATAAGGGAAGTAATCAGCCCATCTGTTTTGTTAACTATTCAGCACTTCAGAAGATGGA
CTCTTTGGACGCCATGGAAGGTGGTGATGTTGAGCTTGAATGGGAAGAAACCACTATGTAAATATTCA
GACCAAAGGATAC

NOV15a, CG50718-02 Protein Sequence	SEQ ID NO: 166	2281 aa	MW at 255030.3kD

MDFLIIFLLLFIGTSETQVDVSNVVPGTRYDITISSISTTYTSPVTRIGASNEPGPPVFLAGERVGSA
GILLSWNTPPNPNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLNPGTTYEIKVAAENSAG
IGVFSDPFLFQTAESAPGKVVDFTGEAVPFSSKLMWYTSATKKKITSFKISVKHNRSGIVVKEVSIRV
ECILSASLPLHCNENSESFLWSTASPSPTLGRVTPPSRTTHSSSTLTQNEISSVKEPISFVVTHLRPY
TTYLFEVSAATTEAGYIDSTIVRTPESVPEGPPQNCVTGNITGKSFSILWDPPTIVTGKFSYRVELYG
PSAGRILDNSTKDLKFAFTNLTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVPGAVFDLQLAEVEST
QVRITWKKPRQPNGIINQYRVKVLVPETGIILENTLLTGNNEINDPMAPEIVNIVQPMVGLYEGSAEM
SSDLHSLATFIYNSHPDKNFPARNRAEDQTSPVVTTRNQYITDIAAEQLTYVLIRLRRFWAETMGFSR
YTIMSSASRDNLTSPGPLSAQNFRVTHVTITEVFLHWDPPDPVFFHHYLITILDVENQSKSIILRTLN
SLSLVLIGLKKYTKYKMRVAASTHVGESSLSEENDIFVRTSEDEPESSPQDVEVIDVTADEIRLKWSP
PEKPNGIIIAYEVLYKNIDTLYMKNTSTTDIILRNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSE
SVPDSAPENITYKNISSGEIELSFLPPSSPNGIIQKYTIYLKRSNGNEERTINTTSLTQNIKGLKKYT
QYIIEVSASTLKGEGVRSAPISILTEEDAPDSPPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVYV
WRSSLKTINVTETSLELSDLDYNVEYSAYVTASTRFGDGKTRSNIISFQTPEGPSDPPKDVYYANLSS
SSIILFWTPPSKPNGIIQYYSVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKLTIFSYYTFWLTA
STSVGNGNKSSDIIEVYTDQDVPEGFVGNLTYESISSTAINVSWVPPAQPNGLVFYYVSLILQQTPRH
VRPPLVTYERSIYFDNLEKYTDYILKITPSTEKGFSDTYTAQLYIKTEEDIPETSPIINTFKNLSSTS
VLLSWDPPVKPNGAIISYDLTLQGPNENYSFITSDNYIILEELSPFTLYSFFAAARTRKGLGPSSILF
FYTDESVPLAPPQNLTLINCTSDFVWLKWSPSPLPGGIVKVYSFKIHEHETDTIYYKNISGFKTEAKL
VGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFTTQESVPDVVQNMQCMATSWQSVLVKWDPPKKANGII
TQYMVTVERNSTKVSPQDHMYTFIKLLANTSYVFKVRASTSAGEGDESTCHVSTLPETVPSVPTNIAF
SDVQSTSATLTWIRPDTILGYFQNYKITTQLRAQKCKEWESEECVEYQKIQYLYEAHLTEETVYGLKK
FRWYRFQVAASTNAGYGNASNWISTKTLPGPPDGPPENVHVVATSPFSISISWSEPAVITGPTCYLID
VKSVDNDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAFTGNISAAYVEGKSSAEMIVTTLESAPK
DPPNNMTFQKIPDEVTKFQLTFLPPSQPNGNIQVYQALVYREDDPTAVQIHNLSIIQKTNTFVIAMLE
GLKGGHTYNISVYAVNSAGAGPKVPMRITMDIKAPARPKTKPTPIYDATGKLLVTSTTITIRMPICYY
SDDHGPIKNVQVLVTETGAQHDGNVTKWYDAYFNKARPYFTNEGFPNPPCTEGKTKFSGNEEIYIIGA
DNACMIPGNEDKICNGPLKPKKQYLFKFRATNIMGQFTDSDYSDPVKTLGEGLSERTVEIILSVTLCI
LSIILLGTAIFAFARIRQKQKEGGTYSPQDAEIIDTKLKLDQLITVADLELKDERLTRPISKKSFLQH
VEELCTNNNLKFQEEFSELPKFLQDLSSTDADLPWNRAKNRFPNIKPYNNNRVKLIADASVPGSDYIN
ASYISGYLCPNEFIATQGPLPGTVGDFWRMVWETRAKTLVMLTQCFEKGRIRCHQYWPEDNKPVTVFG
DIVITKLMEDVQIDWTIRDLKIERHGDCMTVRQCNFTAWPEHGVPENSAPLIHFVKLVRASRAHDTTP
MIVHCSAGVGRTGVFIALDHLTQHINDHDFVDIYGLVAELRSERMCMVQNLAQYIFLHQCILDLLSNK
GSNQPICFVNYSALQKMDSLDAMEGGDVELEWEETTM

NOV15b, CG50718-06 DNA Sequence	SEQ ID NO: 167	6900 bp
	ORF Start: ATG at 1	ORF Stop: TAA at 6898

ATGGATTTTCTTATCATTTTTCTTTTACTTTTTATTGGGACTTCAGAGACACAGGTAGATGTTTCCAA
TGTCGTTCCTGGTACTAGGTACGATATAACCATCTCTTCAATTTCTACAACATACACCTCACCTGTTA
CTAGAATAGTGACAACAAATGTAACAAAACCAGGGCCTCCAGTCTTCCTAGCCGGGGAAAGAGTCGGA
TCTGCTGGGATTCTTCTGTCTTGGAATACACCACCTAATCCAAATGGAAGGATTATATCTTACATTGT
CAAATATAAGGAAGTTTGTCCGTGGATGCAAACAGTATATACACAAGTCAGATCAAAGCCAGACAGTC
TGGAAGTTCTTCTTACTAATCTTAATCCTGGAACAACATATGAAATTAAGGTTGCTGCTGAAAACAGT
GCTGGCATTGGAGTGTTTAGTGATCCATTTCTCTTCCAAACTGCAGAAAGTGCTCCAGGAAAAGTGGT
GAATCTCACAGTTGAGGCCTACAACGCTTCAGCAGTTAAGCTGATTTGGTATTTACCTCGGCAACCAA
ATGGCAAAATTACCAGCTTCAAGATTAGTGTCAAGCATGCCAGAAGTGGGATAGTAGTGAAAGATGTC
TCAATCAGAGTAGAGGACATTTTGACTGGGAAATTGCCAGAATGCAATGAGAATAGTGAATCTTTTTT
ATGGAGTACAGCCAGCCCTTCTCCAACCCTTGGTAGAGTTACACCTCCATCGCGTACCACACATTCAT
CAAGCACGTTGACACAGAATGAGATCAGCTCTGTGTGGAAAGAGCCTATCAGTTTTGTAGTGACACAC
TTGAGACCTTATACAACATATCTTTTTGAAGTTTCAGCTGTTACAACTGAAGCAGGTTATATTGATAG
TACGATTGTCAGAACACCAGAATCAGTGCCTGAAGGACCACCACAAAACTGCGTAACAGGCAACATCA
CAGGAAAGTCCTTTTCAATTTTATGGGACCCACCAACTATAGTAACAGGGAAATTTAGTTATAGAGTT
GAATTATATGGACCATCAGGTCGCATTTTGGATAACAGCACAAAAGACCTCAAGTTTGCATTCACTAA
CCTAACACCATTTACAATGTATGATGTCTATATTGCGGCTGAAACCAGTGCAGGGACTGGGCCCAAGT

CAAATATTTCAGTATTCACTCCACCAGATGTTCCAGGGGCAGTGTTTGATTTACAACTTGCAGAGGTA
GAATCCACGCAAGTAAGAATTACTTGGAAGAAACCACGACAACCAAATGGAATTATTAACCAATACCG
AGTGAAAGTGCTAGTTCCAGAGACAGGAATAATTTTGGAAAATACTTTGCTCACTGGAAATAATGAGT
ATATAAATGACCCCATGGCTCCAGAAATTGTGAACATAGTAGAGCCAATGGTAGGATTATATGAGGGT
TCAGCAGAGATGTCGTCTGACCTTCACTCACTTGCTACATTTATATATAACAGCCATCCAGATAAAAA
CTTTCCTGCAAGGAATAGAGCTGAAGACCAGACTTCACCAGTTGTAACTACAAGGAATCAGTATATTA
CTGACATTGCAGCTGAACAGCTGTCTTATGTTATCAGGAGACTTGTACCTTTCACTGAGCACATGATT
AGTGTATCTGCTTTCACCATCATGGGAGAAGGACCACCAACAGTTCTCAGTGTTAGGACACGTCAGCA
AGTGCCAAGCTCCATTAAAATTATAAACTATAAAAATATTAGTTCTTCATCTATTTTGTTATATTGGG
ATCCTCCAGAATATCCCAATGGAAAAATAACTCACTATACGATTTATGCAATGGAATTGGATACAAAC
AGAGCATTCCAGATAACTACCATAGATAACAGCTTTCTCATAACAGGGTTAAAGAAATACACAAAATA
CAAAATGAGAGTGGCAGCCTCAACCCACGATGGAGAAAGTTCTTTGTCTGAAGAAAATGACATCTTTG
TGAGAACTTCAGAAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTAATTGATGTTACCGCAGAT
GAAATAAGGTTGAAGTGGTCACCACCCGAAAAGCCCAATGGGATCATTATTGCTTATGAAGTGCTATA
TAAAAATATAGATACTTTATATATGAAGAACACATCAACAACAGACATAATATTAAGGAACTTAAGAC
CTCACACCCTCTATAACATTTCTGTAAGGTCTTACACCAGATTTGGTCATGGCAATCAGGTATCTTCT
TTACTCTCTGTAAGGACTTCGGAGACTGTGCCTGATAGTGCACCAGAAAATATCACTTACAAAAATAT
TTCTTCTGGAGAGATTGAGCTATCATTCCTTCCCCCAAGTAGTCCCAATGGAATCATAAAAAAATATA
CAATTTATCTCAAGAGAAGTAATGGAAATGAGGAAAGAACTATAAATACAACCTCTTTAACCCAAAAC
ATTAAAGTACTGAAGAAATATACCCAATATATCATTGAGGTGTCTGCTAGTACACTGAAAGGTGAAGG
AGTTCGGAGTGCTCCCATAAGTATACTGACGGAGGAAGATGCTCCTGATTCTCCCCCTCAAGACTTCT
CTGTAAAACAGTTGTCTGGTGTCACGGTGAAGTTGTCATGGCAACCACCCCTGGAGCCAAATGGAATT
ATCCTTTATTACACAGTTTATGTCTGGAATAGATCATCATTAAAAACTATTAATGTCACTGAAACATC
ATTGGAGTTATCAGATTTGGATTATAATGTTGAATACAGTGCTTATGTAACAGCTAGCACCAGATTTG
GTGATGGGAAAACAAGAAGCAATATCATTAGCTTTCAAACACCAGAGGGAGCACCAAGCGATCCTCCC
AAAGATGTTTATTATGCAAACCTCAGTTCTTCATCAATAATTCTTTTCTGGACACCTCCTTCAAAACC
TAATGGGATTATACAATATTACTCTGTTTATTACAGAAATACTTCAGGTACTTTTATGCAGAATTTTA
CACTCCATGAAGTAACCAATGACTTTGACAATATGACTGTATCCACAATTATAGATAAACTGACAATA
TTCAGCTACTATACATTTTGGTTAACAGCAAGTACTTCAGTTGGAAATGGGAATAAAAGCAGTGACAT
CATTGAAGTATACACAGATCAAGACGTCCCTGAAGGGTTTGTTGGAAACCTGACTTACGAATCCATTT
CGTCAACTGCAATAAATGTAAGCTGGGTCCCACCGGCTCAACCAAACGGTCTAGTCTTCTACTATGTT
TCACTGATCTTACAGCAGACTCCTCGCCATGTGAGACCACCTCTTGTTACATATGAGAGAAGCATATA

TTTTGATAATCTGGAAAAATACACTGATTATATATTAAAAATTACTCCATCAACAGAAAAGGGATTCT
CTGATACCTATACTGCCCAGCTATACATCAAGACTGAAGAAGATGTCCCAGAAACTTCACCAATAATC
AACACTTTTAAAAACCTTTCCTCTACCTCAGTTCTCTTATCATGGGATCCCCCAGTAAAGCCAAATGG
TGCAATAATAAGTTATGATTTAACTTTACAAGGACCAAATGAAAATTATTCTTTCATTACTTCTGATA
ATTACATAATATTGGAAGAGCTTTCACCATTTACATTATATAGCTTTTTTGCTGCCGCAAGAACTAGA
AAAGGACTTGGTCCTTCCAGTATTCTTTTCTTTTACACAGATGAGTCAGTGCCGTTAGCACCTCCACA
AAATTTGACTTTAATCAACTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCCTCTTCCAG
GTGGTATTGTTAAAGTATATAGTTTTAAAATTCATGAACATGAAACTGACACTATATATTATAAGAAT
ATATCAGGATTTAAAACTGAAGCCAAACTTGTTGGACTGGAACCAGTCAGCACCTACTCTATCCGTGT
ATCTGCGTTCACCAAAGTTGGAAATGGCAATCAATTTAGTAATGTAGTAAAATTCACAACCCAAGAAT
CAGTTCCAGATGTCGTGCAGAATATGCAGTGCATGGCAACTAGCTGGCAGTCAGTTTTAGTGAAATGG
GATCCACCCAAAAAGGCAAATGGAATAATAACGCAGTATATGGTAACAGTTGAAAGGAATTCTACAAA
AGTTTCTCCCCAAGATCACATGTACACTTTCATAAAGCTTCTTGCCAATACCTCATATGTCTTTAAAG
TAAGAGCTTCAACCTCAGCTGGTGAAGGTGATGAAAGCACATGCCATGTCAGCACACTACCTGAAACA
GTTCCCAGTGTTCCCACAAATATTGCTTTTTCTGATGTTCAGTCAACTAGTGCAACATTGACATGGAT
AAGACCTGACACTATCCTTGGCTACTTTCAAAATTACAAAATTACCACTCAACTTCGTGCTCAAAAAT
GCAAAGAATGGGAATCCGAAGAATGTGTTGAATATCAAAAAATTCAATACCTCTATGAAGCTCACTTA
ACTGAAGAGACAGTATATGGATTAAAGAAATTTAGATGGTATAGATTCCAAGTGGCTGCCAGCACCAA
TGCTGGCTATGGCAATGCTTCAAACTGGATTTCTACAAAAACTCTGCCTGGCCCTCCAGATGGTCCTC
CTGAAAATGTTCATGTAGTAGCAACATCACCTTTTAGCATCAGCATAAGCTGGAGTGAACCTGCTGTC
ATTACTGGACCAACATGTTATCTGATTGATGTCAAATCGGTAGATAATGATGAATTTAATATATCCTT
CATCAAGTCAAATGAAGAAAATAAAACCATAGAAATTAAAGATTTAGAAATATTCACAAGGTATTCTG
TAGTGATCACTGCATTTACTGGGAACATTAGTGCTGCATATGTAGAAGGGAAGTCAAGTGCTGAAATG
ATTGTTACTACTTTAGAATCAGCCCCAAAGGACCCACCTAACAACATGACATTTCAGAAGATACCAGA
TGAAGTTACAAAATTTCAATTAACGTCCCTTCCTCCTTCTCAACCTAATGGAAATATCCAAGTATATC
AAGCTCTGGTTTACCGAGAAGATGATCCTACTGCTGTCCAGATTCACAACCTCAGTATTATACAGAAA
ACCAACACATTCGTCATTGCAATGCTAGAAGGACTAAAAGGTGGACATACATACAATATCAGTGTTTA

CGCAGTCAATAGTGCTGGTGCAGGTCCAAAGGTTCCGATGAGAATAACCATGGATATCAAAGCTCCAG
CACGACCAAAAACCAAACCAACCCCTATTTATGATGCCACAGGAAAACTGCTTGTGACTTCAACAACA
ATTACAATCAGAATGCCAATATGTTACTACAGTGATGATCATGGACCAATAAAAAATGTACAAGTGCT
TGTGACAGAAACAGGAGCTCAGCATGATGGAAATGTAACAAAGTGGTATGATGCATATTTTAATAAAG
CAAGGCCATATTTTACAAATGAAGGCTTTCCTAACCCTCCATGTACAGAAGGAAAGACAAAGTTTAGT
GGCAATGAAGAAATCTACATCATAGGTGCTGATAATGCATGCATGATTCCTGGCAATGAAGACAAAAT
TTGCAATGGACCACTGAAACCAAAAAAGCAATACTTATTTAAATTTAGAGCTACAAATATTATGGGAC
AATTTACTGACTCTGATTATTCTGACCCTGTTAAGACTTTAGGCGAAGGACTTTCAGAAAGAACCCTA
GAGATCATTCTTTCCGTCACTTTGTGTATCCTTTCAATAATTCTCCTTGGAACAGCTATTTTTGCATT
TGCAAGAATTCGACAGAAGCAGAAAGAAGGTGGCACATACTCTCCTCAGGATGCAGAAATTATTGACA
CTAAATTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTGGAACTGAAGGACGAGAGATTAACGCGG
TTACTTAGTTATAGAAAATCCATCAAGCCAATAAGCAAGAAATCCTTCCTGCAACATGTTGAAGAGCT
TTGCACAAACAACAACCTAAAGTTTCAAGAAGAATTTTCGGAATTACCAAAATTTCTTCAGGATCTTT
CTTCAACTGATGCTGATCTGCCTTGGAATAGAGCAAAAAACCGCTTCCCAAACATAAAACCATATAAT
AATAACAGAGTAAAGCTGATAGCTGACGCTAGTGTTCCAGGTTCGGATTATATTAATGCCAGCTATAT
TTCTGGTTATTTATGTCCAAATGAATTTATTGCTACTCAAGGTCCACTACCAGGAACAGTTGGAGATT
TTTGGAGAATGGTGTGGGAAACCAGAGCAAAAACATTAGTAATGCTAACACAGTGTTTTGAAAAAGGA
CGGATCAGATGCCATCAGTATTGGCCAGAGGACAACAAGCCAGTTACTGTCTTTGGAGATATAGTGAT
TACAAAGCTAATGGAGGATGTTCAAATAGATTGGACTATCAGGGATCTGAAAATTGAAAGGCATGGGG
ATTGCATGACTGTTCGACAGTGTAACTTTACTGCCTGGCCAGAGCATGGGGTTCCTGAGAACAGCGCC
CCTCTAATTCACTTTGTGAAGTTGGTTCGAGCAAGCAGGGCACATGACACCACACCTATGATTGTTCA
CTGTAGTGCTGGAGTTGGAAGAACTGGAGTTTTTATTGCTCTGGACCATTTAACACAACATATAAATG

ACCATGATTTTGTGGATATATATGGACTAGTAGCTGAACTGAGAAGTGAAAGAATGTGCATGGTGCAG
AATCTGGCACAGTATATCTTTTTACACCAGTGCATTCTGGATCTCTTATCAAATAAGGGAAGTAATCA
GCCCATCTGTTTTGTTAACTATTCAGCACTTCAGAAGATGGACTCTTTGGACGCCATGGAAGGTGATG
TTGAGCTTGAATGGGAAGAAACCACTATG**TAA**

NOV15b, CG50718-06 Protein Sequence	SEQ ID NO: 168	2299 aa	MW at 257243.9kD

MDFLIIFLLLFIGTSETQVDVSNVVPGTRYDITISSISTTYTSPVTRIVTTNVTKPGPPVFLAGERVG
SAGILLSWNTPPNPNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLNPGTTYEIKVAAENS
AGIGVFSDPFLFQTAESAPGKVVNLTVEAYNASAVKLIWYLPRQPNGKITSFKISVKHARSGIVVKDV
SIRVEDILTGKLPECNENSESFLWSTASPSPTLGRVTPPSRTTHSSSTLTQNEISSVWKEPISFVVTH
LRPYTTYLFEVSAVTTEAGYIDSTIVRTPESVPEGPPQNCVTGNITGKSFSILWDPPTIVTGKFSYRV
ELYGPSGRILDNSTKDLKFAFTNLTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVPGAVFDLQLAEV
ESTQVRITWKKPRQPNGIINQYRVKVLVPETGIILENTLLTGNNEYINDPMAPEIVNIVEPMVGLYEG
SAEMSSDLHSLATFIYNSHPDKNFPARNRAEDQTSPVVTTRNQYITDIAAEQLSYVIRRLVPFTEHMI
SVSAFTIMGEGPPTVLSVRTRQQVPSSIKIINYKNISSSSILLYWDPPEYPNGKITHYTIYAMELDTN
RAFQITTIDNSFLITGLKKYTKYKMRVAASTHDGESSLSEENDIFVRTSEDEPESSPQDVEVIDVTAD
EIRLKWSPPEKPNGIIIAYEVLYKNIDTLYMKNTSTTDIILRNLRPHTLYNISVRSYTRFGHGNQVSS
LLSVRTSETVPDSAPENITYKNISSGEIELSFLPPSSPNGIIKKYTIYLKRSNGNEERTINTTSLTQN
IKVLKKYTQYIIEVSASTLKGEGVRSAPISILTEEDAPDSPPQDFSVKQLSGVTVKLSWQPPLEPNGI
ILYYTVYVWNRSSLKTINVTETSLELSDLDYNVEYSAYVTASTRFGDGKTRSNIISFQTPEGAPSDPP
KDVYYANLSSSSIILFWTPPSKPNGIIQYYSVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKLTI
FSYYTFWLTASTSVGNGNKSSDIIEVYTDQDVPEGFVGNLTYESISSTAINVSWVPPAQPNGLVFYYV
SLILQQTPRHVRPPLVTYERSIYFDNLEKYTDYILKITPSTEKGFSDTYTAQLYIKTEEDVPETSPII
NTFKNLSSTSVLLSWDPPVKPNGAIISYDLTLQGPNENYSFITSDNYIILEELSPFTLYSFFAAARTR
KGLGPSSILFFYTDESVPLAPPQNLTLINCTSDFVWLKWSPSPLPGGIVKVYSFKIHEHETDTIYYKN
ISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFTTQESVPDVVQNMQCMATSWQSVLVKW
DPPKKANGIITQYMVTVERNSTKVSPQDHMYTFIKLLANTSYVFKVRASTSAGEGDESTCHVSTLPET
VPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKITTQLRAQKCKEWESEECVEYQKIQYLYEAHL
TEETVYGLKKFRWYRFQVAASTNAGYGNASNWISTKTLPGPPDGPPENVHVVATSPFSISISWSEPAV
ITGPTCYLIDVKSVDNDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAFTGNISAAYVEGKSSAEM
IVTTLESAPKDPPNNMTFQKIPDEVTKFQLTSLPPSQPNGNIQVYQALVYREDDPTAVQIHNLSIIQK
TNTFVIAMLEGLKGGHTYNISVYAVNSAGAGPKVPMRITMDIKAPARPKTKPTPIYDATGKLLVTSTT
ITIRMPICYYSDDHGPIKNVQVLVTETGAQHDGNVTKWYDAYFNKARPYFTNEGFPNPPCTEGKTKFS
GNEEIYIIGADNACMIPGNEDKICNGPLKPKKQYLFKFRATNIMGQFTDSDYSDPVKTLGEGLSERTL
EIILSVTLCILSIILLGTAIFAFARIRQKQKEGGTYSPQDAEIIDTKLKLDQLITVADLELKDERLTR
LLSYRKSIKPISKKSFLQHVEELCTNNNLKFQEEFSELPKFLQDLSSTDADLPWNRAKNRFPNIKPYN

NNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLPGTVGDFWRMVWETRAKTLVMLTQCFEKG
RIRCHQYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDCMTVRQCNFTAWPEHGVPENSA
PLIHFVKLVRASRAHDTTPMIVHCSAGVGRTGVFIALDHLTQHINDHDFVDIYGLVAELRSERMCMVQ
NLAQYIFLHQCILDLLSNKGSNQPICFVNYSALQKMDSLDAMEGDVELEWEETTM

NOV15c, 258979883 DNA Sequence	SEQ ID NO: 169	2220 bp
	ORF Start: at 1	ORF Stop: end of sequence

AGATCTCCTGAAGGACCACCACAAAACTGCGTAACAGGCAACATCACAGGAAAGTCCTTTTCAATTTT
ATGGGACCCACCAACTATAGTAACAGGGAAATTTAGTTATAGAGTTGAATTATATGGACCATCAGGTC
GCATTTTGGATAACAGCACAAAAGACCTCAAGTTTGCATTCACTAACCTAACACCATTTACAATGTAT
GATGTCTATATTGCGGCTGAAACCAGTGCAGGGACTGGGCCCAAGTCAAATATTTCAGTATTCACTCC
ACCAGATGTTCCAGGGGCAGTGTTTGATTTACAACTTGCAGAGGTAGAATCCACGCAAGTAAGAATTA
CTTGGAAGAAACCACGACAACCAAATGGAATTATTAACCAATACCGAGTGAAAGTGCTAGTTCCAGAG
ACAGGAATAATTTTGGAAAATACTTTGCTCACTGGAAATAATGAGTATATAAATGACCCCATGGCTCC
AGAAATTGTGAACATAGTAGAGCCAATGGTAGGATTATATGAGGGTTCAGCAGAGATGTCGTCTGACC
TTCACTCACTTGCTACATTTATATATAACAGCCATCCAGATAAAAACTTTCCTGCAAGGAATAGAGCT
GAAGACCAGACTTCACCAGTTGTAACTACAAGGAATCAGTATATTACTGACATTGCAGCTGAACAGCT
GTCTTATGTTATCAGGAGACTTGTACCTTTCACTGAGCACATGATTAGTGTATCTGCTTTCACCATCA
TGGGAGAAGGACCACCAACAGTTCTCAGTGTTAGGACACGTCAGCAAGTGCCAAGCTCCATTAAAATT
ATAAACTATAAAAATATTAGTTCTTCATCTATTTTGTTATATTGGGATCCTCCAGAATATCCCAATGG
AAAAATAACTCACTATACGATTTATGCAATGGAATTGGATACAAACAGAGCATTCCAGATAACTACCA
TAGATAACAGCTTTCTCATAACAGGGTTAAAGAAATACACAAAATACAAAATGAGAGTGGCAGCCTCA
ACCCACGATGGAGAAAGTTCTTTGTCTGAAGAAAATGACATCTTTGTGAGAACTTCAGAAGATGAACC
GGAATCATCACCTCAAGATGTCGAAGTAATTGATGTTACCGCAGATGAAATAAGGTTGAAGTGGTCAC
CACCCGAAAAGCCCAATGGGATCATTATTGCTTATGAAGTGCTATATAAAAATATAGATACTTTATAT
ATGAAGAACACATCAACAACAGACATAATATTAAGGAACTTAAGACCTCACACCCTCTATAACATTTC
TGTAAGGTCTTACACCAGATTTGGTCATGGCAATCAGGTATCTTCTTTACTCTCTGTAAGGACTTCGG
AGACTGTGCCTGATAGTGCACCAGAAAATATCACTTACAAAAATATTTCTTCTGGAGAGATTGAGCTA
TCATTCCTTCCCCCAAGTAGTCCCAATGGAATCATAAAAAAATATACAATTTATCTCAAGAGAAGTAA
TGGAAATGAGGAAAGAACTATAAATACAACCTCTTTAACCCAAAACATTAAAGTACTGAAGAAATATA
CCCAATATATCATTGAGGTGTCTGCTAGTACACTGAAAGGTGAAGGAGTTCGGAGTGCTCCCATAAGT
ATACTGACGGAGGAAGATGCTCCTGATTCTCCCCCTCAAGACTTCTCTGTAAAACAGTTGTCTGGTGT
CACGGTGAAGTTGTCATGGCAACCACCCCTGGAGCCAAATGGAATTATCCTTTATTACACAGTTTATG
TCTGGAATAGATCATCATTAAAAACTATTAATGTCACTGAAACATCATTGGAGTTATCAGATTTGGAT
TATAATGTTGAATACAGTGCTTATGTAACAGCTAGCACCAGATTTGGTGATGGGAAAACAAGAAGCAA
TATCATTAGCTTTCAAACACCAGAGGGAGCACCAAGCGATCCTCCCAAAGATGTTTATTATGCAAACC
TCAGTTCTTCATCAATAATTCTTTTCTGGACACCTCCTTCAAAACCTAATGGGATTATACAATATTAC
TCTGTTTATTACAGAAATACTTCAGGTACTTTTATGCAGAATTTTACACTCCATGAAGTAACCAATGA
CTTTGACAATATGACTGTATCCACAATTATAGATAAACTGACAATATTCAGCTACTATACATTTTGGT
TAACAGCAAGTACTTCAGTTGGAAATGGGAATAAAAGCCTCGAG

NOV15c, 258979883 Protein Sequence	SEQ ID NO: 170	740 aa	MW at 82862.3kD

RSPEGPPQNCVTGNITGKSFSILWDPPTIVTGKFSYRVELYGPSGRILDNSTKDLKFAFTNLTPFTMY
DVYIAAETSAGTGPKSNISVFTPPDVPGAVFDLQLAEVESTQVRITWKKPRQPNGIINQYRVKVLVPE
TGIILENTLLTGNNEYINDPMAPEIVNIVEPMVGLYEGSAEMSSDLHSLATFIYNSHPDKNFPARNRA
EDQTSPVVTTRNQYITDIAAEQLSYVIRRLVPFTEHMISVSAFTIMGEGPPTVLSVRTRQQVPSSIKI
INYKNISSSSILLYWDPPEYPNGKITHYTIYAMELDTNRAFQITTIDNSFLITGLKKYTKYKMRVAAS
THDGESSLSEENDIFVRTSEDEPESSPQDVEVIDVTADEIRLKWSPPEKPNGIIIAYEVLYKNIDTLY
MKNTSTTDIILRNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSAPENITYKNISSGEIEL
SFLPPSSPNGIIKKYTIYLKRSNGNEERTINTTSLTQNIKVLKKYTQYIIEVSASTLKGEGVRSAPIS
ILTEEDAPDSPPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVYVWNRSSLKTINVTETSLELSDLD
YNVEYSAYVTASTRFGDGKTRSNIISFQTPEGAPSDPPKDVYYANLSSSSIILFWTPPSKPNGIIQYY
SVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKLTIFSYYTFWLTASTSVGNGNKSLE

NOV15d, CG50718-01 DNA Sequence	SEQ ID NO: 171	6994 bp
	ORF Start: ATG at 31	ORF Stop: TAA at 6874

TGATTCTACTGGCTGAAAAATGTAATAAAGATGGATTTTCTTATCATTTTTCTTTTACTTTTTATTGG

GACTTCAGAGACACAGGTAGATGTTTCCAATGTCGTTCCTGGTACTAGGTACGATATAACCATCTCTT
CAATTTCTACAACATACACCTCACCTGTTACTAGAATAGGGGCTTCTAATGAACCAGGGCCTCCAGTC
TTCCTAGCCGGGGAAAGAGTCGGATCTGCTGGGATTCTTCTGTCTTGGAATACACCACCTAATCCAAA
TGGAAGGATTATATCTTACATTGTCAAATATAAGGAAGTTTGTCCGTGGATGCAAACAGTATATACAC
AAGTCAGATCAAAGCCAGACAGTCTGGAAGTTCTTCTTACTAATCTTAATCCTGGAACAACATATGAA
ATTAAGGTAGCTGCTGAAAACAGTGCTGGCATTGGAGTGTTTAGTGATCCATTTCTCTTCCAAACTGC
AGAAAGTGCTCCAGGAAAAGTGGTGGATTTCACAGGTGAGGCTGTCCCGTTCAGCAGTAAGCTGATGT
GGTATACCTCGGCAACCAAAAAAAAAATTACCAGCTTCAAGATTAGTGTCAAGCATAACAGAAGTGGG
ATAGTAGTGAAAGAAGTGTCAATCAGAGTGGAGTGCATTTTAAGTGCTTCCCTTCCTTTGCACTGCAA
CGAGAATAGTGAATCTTTTTTATGGAGTACAGCCAGCCCTTCTCCAACCCTTGGTAGAGTTACACCTC
CATCGCGTACCACACATTCATCAAGCACGTTGACACAGAATGAGATCAGCTCTGTGAAAGAGCCTATC
AGTTTTGTAGTGACACACTTGAGACCTTATACAACATATCTTTTTGAAGTTTCAGCTGCTACAACTGA
AGCAGGTTATATTGATAGTACGATTGTCAGAACACCAGAATCAGTGCCTGAAGGACCACCACAAAACT
GCGTAACAGGCAACATCACAGGAAAGTCCTTTTCAATTTTATGGGACCCACCAACTATAGTAACAGGG
AAATTTAGTTATAGAGTTGAATTATATGGACCATCAGCAGGTCGCATTTTGGATAACAGCACAAAAGA
CCTCAAGTTTGCATTCACTAACCTAACACCATTTACAATGTATGATGTCTATATTGCGGCTGAAACCA
GTGCAGGGACTGGGCCCAAGTCAAATATTTCAGTATTCACTCCACCAGATGTTCCAGGGGCAGTGTTT
GATTTACAACTTGCAGAGGTAGAATCCACGCAAGTAAGAATTACTTGGAAGAAACCACGACAACCAAA
TGGAATTATTAACCAATACCGAGTGAAAGTGCTAGTTCCAGAGACAGGAATAATTTTGGAAAATACTT
TGCTCACTGGAAATAATGAGATAAATGACCCCATGGCTCCAGAAATTGTGAACATAGTACAGCCAATG
GTAGGATTATATGAGGGTTCAGCAGAGATGTCGTCTGACCTTCACTCACTTGCTACATTTATATATAA
CAGCCATCCAGATAAAAACTTTCCTGCAAGGAATAGAGCTGAAGACCAGACTTCACCAGTTGTAACTA
CAAGGAATCAGTATATTACTGACATTGCAGCTGAACAGCTGACTTATGTTCTTATCAGATTAAGGAGA
TTTTGGGCTGAGACAATGGGGTTTTCTAGATATACAATCATGTCATCTGCAAGCAGGGACAATTTGAC
TTCCCCAGGCCCTTTGTCAGCCCAAAATTTCAGAGTTACACATGTTACCATAACAGAAGTATTTTTAC
ACTGGGATCCTCCAGATCCTGTATTTTTTCATCATTACCTTATCACTATTTTGGATGTTGAAAACCAA
TCCAAGAGTATTATTTTAAGGACATTAAACAGTTTGTCTCTTGTCCTTATAGGGTTAAAGAAATACAC
AAAATACAAAATGAGAGTGGCAGCCTCAACCCACGTTGGAGAAAGTTCTTTGTCTGAAGAAAATGACA
TCTTTGTGAGAACTTCAGAAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTAATTGATGTTACC
GCAGATGAAATAAGGTTGAAGTGGTCACCACCCGAAAAGCCCAATGGGATCATTATTGCTTATGAAGT
GCTATATAAAAATATAGATACTTTATATATGAAGAACACATCAACAACAGACATAATATTAAGGAACT
TAAGACCTCACACCCTCTATAACATTTCTGTAAGGTCTTACACCAGATTTGGTCATGGCAATCAGGTA
TCTTCTTTACTCTCTGTAAGGACTTCGGAGTCAGTGCCTGATAGTGCACCAGAAAATATCACTTACAA
AAATATTTCTTCTGGAGAGATTGAGCTATCATTCCTTCCCCCAAGTAGTCCCAATGGAATCATACAAA
AATATACAATTTATCTCAAGAGAAGTAATGGAAATGAGGAAAGAACTATAAATACAACCTCTTTAACC
CAAAACATTAAAGGTCTGAAGAAATATACCCAATATATCATTGAGGTGTCTGCTAGTACACTCAAAGG
TGAAGGAGTTCGGAGTGCTCCCATAAGTATACTGACGGAGGAAGATGCTCCTGATTCTCCCCCTCAAG
ACTTCTCTGTAAAACAGTTGTCTGGTGTCACGGTGAAGTTGTCATGGCAACCACCCCTGGAGCCAAAT
GGAATTATCCTTTATTACACAGTTTATGTCTGGAGATCATCATTAAAAACTATTAATGTCACTGAAAC
ATCATTGGAGTTATCAGATTTGGATTATAATGTTGAATACAGTGCTTATGTAACAGCTAGCACCAGAT
TTGGTGATGGGAAAACAAGAAGCAATATCATTAGCTTTCAAACACCAGAGGGACCAAGCGATCCTCCC
AAAGATGTTTATTATGCAAACCTCAGTTCTTCATCAATAATTCTTTTCTGGACACCTCCTTCAAAACC
TAATGGGATTATACAATATTACTCTGTTTATTACAGAAATACTTCAGGTACTTTTATGCAGAATTTTA
CACTCCATGAAGTAACCAATGACTTTGACAATATGACTGTATCCACAATTATAGATAAACTGACAATA
TTCAGCTACTATACATTTTGGTTAACAGCAAGTACTTCAGTTGGAAATGGGAATAAAAGCAGTGACAT
CATTGAAGTATACACAGATCAAGACGTACCTGAAGGGTTTGTTGGAAACCTGACTTACGAATCCATTT
CGTCAACTGCAATAAATGTAAGCTGGGTCCCACCGGCTCAACCAAACGGTCTAGTCTTCTACTATGTT
TCACTGATCTTACAGCAGACTCCTCGCCATGTGAGACCACCTCTTGTTACATATGAGAGAAGCATATA

TTTTGATAATCTGGAAAAATACACTGATTATATATTAAAAATTACTCCATCAACAGAAAAGGGATTCT
CTGATACCTATACTGCCCAGCTATACATCAAGACTGAAGAAGATATCCCAGAAACTTCACCAATAATC
AACACTTTTAAAAACCTTTCCTCTACCTCAGTTCTCTTATCATGGGATCCCCCAGTAAAGCCAAATGG
TGCAATAATAAGTTATGATTTAACTTTACAAGGACCAAATGAAAATTATTCTTTCATTACTTCTGATA
ATTACATAATATTGGAAGAGCTTTCACCATTTACATTATATAGCTTTTTTGCTGCCGCAAGAACTAGA
AAAGGACTTGGTCCTTCCAGTATTCTTTTCTTTTACACAGATGAGTCAGTGCCGTTAGCACCTCCACA
AAATTTGACTTTAATCAACTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCCTCTTCCAG
GTGGTATTGTTAAAGTATATAGTTTTAAAATTCATGAACATGAAACTGACACTATATATTATAAGAAT
ATATCAGGATTTAAAACTGAAGCCAAACTTGTTGGACTGGAACCAGTCAGCACCTACTCTATCCGTGT
ATCTGCGTTCACCAAAGTTGGAAATGGCAATCAATTTAGTAATGTAGTAAAATTCACAACCCAAGAAT
CAGTTCCAGATGTCGTGCAGAATATGCAGTGCATGGCAACTAGCTGGCAGTCAGTTTTAGTGAAATGG

GATCCACCCAAAAAGGCAAATGGAATAATAACGCAGTATATGGTAACAGTTGAAAGGAATTCTACAAA
AGTTTCTCCCCAAGATCACATGTACACTTTCATAAAGCTTCTTGCCAATACCTCATATGTCTTTAAAG
TAAGAGCTTCAACCTCAGCTGGTGAAGGTGATGAAAGCACATGCCATGTCAGCACACTACCTGAAACA
GTTCCCAGTGTTCCCACAAATATTGCTTTTTCTGATGTTCAGTCAACTAGTGCAACATTGACATGGAT
AAGACCTGACACTATCCTTGGCTACTTTCAAAATTACAAAATTACCACTCAACTTCGTGCTCAAAAAT
GCAAAGAATGGGAATCCGAAGAATGTGTTGAATATCAAAAAATTCAATACCTCTATGAAGCTCACTTA
ACTGAAGAGACAGTATATGGATTAAAGAAATTTAGATGGTATAGATTCCAAGTGGCTGCCAGCACCAA
TGCTGGCTATGGCAATGCTTCAAACTGGATTTCTACAAAAACTCTGCCTGGCCCTCCAGATGGTCCTC
CTGAAAATGTTCATGTAGTAGCAACATCACCTTTTAGCATCAGCATAAGCTGGAGTGAACCTGCTGTC
ATTACTGGACCAACATGTTATCTGATTGATGTCAAATCGGTAGATAATGATGAATTTAATATATCCTT
CATCAAGTCAAATGAAGAAAATAAAACCATAGAAATTAAAGATTTAGAAATATTCACAAGGTATTCTG
TAGTGATCACTGCATTTACTGGGAACATTAGTGCTGCATATGTAGAAGGGAAGTCAAGTGCTGAAATG
ATTGTTACTACTTTAGAATCAGCCCCAAAGGACCCACCTAACAACATGACATTTCAGAAGATACCAGA
TGAAGTTACAAAATTTCAATTAACGTTCCTTCCTCCTTCTCAACCTAATGGAAATATCCAAGTATATC
AAGCTCTGGTTTACCGAGAAGATGATCCTACTGCTGTCCAGATTCACAACCTCAGTATTATACAGAAA
ACCAACACATTCGTCATTGCAATGCTAGAAGGACTAAAAGGTGGACATACATACAATATCAGTGTTTA
CGCAGTCAATAGTGCTGGTGCAGGTCCAAAGGTTCCGATGAGAATAACCATGGATATCAAAGCTCCAG
CACGACCAAAAACCAAACCAACCCCTATTTATGATGCCACAGGAAAACTGCTTGTGACTTCAACAACA
ATTACAATCAGAATGCCAATATGTTACTACAGTGATGATCATGGACCAATAAAAAATGTACAAGTGCT
TGTGACAGAAACAGGAGCTCAGCATGATGGAAATGTAACAAAGTGGTATGATGCATATTTTAATAAAG
CAAGGCCATATTTTACAAATGAAGGCTTTCCTAACCCTCCATGTACAGAAGGAAAGACAAAGTTTAGT
GGCAATGAAGAAATCTACATCATAGGTGCTGATAATGCATGCATGATTCCTGGCAATGAAGACAAAAT
TTGCAATGGACCACTGAAACCAAAAAAGCAATACTTATTTAAATTTAGAGCTACAAATATTATGGGAC
AATTTACTGACTCTGATTATTCTGACCCTGTTAAGACTTTAGGCGAAGGACTTTCAGAAAGAACCGTA
GAGATCATTCTTTCCGTCACTTTGTGTATCCTTTCAATAATTCTCCTTGGAACAGCTATTTTTGCATT
TGCAAGAATTCGACAGAAGCAGAAAGAAGGTGGCACATACTCTCCTCAGGATGCAGAAATTATTGACA
CTAAATTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTGGAACTGAAGGACGAGAGATTAACGCGG
CCAATAAGCAAGAAATCCTTCCTGCAACATGTTGAAGAGCTTTGCACAAACAACAACCTAAAGTTTCA
AGAAGAATTTTCGGAATTACCAAAATTTCTTCAGGATCTTTCTTCAACTGATGCTGATCTGCCTTGGA
ATAGAGCAAAAAACCGCTTCCCAAACATAAAACCATATAATAATAACAGAGTAAAGCTGATAGCTGAC
GCTAGTGTTCCAGGTTCGGATTATATTAATGCCAGCTATATTTCTGGTTATTTATGTCCAAATGAATT
TATTGCTACTCAAGGTCCACTACCAGGAACAGTTGGAGATTTTTGGAGAATGGTGTGGGAAACCAGAG
CAAAAACATTAGTAATGCTAACACAGTGTTTTGAAAAAGGACGGATCAGATGCCATCAGTATTGGCCA
GAGGACAACAAGCCAGTTACTGTCTTTGGAGATATAGTGATTACAAAGCTAATGGAGGATGTTCAAAT
AGATTGGACTATCAGGGATCTGAAAATTGAAAGGCATGGGGATTGCATGACTGTTCGACAGTGTAACT
TTACTGCCTGGCCAGAGCATGGGGTTCCTGAGAACAGCGCCCCTCTAATTCACTTTGTGAAGTTGGTT
CGAGCAAGCAGGGCACATGACACCACACCTATGATTGTTCACTGCAGTGCTGGAGTTGGAAGAACTGG
AGTTTTTATTGCTCTGGACCATTTAACACAACATATAAATGACCATGATTTTGTGGATATATATGGAC

TAGTAGCTGAACTGAGAAGTGAAAGAATGTGCATGGTGCAGAATCTGGCACAGTATATCTTTTTACAC
CAGTGCATTCTGGATCTCTTATCAAATAAGGGAAGTAATCAGCCCATCTGTTTTGTTAACTATTCAGC
ACTTCAGAAGATGGACTCTTTGGACGCCATGGAAGGTGGTGATGTTGAGCTTGAATGGGAAGAAACCA
CTATGTAAATATTCAGACCAAAGGATACAATTGGAAGAGATTTTTAAATCCCAGGGGCCAAAGTTACC
CCCTCATTCTTCCGAATTGAAATGTGCAACCTTAAAGAAATATCTATGCTTCTCTCAC

NOV15d, CG50718-01 Protein Sequence	SEQ ID NO: 172	2281 aa	MW at 255030.3kD

MDFLIIFLLLFIGTSETQVDVSNVVPGTRYDITISSISTTYTSPVTRIGASNEPGPPVFLAGERVGSA
GILLSWNTPPNPNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLNPGTTYEIKVAAENSAG
IGVFSDPFLFQTAESAPGKVVDFTGEAVPFSSKLMWYTSATKKKITSFKISVKHNRSGIVVKEVSIRV
ECILSASLPLHCNENSESFLWSTASPSPTLGRVTPPSRTTHSSSTLTQNEISSVKEPISFVVTHLRPY
TTYLFEVSAATTEAGYIDSTIVRTPESVPEGPPQNCVTGNITGKSFSILWDPPTIVTGKFSYRVELYG
PSAGRILDNSTKDLKFAFTNLTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVPGAVFDLQLAEVEST
QVRITWKKPRQPNGIINQYRVKVLVPETGIILENTLLTGNNEINDPMAPEIVNIVQPMVGLYEGSAEM
SSDLHSLATFIYNSHPDKNFPARNRAEDQTSPVVTTRNQYITDIAAEQLTYVLIRLRRFWAETMGFSR
YTIMSSASRDNLTSPGPLSAQNFRVTHVTITEVFLHWDPPDPVFFHHYLITILDVENQSKSIILRTLN
SLSLVLIGLKKYTKYKMRVAASTHVGESSLSEENDIFVRTSEDEPESSPQDVEVIDVTADEIRLKWSP
PEKPNGIIIAYEVLYKNIDTLYMKNTSTTDIILRNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSE
SVPDSAPENITYKNISSGEIELSFLPPSSPNGIIQKYTIYLKRSNGNEERTINTTSLTQNIKGLKKYT
QYIIEVSASTLKGEGVRSAPISILTEEDAPDSPPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVYV

WRSSLKTINVTETSLELSDLDYNVEYSAYVTASTRFGDGKTRSNIISFQTPEGPSDPPKDVYYANLSS
SSIILFWTPPSKPNGIIQYYSVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKLTIFSYYTFWLTA
STSVGNGNKSSDIIEVYTDQDVPEGFVGNLTYESISSTAINVSWVPPAQPNGLVFYYVSLILQQTPRH
VRPPLVTYERSIYFDNLEKYTDYILKITPSTEKGFSDTYTAQLYIKTEEDIPETSPIINTFKNLSSTS
VLLSWDPPVKPNGAIISYDLTLQGPNENYSFITSDNYIILEELSPFTLYSFFAAARTRKGLGPSSILF
FYTDESVPLAPPQNLTLINCTSDFVWLKWSPSPLPGGIVKVYSFKIHEHETDTIYYKNISGFKTEAKL
VGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFTTQESVPDVVQNMQCMATSWQSVLVKWDPPKKANGII
TQYMVTVERNSTKVSPQDHMYTFIKLLANTSYVFKVRASTSAGEGDESTCHVSTLPETVPSVPTNIAF
SDVQSTSATLTWIRPDTILGYFQNYKITTQLRAQKCKEWESEECVEYQKIQYLYEAHLTEETVYGLKK
FRWYRFQVAASTNAGYGNASNWISTKTLPGPPDGPPENVHVVATSPFSISISWSEPAVITGPTCYLID
VKSVDNDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAFTGNISAAYVEGKSSAEMIVTTLESAPK
DPPNNMTFQKIPDEVTKFQLTFLPPSQPNGNIQVYQALVYREDDPTAVQIHNLSIIQKTNTFVIAMLE
GLKGGHTYNISVYAVNSAGAGPKVPMRITMDIKAPARPKTKPTPIYDATGKLLVTSTTITIRMPICYY
SDDHGPIKNVQVLVTETGAQHDGNVTKWYDAYFNKARPYFTNEGFPNPPCTEGKTKFSGNEEIYIIGA
DNACMIPGNEDKICNGPLKPKKQYLFKFRATNIMGQFTDSDYSDPVKTLGEGLSERTVEIILSVTLCI
LSIILLGTAIFAFARIRQKQKEGGTYSPQDAEIIDTKLKLDQLITVADLELKDERLTRPISKKSFLQH
VEELCTNNNLKFQEEFSELPKFLQDLSSTDADLPWNRAKNRFPNIKPYNNNRVKLIADASVPGSDYIN
ASYISGYLCPNEFIATQGPLPGTVGDFWRMVWETRAKTLVMLTQCFEKGRIRCHQYWPEDNKPVTVFG
DIVITKLMEDVQIDWTIRDLKIERHGDCMTVRQCNFTAWPEHGVPENSAPLIHFVKLVRASRAHDTTP
MIVHCSAGVGRTGVFIALDHLTQHINDHDFVDIYGLVAELRSERMCMVQNLAQYIFLHQCILDLLSNK
GSNQPICFVNYSALQKMDSLDAMEGGDVELEWEETTM

NOV15e, CG50718-03 DNA Sequence	SEQ ID NO: 173	2739 bp
	ORF Start: at 7	ORF Stop: at 2734

AGATCTCAGGTAGATGTTTCCAATGTCGTTCCTGGTACTAGGTACGATATAACCATCTCTTCAATTTC
TACAACATACACCTCACCTGTTACTAGAATAGTGACAACAAATGTAACAAAACCAGGGCCTCCAGTCT
TCCTAGCCGGGGAAAGAGTCGGATCTGCTGGGATTCTTCTGTCTTGGAATACACCACCTAATCCAAAT
GGAAGGATTATATCTTACATTGTCAAATATAAGGAAGTTTGTCCGTGGATGCAAACAGTATATACACA
AGTCAGATCAAAGCCAGACAGTCTGGAAGTTCTTCTTACTAATCTTAATCCTGGAACAACATATGAAA
TTAAGGTTGCTGCTGAAAACAGTGCTGGCATTGGAGTGTTTAGTGATCCATTTCTCTTCCAAACTGCA
GAAAGTGCTCCAGGAAAAGTGGTGAATCTCACAGTTGAGGCCTACAACGCTTCAGCAGTTAAGCTGAT
TTGGTATTTACCTCGGCAACCAAATGGCAAAATTACCAGCTTCAAGATTAGTGTCAAGCATGCCAGAA
GTGGGATAGTAGTGAAAGATGTCTCAATCAGAGTAGAGGACATTTTGACTGGGAAATTGCCAGAATGC
AATGAGAATAGTGAATCTTTTTTATGGAGTACAGCCAGCCCTTCTCCAACCCTTGGTAGAGTTACACC
TCCATCGCGTACCACACATTCATCAAGCACGTTGACACAGAATGAGATCAGCTCTGTGTGGAAAGAGC
CTATCAGTTTTGTAGTGACACACTTGAGACCTTATACAACATATCTTTTTGAAGTTTCAGCTGTTACA
ACTGAAGCAGGTTATATTGATAGTACGATTGTCAGAACACCAGAATCAGTGCCTGAAGGACCACCACA
AAACTGCGTAACAGGCAACATCACAGGAAAGTCCTTTTCAATTTTATGGGACCCACCAACTATAGTAA
CAGGGAAATTTAGTTATAGAGTTGAATTATATGGACCATCAGGTCGCATTTTGGATAACAGCACAAAA
GACCTCAAGTTTGCATTCACTAACCTAACACCATTTACAATGTATGATGTCTATATTGCGGCTGAAAC
CAGTGCAGGGACTGGGCCCAAGTCAAATATTTCAGTATTCACTCCACCAGATGTTCCAGGGGCAGTGT
TTGATTTACAACTTGCAGAGGTAGAATCCACGCAAGTAAGAATTACTTGGAAGAAACCACGACAACCA
AATGGAATTATTAACCAATACCGAGTGAAAGTGCTAGTTCCAGAGACAGGAATAATTTTGGAAAATAC
TTTGCTCACTGGAAATAATGAGTATATAAATGACCCCATGGCTCCAGAAATTGTGAACATAGTAGAGC
CAATGGTAGGATTATATGAGGGTTCAGCAGAGATGTCGTCTGACCTTCACTCACTTGCTACATTTATA
TATAACAGCCATCCAGATAAAAACTTTCCTGCAAGGAATAGAGCTGAAGACCAGACTTCACCAGTTGT
AACTACAAGGAATCAGTATATTACTGACATTGCAGCTGAACAGCTGTCTTATGTTATCAGGAGACTTG
TACCTTTCACTGAGCACATGATTAGTGTATCTGCTTTCACCATCATGGGAGAAGGACCACCAACAGTT
CTCAGTGTTAGGACACGTCAGCAAGTGCCAAGCTCCATTAAAATTATAAACTATAAAAATATTAGTTC
TTCATCTATTTTGTTATATTGGGATCCTCCAGAATATCCCAATGGAAAAATAACTCACTATACGATTT
ATGCAATGGAATTGGATACAAACAGAGCATTCCAGATAACTACCATAGATAACAGCTTTCTCATAACA
GGGTTAAAGAAATACACAAAATACAAAATGAGAGTGGCAGCCTCAACCCACGATGGAGAAAGTTCTTT
GTCTGAAGAAAATGACATCTTTGTGAGAACTTCAGAAGATGAACCGGAATCATCACCTCAAGATGTCG
AAGTAATTGATGTTACCGCAGATGAAATAAGGTTGAAGTGGTCACCACCCGAAAAGCCCAATGGGATC
ATTATTGCTTATGAAGTGCTATATAAAAATATAGATACTTTATATATGAAGAACACATCAACAACAGA
CATAATATTAAGGAACTTAAGACCTCACACCCTCTATAACATTTCTGTAAGGTCTTACACCAGATTTG
GTCATGGCAATCAGGTATCTTCTTTACTCTCTGTAAGGACTTCGGAGACTGTGCCTGATAGTGCACCA
GAAAATATCACTTACAAAAATATTTCTTCTGGAGAGATTGAGCTATCATTCCTTCCCCCAAGTAGTCC
CAATGGAATCATACAAAAATATACAATTTATCTCAAGAGAAGTAATGGAAATGAGGAAAGAACTATAA

ATACAACCTCTTTAACCCAAAACATTAAAGGACTGAAGAAATATACCCAATATATCATTGAGGTGTCT
GCTAGTACACTCAAAGGTGAAGGAGTTCGGAGTGCTCCCATAAGTATACTGACGGAGGAAGATGCTCC
TGATTCTCCCCCTCAAGACTTCTCTGTAAAACAGTTGTCTGGTGTCACGGTGAAGTTGTCATGGCAAC
CACCCCTGGAGCCAAATGGAATTATCCTTTATTACACAGTTTATGTCTGGAATAGATCATCATTAAAA
ACTATTAATGTCACTGAAACATCATTGGAGTTATCAGATTTGGATTATAATGTTGAATACAGTGCTTA
TGTAACAGCTAGCCTCGAG

NOV15e, CG50718-03 Protein Sequence	SEQ ID NO: 174	909 aa	MW at 100812.8kD

QVDVSNVVPGTRYDITISSISTTYTSPVTRIVTTNVTKPGPPVFLAGERVGSAGILLSWNTPPNPNGR
IISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLNPGTTYEIKVAAENSAGIGVFSDPFLFQTAES
APGKVVNLTVEAYNASAVKLIWYLPRQPNGKITSFKISVKHARSGIVVKDVSIRVEDILTGKLPECNE
NSESFLWSTASPSPTLGRVTPPSRTTHSSSTLTQNEISSVWKEPISFVVTHLRPYTTYLFEVSAVTTE
AGYIDSTIVRTPESVPEGPPQNCVTGNITGKSFSILWDPPTIVTGKFSYRVELYGPSGRILDNSTKDL
KFAFTNLTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVPGAVFDLQLAEVESTQVRITWKKPRQPNG
IINQYRVKVLVPETGIILENTLLTGNNEYINDPMAPEIVNIVEPMVGLYEGSAEMSSDLHSLATFIYN
SHPDKNFPARNRAEDQTSPVVTTRNQYITDIAAEQLSYVIRRLVPFTEHMISVSAFTIMGEGPPTVLS
VRTRQQVPSSIKIINYKNISSSSILLYWDPPEYPNGKITHYTIYAMELDTNRAFQITTIDNSFLITGL
KKYTKYKMRVAASTHDGESSLSEENDIFVRTSEDEPESSPQDVEVIDVTADEIRLKWSPPEKPNGIII
AYEVLYKNIDTLYMKNTSTTDIILRNLRPHTLYNISVRSYTRFGHGNQVSSLLSVRTSETVPDSAPEN
ITYKNISSGEIELSFLPPSSPNGIIQKYTIYLKRSNGNEERTINTTSLTQNIKGLKKYTQYIIEVSAS
TLKGEGVRSAPISILTEEDAPDSPPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVYVWNRSSLKTI
NVTETSLELSDLDYNVEYSAYVTAS

NOV15f, CG50718-04 DNA Sequence	SEQ ID NO: 175	2565 bp
	ORF Start: at 7	ORF Stop: at 2560

AGATCTCCTGAAGGGTTTGTTGGAAACCTGACTTACGAATCCATTTCGTCAACTGCAATAAATGTAAG
CTGGGTCCCACCGGCTCAACCAAACGGTCTAGTCTTCTACTATGTTTCACTGATCTTACAGCAGACTC
CTCGCCATGTGAGACCACCTCTTGTTACATATGAGAGAAGCATATATTTTGATAATCTGGAAAAATAC
ACTGATTATATATTAAAAATTACTCCATCAACAGAAAAGGGATTCTCTGATACCTATACTGCCCAGCT
ATACATCAAGACTGAAGAAGATGTCCCAGAAACTTCACCAATAATCAACACTTTTAAAAACCTTTCCT
CTACCTCAGTTCTCTTATCATGGGATCCCCCAGTAAAGCCAAATGGTGCAATAATAAGTTATGATTTA
ACTTTACAAGGACCAAATGAAAATTATTCTTTCATTACTTCTGATAATTACATAATATTGGAAGAGCT
TTCACCATTTACATTATATAGCTTTTTTGCTGCCGCAAGAACTAGAAAAGGACTTGGTCCTTCCAGTA
TTCTTTTCTTTTACACAGATGAGTCAGTGCCGTTAGCACCTCCACAAAATTTGACTTTAATCAACTGT
ACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCCTCTTCCAGGTGGTATTGTTAAAGTATATAG
TTTTAAAATTCATGAACATGAAACTGACACTATATATTATAAGAATATATCAGGATTTAAAACTGAAG
CCAAACTTGTTGGACTGGAACCAGTCAGCACCTACTCTATCCGTGTATCTGCGTTCACCAAAGTTGGA
AATGGCAATCAATTTAGTAATGTAGTAAAATTCACAACCCAAGAATCAGTTCCAGATGTCGTGCAGAA
TATGCAGTGCATGGCAACTAGCTGGCAGTCAGTTTTAGTGAAATGGGATCCACCCAAAAAGGCAAATG
GAATAATAACGCAGTATATGGTAACAGTTGAAAGGAATTCTACAAAAGTTTCTCCCCAAGATCACATG
TACACTTTCATAAAGCTTCTTGCCAATACCTCATATGTCTTTAAAGTAAGAGCTTCAACCTCAGCTGG
TGAAGGTGATGAAAGCACATGCCATGTCAGCACACTACCTGAAACAGTTCCCAGTGTTCCCACAAATA
TTGCTTTTTCTGATGTTCAGTCAACTAGTGCAACATTGACATGGATAAGACCTGACACTATCCTTGGC
TACTTTCAAAATTACAAAATTACCACTCAACTTCGTGCTCAAAAATGCAAAGAATGGGAATCCGAAGA
ATGTGTTGAATATCAAAAAATTCAATACCTCTATGAAGCTCACTTAACTGAAGAGACAGTATATGGAT
TAAAGAAATTTAGATGGTATAGATTCCAAGTGGCTGCCAGCACCAATGCTGGCTATGGCAATGCTTCA
AACTGGATTTCTACAAAAACTCTGCCTGGCCCTCCAGATGGTCCTCCTGAAAATGTTCATGTAGTAGC
AACATCACCTTTTAGCATCAGCATAAGCTGGAGTGAACCTGCTGTCATTACTGGACCAACATGTTATC
TGATTGATGTCAAATCGGTAGATAATGATGAATTTAATATATCCTTCATCAAGTCAAATGAAGAAAAT
AAAACCATAGAAATTAAAGATTTAGAAATATTCACAAGGTATTCTGTAGTGATCACTGCATTTACTGG
GAACATTAGTGCTGCATATGTAGAAGGGAAGTCAAGTGCTGAAATGATTGTTACTACTTTAGAATCAG
CCCCAAAGGACCCACCTAACAACATGACATTTCAGAAGATACCAGATGAAGTTACAAAATTTCAATTA
ACGTCCCTTCCTCCTTCTCAACCTAATGGAAATATCCAAGTATATCAAGCTCTGGTTTACCGAGAAGA
TGATCCTACTGCTGTCCAGATTCACAACCTCAGTATTATACAGAAAACCAACACATTCGTCATTGCAA
TGCTAGAAGGACTAAAAGGTGGACATACATACAATATCAGTGTTTACGCAGTCAATAGTGCTGGTGCA
GGTCCAAAGGTTCCGATGAGAATAACCATGGATATCAAAGCTCCAGCACGACCAAAAACCAAACCAAC
CCCTATTTATGATGCCACAGGAAAACTGCTTGTGACTTCAACAACAATTACAATCAGAATGCCAATAT
GTTACTACAGTGATGATCATGGACCAATAAAAAATGTACAAGTGCTTGTGACAGAAACAGGAGCTCAG

CATGATGGAAATGTAACAAAGTGGTATGATGCATATTTTAATAAAGCAAGGCCATATTTTACAAATGA
AGGCTTTCCTAACCCTCCATGTACAGAAGGAAAGACAAAGTTTAGTGGCAATGAAGAAATCTACATCA
TAGGTGCTGATAATGCATGCATGATTCCTGGCAATGAAGACAAAATTTGCAATGGACCACTGAAACCA
AAAAAGCAATACTTATTTAAATTTAGAGCTACAAATATTATGGGACAATTTACTGACTCTGATTATTC
TGACCCTGTTAAGACTTTAGGCGAAGGACTTTCAGAAAGAACCCTCGAG

NOV15f, CG50718-04 Protein Sequence	SEQ ID NO: 176	851 aa	MW at 95001.6kD

PEGFVGNLTYESISSTAINVSWVPPAQPNGLVFYYVSLILQQTPRHVRPPLVTYERSIYFDNLEKYTD
YILKITPSTEKGFSDTYTAQLYIKTEEDVPETSPIINTFKNLSSTSVLLSWDPPVKPNGAIISYDLTL
QGPNENYSFITSDNYIILEELSPFTLYSFFAAARTRKGLGPSSILFFYTDESVPLAPPQNLTLINCTS
DFVWLKWSPSPLPGGIVKVYSFKIHEHETDTIYYKNISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNG
NQFSNVVKFTTQESVPDVVQNMQCMATSWQSVLVKWDPPKKANGIITQYMVTVERNSTKVSPQDHMYT
FIKLLANTSYVFKVRASTSAGEGDESTCHVSTLPETVPSVPTNIAFSDVQSTSATLTWIRPDTILGYF
QNYKITTQLRAQKCKEWESEECVEYQKIQYLYEAHLTEETVYGLKKFRWYRFQVAASTNAGYGNASNW
ISTKTLPGPPDGPPENVHVVATSPFSISISWSEPAVITGPTCYLIDVKSVDNDEFNISFIKSNEENKT
IEIKDLEIFTRYSVVITAFTGNISAAYVEGKSSAEMIVTTLESAPKDPPNNMTFQKIPDEVTKFQLTS
LPPSQPNGNIQVYQALVYREDDPTAVQIHNLSIIQKTNTFVIAMLEGLKGGHTYNISVYAVNSAGAGP
KVPMRITMDIKAPARPKTKPTPIYDATGKLLVTSTTITIRMPICYYSDDHGPIKNVQVLVTETGAQHD
GNVTKWYDAYFNKARPYFTNEGFPNPPCTEGKTKFSGNEEIYIIGADNACMIPGNEDKICNGPLKPKK
QYLFKFRATNIMGQFTDSDYSDPVKTLGEGLSERT

NOV15g, CG50718-05 DNA Sequence	SEQ ID NO: 177	6903 bp
	ORF Start: ATG at 1	ORF Stop: TAA at 6901

ATGGATTTTCTTATCATTTTTCTTTTACTTTTTATTGGGACTTCAGAGACACAGGTAGATGTTTCCAA
TGTCGTTCCTGGTACTAGGTACGATATAACCATCTCTTCAATTTCTACAACATACACCTCACCTGTTA
CTAGAATAGTGACAACAAATGTAACAGAACCAGGGCCTCCAGTCTTCCTAGCCGGGGAAAGAGTCGGA
TCTGCTGGGATTCTTCTGTCTTGGAATACACCACCTAATCCAAATGGAAGGATTATATCTTACATTGT
CAAATATAAGGAAGTTTGTCCGTGGATGCAAACAGTATATACACAAGTCAGATCAAAGCCAGACAGTC
TGGAAGTTCTTCTTACTAATCTTAATCCTGGAACAACATATGAAATTAAGGTAGCTGCTGAAAACAGT
GCTGGCATTGGAGTGTTTAGTGATCCATTTCTCTTCCAAACTGCAGAAAGTCCAGCTCCAGGAAAAGT
GGTGAATCTCACAGTTGAGGCCTACAACGCTTCAGCAGTTAAGCTGATTTGGTATTTACCTCGGCAAC
CAAATGGCAAAATTACCAGCTTCAAGATTAGTGTCAAGCATGCCAGAAGTGGGATAGTAGTGAAAGAT
GTCTCAATCAGAGTAGAGGACATTTTGACTGGGAAATTGCCAGAATGCAATGTAGAGAATAGTGAATC
TTTTTTATGGAGTACAGCCAGCCCTTCTCCAACCCTTGGTAGAGTTACACCTCCATCGCGTACCACAC
ATTCATCAAGCACGTTGACACAGAATGAGATCAGCTCTGTGTGGAAAGAGCCTATCAGTTTTGTAGTG
ACACACTTGAGACCTTATACAACATATCTTTTTGAAGTTTCAGCTGCTACAACTGAAGCAGGTTATAT
TGATAGTACGATTGTCAGAACACCAGAATCAGTGCCTGAAGGACCACCACAAAACTGCGTAACAGGCA
ACATCACAGGAAAGTCCTTTTCAATTTTATGGGACCCACCAACTATAGTAACAGGGAAATTTAGTTAT
AGAGTTGAATTATATGGACCATCAGGTCGCATTTTGGATAACAGCACAAAAGACCTCAAGTTTGCATT
CACTAACCTAACACCATTTACAATGTATGATGTCTATATTGCGGCTGAAACCAGTGCAGGGACTGGGC
CCAAGTCAAATATTTCAGTATTCACTCCACCAGATGTTCCAGGGGCAGTGTTTGATTTACAACTTGCA
GAGGTAGAATCCACGCAAGTAAGAATTACTTGGAAGAAACCACGACAACCAAATGGAATTATTAACCA
ATACCGAGTGAAAGTGCTAGTTCCAGAGACAGGAATAATTTTGGAAAATACTTTGCTCACTGGAAATA
ATGAGATAAATGACCCCATGGCTCCAGAAATTGTGAACATAGTAGAGCCAATGGTAGGATTATATGAG
GGTTCAGCAGAGATGTCGTCTGACCTTCACTCACTTGCTACATTTATATATAACAGCCATCCAGATAA
AAACTTTCCTGCAAGGAATAGAGCTGAAGACCAGACTTCACCAGTTGTAACTACAAGGAATCAGTATA
TTACTGACATTGCAGCTGAACAGCTGTCTTATGTTATCAGGAGACTTGTACCTTTCACTGAGCACATG
ATTAGTGTATCTGCTTTCACCATCATGGGAGAAGGACCACCAACAGTTCTCAGTGTTAGGACACGTCA
GCAAGTGCCAAGCTCCATTAAAATTATAAACTATAAAAATATTAGTTCTTCATCTATTTTGTTATATT
GGGATCCTCCAGAATATCCCAATGGAAAAATAACTCACTATACGATTTATGCAATGGAATTGGATACA
AACAGAGCATTCCAGATAACTACCATAGATAACAGCTTTCTCATAACAGGTATAGGGTTAAAGAAATA
CACAAAATACAAAATGAGAGTGGCAGCCTCAACCCACGTTGGAGAAAGTTCTTTGTCTGAAGAAAATG
ACATCTTTGTGAGAACTTCAGAAGATGAACCGGAATCATCACCTCAAGATGTCGAAGTAATTGATGTT
ACCGCAGATGAAATAAGGTTGAAGTGGTCACCACCCGAAAAGCCCAATGGGATCATTATTGCTTATGA
AGTGCTATATAAAAATATAGATACTTTATATATGAAGAACACATCAACAACAGACATAATATTAAGGA
ACTTAAGACCTCACACCCTCTATAACATTTCTGTAAGGTCTTACACCAGATTTGGTCATGGCAATCAG
GTATCTTCTTTACTCTCTGTAAGGACTTCGGAGACTGTGCCTGATAGTGCACCAGAAAATATCACTTA
CAAAAATATTTCTTCTGGAGAGATTGAGCTATCATTCCTTCCCCCAAGTAGTCCCAATGGAATCATAC

AAAAATATACAATTTATCTCAAGAGAAGTAATGGAAATGAGGAAAGAACTATAAATACAACCTCTTTA
ACCCAAAACATTCTGAAGAAATATACCCAATATATCATTGAGGTGTCTGCTAGTACACTCAAAGGTGA
AGGAGTTCGGAGTGCTCCCATAAGTATACTGACGGAGGAAGATGCTCCTGATTCTCCCCCTCAAGACT
TCTCTGTAAAACAGTTGTCTGGTGTCACGGTGAAGTTGTCATGGCAACCACCCCTGGAGCCAAATGGA
ATTATCCTTTATTACACAGTTTATGTCTGGAGGAATAGATCATCATTAAAAACTATTAATGTCACTGA
AACATCATTGGAGTTATCAGATTTGGATTATAATGTTGAATACAGTGCTTATGTAACAGCTAGCACCA
GATTTGGTGATGGGAAAACAAGAAGCAATATCATTAGCTTTCAAACACCAGAGGGACCAAGCGATCCT
CCCAAAGATGTTTATTATGCAAACCTCAGTTCTTCATCAATAATTCTTTTCTGGACACCTCCTTCAAA
ACCTAATGGGATTATACAATATTACTCTGTTTATTACAGAAATACTTCAGGTACTTTTATGCAGAATT
TTACACTCCATGAAGTAACCAATGACTTTGACAATATGACTGTATCCACAATTATAGATAAACTGACA
ATATTCAGCTACTATACATTTTGGTTAACAGCAAGTACTTCAGTTGGAAATGGGAATAAAAGCAGTGA
CATCATTGAAGTATACACAGATCAAGACGTCCCTGAAGGGTTTGTTGGAAACCTGACTTACGAATCCA
TTTCGTCAACTGCAATAAATGTAAGCTGGGTCCCACCGGCTCAACCAAACGGTCTAGTCTTCTACTAT
GTTTCACTGATCTTACAGCAGACTCCTCGCCATGTGAGACCACCTCTTGTTACATATGAGAGAAGCAT

ATATTTTGATAATCTGGAAAAATACACTGATTATATATTAAAAATTACTCCATCAACAGAAAAGGGAT
TCTCTGATACCTATACTGCCCAGCTATACATCAAGACTGAAGAAGATGTCCCAGAAACTTCACCAATA
ATCAACACTTTTAAAAACCTTTCCTCTACCTCAGTTCTCTTATCATGGGATCCCCCAGTAAAGCCAAA
TGGTGCAATAATAAGTTATGATTTAACTTTACAAGGACCAAATGAAAATTATTCTTTCATTACTTCTG
ATAATTACATAATATTGGAAGAGCTTTCACCATTTACATTATATAGCTTTTTTGCTGCCGCAAGAACT
AGAAAAGGACTTGGTCCTTCCAGTATTCTTTTCTTTTACACAGATGAGTCAGTGCCGTTAGCACCTCC
ACAAAATTTGACTTTAATCAACTGTACTTCAGACTTTGTATGGCTGAAATGGAGCCCAAGTCCTCTTC
CAGGTGGTATTGTTAAAGTATATAGTTTTAAAATTCATGAACATGAAACTGACACTATATATTATAAG
AATATATCAGGATTTAAAACTGAAGCCAAACTTGTTGGACTGGAACCAGTCAGCACCTACTCTATCCG
TGTATCTGCGTTCACCAAAGTTGGAAATGGCAATCAATTTAGTAATGTAGTAAAATTCACAACCCAAG
AATCAGTTCCAGATGTCGTGCAGAATATGCAGTGCATGGCAACTAGCTGGCAGTCAGTTTTAGTGAAA
TGGGATCCACCCAAAAAGGCAAATGGAATAATAACGCAGTATATGGTAACAGTTGAAAGGAATTCTAC
AAAAGTTTCTCCCCAAGATCACATGTACACTTTCATAAAGCTTCTTGCCAATACCTCATATGTCTTTA
AAGTAAGAGCTTCAACCTCAGCTGGTGAAGGTGATGAAAGCACATGCCATGTCAGCACACTACCTGAA
ACAGTTCCCAGTGTTCCCACAAATATTGCTTTTTCTGATGTTCAGTCAACTAGTGCAACATTGACATG
GATAAGACCTGACACTATCCTTGGCTACTTTCAAAATTACAAAATTACCACTCAACTTCGTGCTCAAA
AATGCAAAGAATGGGAATCCGAAGAATGTGTTGAATATCAAAAAATTCAATACCTCTATGAAGCTCAC
TTAACTGAAGAGACAGTATATGGATTAAAGAAATTTAGATGGTATAGATTCCAAGTGGCTGCCAGCAC
CAATGCTGGCTATGGCAATGCTTCAAACTGGATTTCTACAAAAACTCTGCCTGGCCCTCCAGATGGTC
CTCCTGAAAATGTTCATGTAGTAGCAACATCACCTTTTAGCATCAGCATAAGCTGGAGTGAACCTGCT
GTCATTACTGGACCAACATGTTATCTGATTGATGTCAAATCGGTAGATAATGATGAATTTAATATATC
CTTCATCAAGTCAAATGAAGAAAATAAAACCATAGAAATTAAAGATTTAGAAATATTCACAAGGTATT
CTGTAGTGATCACTGCATTTACTGGGAACATTAGTGCTGCATATGTAGAAGGGAAGTCAAGTGCTGAA
ATGATTGTTACTACTTTAGAATCAGCCCCAAAGGACCCACCTAACAACATGACATTTCAGAAGATACC
AGATGAAGTTACAAAATTTCAATTAACGTCCCTTCCTCCTTCTCAACCTAATGGAAATATCCAAGTAT
ATCAAGCTCTGGTTTACCGAGAAGATGATCCTACTGCTGTCCAGATTCACAACCTCAGTATTATACAG
AAAACCAACACATTCGTCATTGCAATGCTAGAAGGACTAAAAGGTGGACATACATACAATATCAGTGT
TTACGCAGTCAATAGTGCTGGTGCAGGTCCAAAGGTTCCGATGAGAATAACCATGGATATCAAAGCTC
CAGCACGACCAAAAACCAAACCAACCCCTATTTATGATGCCACAGGAAAACTGCTTGTGACTTCAACA
ACAATTACAATCAGAATGCCAATATGTTACTACAGTGATGATCATGGACCAATAAAAAATGTACAAGT
GCTTGTGACAGAAACAGGAGCTCAGCATGATGGAAATGTAACAAAGTGGTATGATGCATATTTTAATA
AAGCAAGGCCATATTTTACAAATGAAGGCTTTCCTAACCCTCCATGTACAGAAGGAAAGACAAAGTTT
AGTGGCAATGAAGAAATCTACATCATAGGTGCTGATAATGCATGCATGATTCCTGGCAATGAAGACAA
AATTTGCAATGGACCACTGAAACCAAAAAAGCAATACTTATTTAAATTTAGAGCTACAAATATTATGG
GACAATTTACTGACTCTGATTATTCTGACCCTGTTAAGACTTTAGGCGAAGGACTTTCAGAAAGAACC
CTAGAGATCATTCTTTCCGTCACTTTGTGTATCCTTTCAATAATTCTCCTTGGAACAGCTATTTTTGC
ATTTGCAAGAATTCGACAGAAGCAGAAAGAAGGTGGCACATACTCTCCTCAGGATGCAGAAATTATTG
ACACTAAATTGAAGCTGGATCAGCTCATCACAGTGGCAGACCTGGAACTGAAGGACGAGAGATTAACG
CGGTTACTTAGTTATAGAAAATCCATCAAGCCAATAAGCAAGAAATCCTTCCTGCAACATGTTGAAGA
GCTTTGCACAAACAACAACCTAAAGTTTCAAGAAGAATTTTCGGAATTACCAAAATTTCTTCAGGATC
TTTCTTCAACTGATGCTGATCTGCCTTGGAATAGAGCAAAAAACCGCTTCCCAAACATAAAACCATAT
AATAATAACAGAGTAAAGCTGATAGCTGACGCTAGTGTTCCAGGTTCGGATTATATTAATGCCAGCTA
TATTTCTGGTTATTTATGTCCAAATGAATTTATTGCTACTCAAGGTCCACTACCAGGAACAGTTGGAG
ATTTTTGGAGAATGGTGTGGGAAACCAGAGCAAAAACATTAGTAATGCTAACACAGTGTTTTGAAAAA
GGACGGATCAGATGCCATCAGTATTGGCCAGAGGACAACAAGCCAGTTACTGTCTTTGGAGATATAGT

GATTACAAAGCTAATGGAGGATGTTCAAATAGATTGGACTATCAGGGATCTGAAAATTGAAAGGCATG
GGGATTGCATGACTGTTCGACAGTGTAACTTTACTGCCTGGCCAGAGCATGGGGTTCCTGAGAACAGC
GCCCCTCTAATTCACTTTGTGAAGTTGGTTCGAGCAAGCAGGGCACATGACACCACACCTATGATTGT
TCACTGTAGTGCTGGAGTTGGAAGAACTGGAGTTTTTATTGCTCTGGACCATTTAACACAACATATAA

ATGACCATGATTTTGTGGATATATATGGACTAGTAGCTGAACTGAGAAGTGAAAGAATGTGCATGGTG
CAGAATCTGGCACAGTATATCTTTTTACACCAGTGCATTCTGGATCTCTTATCAAATAAGGGAAGTAA
TCAGCCCATCTGTTTTGTTAACTATTCAGCACTTCAGAAGATGGACTCTTTGGACGCCATGGAAGGTG
ATGTTGAGCTTGAATGGGAAGAAACCACTATGTAA

NOV15g, CG50718-05 Protein Sequence	SEQ ID NO: 178	2300 aa	MW at 257261.9kD

MDFLIIFLLLFIGTSETQVDVSNVVPGTRYDITISSISTTYTSPVTRIVTTNVTEPGPPVFLAGERVG
SAGILLSWNTPPNPNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLNPGTTYEIKVAAENS
AGIGVFSDPFLFQTAESPAPGKVVNLTVEAYNASAVKLIWYLPRQPNGKITSFKISVKHARSGIVVKD
VSIRVEDILTGKLPECNVENSESFLWSTASPSPTLGRVTPPSRTTHSSSTLTQNEISSVWKEPISFVV
THLRPYTTYLFEVSAATTEAGYIDSTIVRTPESVPEGPPQNCVTGNITGKSFSILWDPPTIVTGKFSY
RVELYGPSGRILDNSTKDLKFAFTNLTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVPGAVFDLQLA
EVESTQVRITWKKPRQPNGIINQYRVKVLVPETGIILENTLLTGNNEINDPMAPEIVNIVEPMVGLYE
GSAEMSSDLHSLATFIYNSHPDKNFPARNRAEDQTSPVVTTRNQYITDIAAEQLSYVIRRLVPFTEHM
ISVSAFTIMGEGPPTVLSVRTRQQVPSSIKIINYKNISSSSILLYWDPPEYPNGKITHYTIYAMELDT
NRAFQITTIDNSFLITGIGLKKYTKYKMRVAASTHVGESSLSEENDIFVRTSEDEPESSPQDVEVIDV
TADEIRLKWSPPEKPNGIIIAYEVLYKNIDTLYMKNTSTTDIILRNLRPHTLYNISVRSYTRFGHGNQ
VSSLLSVRTSETVPDSAPENITYKNISSGEIELSFLPPSSPNGIIQKYTIYLKRSNGNEERTINTTSL
TQNILKKYTQYIIEVSASTLKGEGVRSAPISILTEEDAPDSPPQDFSVKQLSGVTVKLSWQPPLEPNG
IILYYTVYVWRNRSSLKTINVTETSLELSDLDYNVEYSAYVTASTRFGDGKTRSNIISFQTPEGPSDP
PKDVYYANLSSSSIILFWTPPSKPNGIIQYYSVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKLT
IFSYYTFWLTASTSVGNGNKSSDIIEVYTDQDVPEGFVGNLTYESISSTAINVSWVPPAQPNGLVFYY
VSLILQQTPRHVRPPLVTYERSIYFDNLEKYTDYILKITPSTEKGFSDTYTAQLYIKTEEDVPETSPI
INTFKNLSSTSVLLSWDPPVKPNGAIISYDLTLQGPNENYSFITSDNYIILEELSPFTLYSFFAAART
RKGLGPSSILFFYTDESVPLAPPQNLTLINCTSDFVWLKWSPSPLPGGIVKVYSFKIHEHETDTIYYK
NISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFTTQESVPDVVQNMQCMATSWQSVLVK
WDPPKKANGIITQYMVTVERNSTKVSPQDHMYTFIKLLANTSYVFKVRASTSAGEGDESTCHVSTLPE
TVPSVPTNIAFSDVQSTSATLTWIRPDTILGYFQNYKITTQLRAQKCKEWESEECVEYQKIQYLYEAH
LTEETVYGLKKFRWYRFQVAASTNAGYGNASNWISTKTLPGPPDGPPENVHVVATSPFSISISWSEPA
VITGPTCYLIDVKSVDNDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAFTGNISAAYVEGKSSAE
MIVTTLESAPKDPPNNMTFQKIPDEVTKFQLTSLPPSQPNGNIQVYQALVYREDDPTAVQIHNLSIIQ
KTNTFVIAMLEGLKGGHTYNISVYAVNSAGAGPKVPMRITMDIKAPARPKTKPTPIYDATGKLLVTST
TITIRMPICYYSDDHGPIKNVQVLVTETGAQHDGNVTKWYDAYFNKARPYFTNEGFPNPPCTEGKTKF
SGNEEIYIIGADNACMIPGNEDKICNGPLKPKKQYLFKFRATNIMGQFTDSDYSDPVKTLGEGLSERT
LEIILSVTLCILSIIILLGTAIFAFARIRQKQKEGGTYSPQDAEIIDTKLKLDQLITVADLELKDERLT
RLLSYRKSIKPISKKSFLQHVEELCTNNNLKFQEEFSELPKFLQDLSSTDADLPWNRAKNRFPNIKPY
NNNRVKLIADASVPGSDYINASYISGYLCPNEFIATQGPLPGTVGDFWRMVWETRAKTLVMLTQCFEK
GRIRCHQYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDCMTVRQCNFTAWPEHGVPENS
APLIHFVKLVRASRAHDTTPMIVHCSAGVGRTGVFIALDHLTQHINDHDFVDIYGLVAELRSERMCMV
QNLAQYIFLHQCILDLLSNKGSNQPICFVNYSALQKMDSLDAMEGDVELEWEETTM

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 15B.

Table 15B. Comparison of the NOV15 protein sequences.

```
NOV15a  MDFLIIFLLLFIGTSETQVDVSNVVPGTRYDITISSISTTYTSPVTRIGASNEPGPPVFL
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  MDFLIIFLLLFIGTSETQVDVSNVVPGTRYDITISSISTTYTSPVTRIGASNEPGPPVFL
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
```

```
NOV15g  ------------------------------------------------------------

NOV15a  AGERVGSAGILLSWNTPPNPNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLN
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  AGERVGSAGILLSWNTPPNPNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLN
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  PGTTYEIKVAAENSAGIGVFSDPFLFQTAESAPGKVVDFTGEAVPFSSKLMWYTSATKKK
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  PGTTYEIKVAAENSAGIGVFSDPFLFQTAESAPGKVVDFTGEAVPFSSKLMWYTSATKKK
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  ITSFKISVKHNRSGIVVKEVSIRVECILSASLPLHCNENSESFLWSTASPSPTLGRVTPP
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  ITSFKISVKHNRSGIVVKEVSIRVECILSASLPLHCNENSESFLWSTASPSPTLGRVTPP
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  SRTTHSSSTLTQNEISSVKEPISFVVTHLRPYTTYLFEVSAATTEAGYIDSTIVRTPESV
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  SRTTHSSSTLTQNEISSVKEPISFVVTHLRPYTTYLFEVSAATTEAGYIDSTIVRTPESV
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  PEGPPQNCVTGNITGKSFSILWDPPTIVTGKFSYRVELYGPSAGRILDNSTKDLKFAFTN
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  PEGPPQNCVTGNITGKSFSILWDPPTIVTGKFSYRVELYGPSAGRILDNSTKDLKFAFTN
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  LTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVPGAVFDLQLAEVESTQVRITWKKPRQP
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  LTPFTMYDVYIAAETSAGTGPKSNISVFTPPDVPGAVFDLQLAEVESTQVRITWKKPRQP
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  NGIINQYRVKVLVPETGIILENTLLTGNNEINDPMAPEIVNIVQPMVGLYEGSAEMSSDL
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  NGIINQYRVKVLVPETGIILENTLLTGNNEINDPMAPEIVNIVQPMVGLYEGSAEMSSDL
```

```
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  HSLATFIYNSHPDKNFPARNRAEDQTSPVVTTRNQYITDIAAEQLTYVLIRLRRFWAETM
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  HSLATFIYNSHPDKNFPARNRAEDQTSPVVTTRNQYITDIAAEQLTYVLIRLRRFWAETM
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  GFSRYTIMSSASRDNLTSPGPLSAQNFRVTHVTITEVFLHWDPPDPVFFHHYLITILDVE
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  GFSRYTIMSSASRDNLTSPGPLSAQNFRVTHVTITEVFLHWDPPDPVFFHHYLITILDVE
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  NQSKSIILRTLNSLSLVLIGLKKYTKYKMRVAASTHVGESSLSEENDIFVRTSEDEPESS
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  NQSKSIILRTLNSLSLVLIGLKKYTKYKMRVAASTHVGESSLSEENDIFVRTSEDEPESS
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  PQDVEVIDVTADEIRLKWSPPEKPNGIIIAYEVLYKNIDTLYMKNTSTTDIILRNLRPHT
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  PQDVEVIDVTADEIRLKWSPPEKPNGIIIAYEVLYKNIDTLYMKNTSTTDIILRNLRPHT
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  LYNISVRSYTRFGHGNQVSSLLSVRTSESVPDSAPENITYKNISSGEIELSFLPPSSPNG
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  LYNISVRSYTRFGHGNQVSSLLSVRTSESVPDSAPENITYKNISSGEIELSFLPPSSPNG
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  IIQKYTIYLKRSNGNEERTINTTSLTQNIKGLKKYTQYIIEVSASTLKGEGVRSAPISIL
NOV15b  ------------------------------------------------------------
NOV15c  ------------------------------------------------------------
NOV15d  IIQKYTIYLKRSNGNEERTINTTSLTQNIKGLKKYTQYIIEVSASTLKGEGVRSAPISIL
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ------------------------------------------------------------

NOV15a  TEEDAPDSPPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVYVWRSSLKTINVTETSLE
NOV15b  ----------------------------------------------MDFLIIFLLLFI
```

```
NOV15c  ------------------------------------------------------------
NOV15d  TEEDAPDSPPQDFSVKQLSGVTVKLSWQPPLEPNGIILYYTVYVWRSSLKTINVTETSLE
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ----------------------------------------------MDFLIIFLLLFI

NOV15a  LSDLDYNVEYSAYVTASTRFGDGKTRSNIISFQTPEGPSDPPKDVYYANLSSSSIILFWT
NOV15b  GTSETQVDVSNVVPGTRYDITISSISTTYTSPVTRIVTTNVTKPGPPVFLAGERVGSAGI
NOV15c  ------------------------------------------------------------
NOV15d  LSDLDYNVEYSAYVTASTRFGDGKTRSNIISFQTPEGPSDPPKDVYYANLSSSSIILFWT
NOV15e  -----QVDVSNVVPGTRYDITISSISTTYTSPVTRIVTTNVTKPGPPVFLAGERVGSAGI
NOV15f  ------------------------------------------------------------
NOV15g  GTSETQVDVSNVVPGTRYDITISSISTTYTSPVTRIVTTNVTEPGPPVFLAGERVGSAGI

NOV15a  PPSKPNGIIQYYSVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKLTIFSYYTFWLTA
NOV15b  LLSWNTPPNPNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLNPGTTYEIKVA
NOV15c  ------------------------------------------------------------
NOV15d  PPSKPNGIIQYYSVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKLTIFSYYTFWLTA
NOV15e  LLSWNTPPNPNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLNPGTTYEIKVA
NOV15f  ------------------------------------------------------------
NOV15g  LLSWNTPPNPNGRIISYIVKYKEVCPWMQTVYTQVRSKPDSLEVLLTNLNPGTTYEIKVA

NOV15a  STSVGNGNKSSDIIEVYTDQDVPEGFVGNLTYESISSTAINVSWVPPAQPNG--------
NOV15b  AENSAGIGVFSDPFLFQTAES-APGKVVNLTVEAYNASAVKLIWYLPRQPNGKITSFKIS
NOV15c  ------------------------------------------------------------
NOV15d  STSVGNGNKSSDIIEVYTDQDVPEGFVGNLTYESISSTAINVSWVPPAQPNG--------
NOV15e  AENSAGIGVFSDPFLFQTAES-APGKVVNLTVEAYNASAVKLIWYLPRQPNGKITSFKIS
NOV15f  ----------------------PEGFVGNLTYESISSTAINVSWVPPAQPNG--------
NOV15g  AENSAGIGVFSDPFLFQTAESPAPGKVVNLTVEAYNASAVKLIWYLPRQPNGKITSFKIS

NOV15a  -------------------------LVFYY--------VSLILQQTPRHVRPPLVTYERS
NOV15b  VKHARSGIVVKDVSIRVEDILTGKLPECNE-NSESFLWSTASPSPTLGRVTPPSRTTHSS
NOV15c  ------------------------------------------------------------
NOV15d  -------------------------LVFYY--------VSLILQQTPRHVRPPLVTYERS
NOV15e  VKHARSGIVVKDVSIRVEDILTGKLPECNE-NSESFLWSTASPSPTLGRVTPPSRTTHSS
NOV15f  -------------------------LVFYY--------VSLILQQTPRHVRPPLVTYERS
NOV15g  VKHARSGIVVKDVSIRVEDILTGKLPECNVENSESFLWSTASPSPTLGRVTPPSRTTHSS

NOV15a  IYFDN--------------LEKYTDYILKITPSTEKGFSDTYTAQLYIKTEEDIPETSPI
NOV15b  STLTQNEISSVWKEPISFVVTHLRPYTTYLFEVSAVTTEAGYIDSTIVRTPESVPEGPPQ
NOV15c  ----------------------------------------------------RSPEGPPQ
NOV15d  IYFDN--------------LEKYTDYILKITPSTEKGFSDTYTAQLYIKTEEDIPETSPI
NOV15e  STLTQNEISSVWKEPISFVVTHLRPYTTYLFEVSAVTTEAGYIDSTIVRTPESVPEGPPQ
NOV15f  IYFDN--------------LEKYTDYILKITPSTEKGFSDTYTAQLYIKTEEDVPETSPI
NOV15g  STLTQNEISSVWKEPISFVVTHLRPYTTYLFEVSAATTEAGYIDSTIVRTPESVPEGPPQ

NOV15a  INTFKNLSSTSVLLSWDPPVKPNGAIISYDLTLQGPN-ENYSFITSDNYIILEELSPFTL
NOV15b  NCVTGNITGKSFSILWDPPTIVT-GKFSYRVELYGPSGRILDNSTKDLKFAFTNLTPFTM
NOV15c  NCVTGNITGKSFSILWDPPTIVT-GKFSYRVELYGPSGRILDNSTKDLKFAFTNLTPFTM
NOV15d  INTFKNLSSTSVLLSWDPPVKPNGAIISYDLTLQGPN-ENYSFITSDNYIILEELSPFTL
NOV15e  NCVTGNITGKSFSILWDPPTIVT-GKFSYRVELYGPSGRILDNSTKDLKFAFTNLTPFTM
NOV15f  INTFKNLSSTSVLLSWDPPVKPNGAIISYDLTLQGPN-ENYSFITSDNYIILEELSPFTL
NOV15g  NCVTGNITGKSFSILWDPPTIVT-GKFSYRVELYGPSGRILDNSTKDLKFAFTNLTPFTM
```

```
NOV15a  YSFFAAARTRKGLGPSSILFFYTDESVPLAPPQNLTLINCTSDFVWLKWSPSPLPGGIVK
NOV15b  YDVYIAAETSAGTGPKSNISVFTPPDVPGA-VFDLQLAEVESTQVRITWKKPRQPNGIIN
NOV15c  YDVYIAAETSAGTGPKSNISVFTPPDVPGA-VFDLQLAEVESTQVRITWKKPRQPNGIIN
NOV15d  YSFFAAARTRKGLGPSSILFFYTDESVPLAPPQNLTLINCTSDFVWLKWSPSPLPGGIVK
NOV15e  YDVYIAAETSAGTGPKSNISVFTPPDVPGA-VFDLQLAEVESTQVRITWKKPRQPNGIIN
NOV15f  YSFFAAARTRKGLGPSSILFFYTDESVPLAPPQNLTLINCTSDFVWLKWSPSPLPGGIVK
NOV15g  YDVYIAAETSAGTGPKSNISVFTPPDVPGA-VFDLQLAEVESTQVRITWKKPRQPNGIIN

NOV15a  VYSFKIHEHETDTIYYKNISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFT
NOV15b  QYRVKVLVPETGIILENTLLTGNNEYINDPMAPEIVNIVEPMVGLYEGSAEMSSDLHSLA
NOV15c  QYRVKVLVPETGIILENTLLTGNNEYINDPMAPEIVNIVEPMVGLYEGSAEMSSDLHSLA
NOV15d  VYSFKIHEHETDTIYYKNISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFT
NOV15e  QYRVKVLVPETGIILENTLLTGNNEYINDPMAPEIVNIVEPMVGLYEGSAEMSSDLHSLA
NOV15f  VYSFKIHEHETDTIYYKNISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFT
NOV15g  QYRVKVLVPETGIILENTLLTGNN-EINDPMAPEIVNIVEPMVGLYEGSAEMSSDLHSLA

NOV15a  TQESVPDVVQNMQCMATSWQSVLVKWDPPKKANGIITQYMVTVERNSTKVSPQDHMYTFI
NOV15b  TFIYNSHPDK--------------NFPARNRAEDQTSPVVTTRNQYITDIAAEQLSYVIR
NOV15c  TFIYNSHPDK--------------NFPARNRAEDQTSPVVTTRNQYITDIAAEQLSYVIR
NOV15d  TQESVPDVVQNMQCMATSWQSVLVKWDPPKKANGIITQYMVTVERNSTKVSPQDHMYTFI
NOV15e  TFIYNSHPDK--------------NFPARNRAEDQTSPVVTTRNQYITDIAAEQLSYVIR
NOV15f  TQESVPDVVQNMQCMATSWQSVLVKWDPPKKANGIITQYMVTVERNSTKVSPQDHMYTFI
NOV15g  TFIYNSHPDK--------------NFPARNRAEDQTSPVVTTRNQYITDIAAEQLSYVIR

NOV15a  KLLANTSYVFKVRASTSAGEGDESTCHVSTLPETVPSVPTNIAFSDVQSTSATLTWIRPD
NOV15b  RLVPFTEHMISVSAFTIMGEG-PPTVLSVRTRQQVPSSIKIINYKNISSSSILLYWDPPE
NOV15c  RLVPFTEHMISVSAFTIMGEG-PPTVLSVRTRQQVPSSIKIINYKNISSSSILLYWDPPE
NOV15d  KLLANTSYVFKVRASTSAGEGDESTCHVSTLPETVPSVPTNIAFSDVQSTSATLTWIRPD
NOV15e  RLVPFTEHMISVSAFTIMGEG-PPTVLSVRTRQQVPSSIKIINYKNISSSSILLYWDPPE
NOV15f  KLLANTSYVFKVRASTSAGEGDESTCHVSTLPETVPSVPTNIAFSDVQSTSATLTWIRPD
NOV15g  RLVPFTEHMISVSAFTIMGEG-PPTVLSVRTRQQVPSSIKIINYKNISSSSILLYWDPPE

NOV15a  TILGYFQNYKITTQLRAQKCKEWESEECVEYQKIQYLYEAHLTEETVYGLKKFRWYRFQV
NOV15b  YPNGKITHYTIY------------AMELDTNRAFQITTIDNSFLIT--GLKKYTKYKMRV
NOV15c  YPNGKITHYTIY------------AMELDTNRAFQITTIDNSFLIT--GLKKYTKYKMRV
NOV15d  TILGYFQNYKITTQLRAQKCKEWESEECVEYQKIQYLYEAHLTEETVYGLKKFRWYRFQV
NOV15e  YPNGKITHYTIY------------AMELDTNRAFQITTIDNSFLIT--GLKKYTKYKMRV
NOV15f  TILGYFQNYKITTQLRAQKCKEWESEECVEYQKIQYLYEAHLTEETVYGLKKFRWYRFQV
NOV15g  YPNGKITHYTIY------------AMELDTNRAFQITTIDNSFLITGIGLKKYTKYKMRV

NOV15a  AASTNAGYGNAS--NWISTKTLPGPPDGPPENVHVVATSPFSISISWSEPAVITGPTCYL
NOV15b  AASTHDGESSLSEENDIFVRTSEDEPESSPQDVEVIDVTADEIRLKWSPPEKPNG----I
NOV15c  AASTHDGESSLSEENDIFVRTSEDEPESSPQDVEVIDVTADEIRLKWSPPEKPNG----I
NOV15d  AASTNAGYGNAS--NWISTKTLPGPPDGPPENVHVVATSPFSISISWSEPAVITGPTCYL
NOV15e  AASTHDGESSLSEENDIFVRTSEDEPESSPQDVEVIDVTADEIRLKWSPPEKPNG----I
NOV15f  AASTNAGYGNAS--NWISTKTLPGPPDGPPENVHVVATSPFSISISWSEPAVITGPTCYL
NOV15g  AASTHVGESSLSEENDIFVRTSEDEPESSPQDVEVIDVTADEIRLKWSPPEKPNG----I

NOV15a  IDVKSVDNDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAFTGNISAAYVEGKSSAEM
NOV15b  IIAYEVLYKNIDTLYMK-NTSTTDIILRNLRPHTLYNISVRSYTRFGHGNQVS----SLL
NOV15c  IIAYEVLYKNIDTLYMK-NTSTTDIILRNLRPHTLYNISVRSYTRFGHGNQVS----SLL
NOV15d  IDVKSVDNDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAFTGNISAAYVEGKSSAEM
NOV15e  IIAYEVLYKNIDTLYMK-NTSTTDIILRNLRPHTLYNISVRSYTRFGHGNQVS----SLL
NOV15f  IDVKSVDNDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAFTGNISAAYVEGKSSAEM
```

```
NOV15g  IIAYEVLYKNIDTLYMK-NTSTTDIILRNLRPHTLYNISVRSYTRFGHGNQVS----SLL

NOV15a  IVTTLESAPKDPPNNMTFQKIPDEVTKFQLTFLPPSQPNGNIQVYQALVYREDDPTAVQI
NOV15b  SVRTSETVPDSAPENITYKNIS--SGEIELSFLPPSSPNGIIKKYTIYLKRSNGNEERTI
NOV15c  SVRTSETVPDSAPENITYKNIS--SGEIELSFLPPSSPNGIIKKYTIYLKRSNGNEERTI
NOV15d  IVTTLESAPKDPPNNMTFQKIPDEVTKFQLTFLPPSQPNGNIQVYQALVYREDDPTAVQI
NOV15e  SVRTSETVPDSAPENITYKNIS--SGEIELSFLPPSSPNGIIQKYTIYLKRSNGNEERTI
NOV15f  IVTTLESAPKDPPNNMTFQKIPDEVTKFQLTSLPPSQPNGNIQVYQALVYREDDPTAVQI
NOV15g  SVRTSETVPDSAPENITYKNIS--SGEIELSFLPPSSPNGIIQKYTIYLKRSNGNEERTI

NOV15a  HNLSIIQKTNTFVIAMLEGLKGGHTYNISVYAVNSAG-AGPKVPMRITMDIKAP-ARPKT
NOV15b  NTTSLTQNIK--------VLKKYTQYIIEVSASTLKGEGVRSAPISILTEEDAPDSPPQD
NOV15c  NTTSLTQNIK--------VLKKYTQYIIEVSASTLKGEGVRSAPISILTEEDAPDSPPQD
NOV15d  HNLSIIQKTNTFVIAMLEGLKGGHTYNISVYAVNSAG-AGPKVPMRITMDIKAP-ARPKT
NOV15e  NTTSLTQNIK--------GLKKYTQYIIEVSASTLKGEGVRSAPISILTEEDAPDSPPQD
NOV15f  HNLSIIQKTNTFVIAMLEGLKGGHTYNISVYAVNSAG-AGPKVPMRITMDIKAP-ARPKT
NOV15g  NTTSLTQNI----------LKKYTQYIIEVSASTLKGEGVRSAPISILTEEDAPDSPPQD

NOV15a  KPTPIYDATGKLLVTSTTITIRMPICYYSDDHGP--IKNVQVLVTETGAQH-DGNVTKWY
NOV15b  FSVKQLSGVTVKLSWQPPLEPNGIILYYTVYVWN-RSSLKTINVTETSLELSDLDYNVEY
NOV15c  FSVKQLSGVTVKLSWQPPLEPNGIILYYTVYVWN-RSSLKTINVTETSLELSDLDYNVEY
NOV15d  KPTPIYDATGKLLVTSTTITIRMPICYYSDDHGP--IKNVQVLVTETGAQH-DGNVTKWY
NOV15e  FSVKQLSGVTVKLSWQPPLEPNGIILYYTVYVWN-RSSLKTINVTETSLELSDLDYNVEY
NOV15f  KPTPIYDATGKLLVTSTTITIRMPICYYSDDHGP--IKNVQVLVTETGAQH-DGNVTKWY
NOV15g  FSVKQLSGVTVKLSWQPPLEPNGIILYYTVYVWRNRSSLKTINVTETSLELSDLDYNVEY

NOV15a  DAYFNKARPYFTNEGFPNPP--CT-EGKTKFSGNEEIYIIGADNACMIPGNEDKICNGPL
NOV15b  SAYVTASTRFGDGKTRSNIISFQTPEGAPSDPPKDVYYANLSSSSIILFWTPPSKPNGII
NOV15c  SAYVTASTRFGDGKTRSNIISFQTPEGAPSDPPKDVYYANLSSSSIILFWTPPSKPNGII
NOV15d  DAYFNKARPYFTNEGFPNPP--CT-EGKTKFSGNEEIYIIGADNACMIPGNEDKICNGPL
NOV15e  SAYVTAS-------------------------------------------------------
NOV15f  DAYFNKARPYFTNEGFPNPP--CT-EGKTKFSGNEEIYIIGADNACMIPGNEDKICNGPL
NOV15g  SAYVTASTRFGDGKTRSNIISFQTPEGP-SDPPKDVYYANLSSSSIILFWTPPSKPNGII

NOV15a  KPKKQYLFKFRATNIMGQFTDSDYSDPVKTLGEGLSERTVEIILSVTLCILSIILLGTAI
NOV15b  QYYSVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKLTIFSYYTFWLTASTSVGNGNK
NOV15c  QYYSVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKLTIFSYYTFWLTASTSVGNGNK
NOV15d  KPKKQYLFKFRATNIMGQFTDSDYSDPVKTLGEGLSERTVEIILSVTLCILSIILLGTAI
NOV15e  -------------------------------------------------------------
NOV15f  KPKKQYLFKFRATNIMGQFTDSDYSDPVKTLGEGLSERT----------------------
NOV15g  QYYSVYYRNTSGTFMQNFTLHEVTNDFDNMTVSTIIDKLTIFSYYTFWLTASTSVGNGNK

NOV15a  FAFARIRQKQKEGGTYSPQDAEIIDTKLKLDQLITVADLELKDERLTRPISKKSFLQHVE
NOV15b  SSDIIEVYTDQDVPEGFVGNLTYESISSTAINVSWVPPAQPNGLVFYYVSLILQQTPRHV
NOV15c  SLE----------------------------------------------------------
NOV15d  FAFARIRQKQKEGGTYSPQDAEIIDTKLKLDQLITVADLELKDERLTRPISKKSFLQHVE
NOV15e  -------------------------------------------------------------
NOV15f  -------------------------------------------------------------
NOV15g  SSDIIEVYTDQDVPEGFVGNLTYESISSTAINVSWVPPAQPNGLVFYYVSLILQQTPRHV

NOV15a  ELCTNNNLKFQEEFSELPKFLQDLSSTDADLPWNRAKNRFPNIKPYNNNRVKLIADASVP
NOV15b  RPPLVTYERSIYFDNLEKYTDYILKITPSTEKGFSDTYTAQLYIKTEEDVPETSPIINTF
NOV15c  -------------------------------------------------------------
NOV15d  ELCTNNNLKFQEEFSELPKFLQDLSSTDADLPWNRAKNRFPNIKPYNNNRVKLIADASVP
```

```
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  RPPLVTYERSIYFDNLEKYTDYILKITPSTEKGFSDTYTAQLYIKTEEDVPETSPIINTF

NOV15a  GSDYINASYISGYLCPNEFIATQGPLPGTVGDFWRMVWETRAKTLVMLTQCFEKGRIRCH
NOV15b  KNLSSTSVLLSWDPPVKPNGAIISYDLTLQGPNENYSFITSDNYIILEELSPFTLYSFFA
NOV15c  ------------------------------------------------------------
NOV15d  GSDYINASYISGYLCPNEFIATQGPLPGTVGDFWRMVWETRAKTLVMLTQCFEKGRIRCH
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  KNLSSTSVLLSWDPPVKPNGAIISYDLTLQGPNENYSFITSDNYIILEELSPFTLYSFFA

NOV15a  QYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDCMTVRQCNFTAWPEHGVPE
NOV15b  AARTRKGLGPSSILFFYTDESVPLAPPQNLTLINCTSDFVWLKWSPSPLPGGIVKVYSFK
NOV15c  ------------------------------------------------------------
NOV15d  QYWPEDNKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDCMTVRQCNFTAWPEHGVPE
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  AARTRKGLGPSSILFFYTDESVPLAPPQNLTLINCTSDFVWLKWSPSPLPGGIVKVYSFK

NOV15a  NSAPLIHFVKLVRASRAHDTTPMIVHCSAGVGRTGVFIALDHLTQHINDHDFVDIYGLVA
NOV15b  IHEHETDTIYYKNISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFTTQESV
NOV15c  ------------------------------------------------------------
NOV15d  NSAPLIHFVKLVRASRAHDTTPMIVHCSAGVGRTGVFIALDHLTQHINDHDFVDIYGLVA
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  IHEHETDTIYYKNISGFKTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVKFTTQESV

NOV15a  ELRSERMCMVQNLAQYIFLHQCILDLLSNKGSNQPICFVNYSALQKMDSLDAMEGGDVEL
NOV15b  PDVVQNMQCMATSWQSVLVKWDPPKKANGIITQYMVTVERNSTKVSPQDHMYTFIKLLAN
NOV15c  ------------------------------------------------------------
NOV15d  ELRSERMCMVQNLAQYIFLHQCILDLLSNKGSNQPICFVNYSALQKMDSLDAMEGGDVEL
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  PDVVQNMQCMATSWQSVLVKWDPPKKANGIITQYMVTVERNSTKVSPQDHMYTFIKLLAN

NOV15a  EWEETTM-----------------------------------------------------
NOV15b  TSYVFKVRASTSAGEGDESTCHVSTLPETVPSVPTNIAFSDVQSTSATLTWIRPDTILGY
NOV15c  ------------------------------------------------------------
NOV15d  EWEETTM-----------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  TSYVFKVRASTSAGEGDESTCHVSTLPETVPSVPTNIAFSDVQSTSATLTWIRPDTILGY

NOV15a  ------------------------------------------------------------
NOV15b  FQNYKITTQLRAQKCKEWESEECVEYQKIQYLYEAHLTEETVYGLKKFRWYRFQVAASTN
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  FQNYKITTQLRAQKCKEWESEECVEYQKIQYLYEAHLTEETVYGLKKFRWYRFQVAASTN

NOV15a  ------------------------------------------------------------
NOV15b  AGYGNASNWISTKTLPGPPDGPPENVHVVATSPFSISISWSEPAVITGPTCYLIDVKSVD
```

```
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  AGYGNASNWISTKTLPGPPDGPPENVHVVATSPFSISISWSEPAVITGPTCYLIDVKSVD

NOV15a  ------------------------------------------------------------
NOV15b  NDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAFTGNISAAYVEGKSSAEMIVTTLES
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  NDEFNISFIKSNEENKTIEIKDLEIFTRYSVVITAFTGNISAAYVEGKSSAEMIVTTLES

NOV15a  ------------------------------------------------------------
NOV15b  APKDPPNNMTFQKIPDEVTKFQLTSLPPSQPNGNIQVYQALVYREDDPTAVQIHNLSIIQ
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  APKDPPNNMTFQKIPDEVTKFQLTSLPPSQPNGNIQVYQALVYREDDPTAVQIHNLSIIQ

NOV15a  ------------------------------------------------------------
NOV15b  KTNTFVIAMLEGLKGGHTYNISVYAVNSAGAGPKVPMRITMDIKAPARPKTKPTPIYDAT
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  KTNTFVIAMLEGLKGGHTYNISVYAVNSAGAGPKVPMRITMDIKAPARPKTKPTPIYDAT

NOV15a  ------------------------------------------------------------
NOV15b  GKLLVTSTTITIRMPICYYSDDHGPIKNVQVLVTETGAQHDGNVTKWYDAYFNKARPYFT
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  GKLLVTSTTITIRMPICYYSDDHGPIKNVQVLVTETGAQHDGNVTKWYDAYFNKARPYFT

NOV15a  ------------------------------------------------------------
NOV15b  NEGFPNPPCTEGKTKFSGNEEIYIIGADNACMIPGNEDKICNGPLKPKKQYLFKFRATNI
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  NEGFPNPPCTEGKTKFSGNEEIYIIGADNACMIPGNEDKICNGPLKPKKQYLFKFRATNI

NOV15a  ------------------------------------------------------------
NOV15b  MGQFTDSDYSDPVKTLGEGLSERTLEIILSVTLCILSIILLGTAIFAFARIRQKQKEGGT
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  MGQFTDSDYSDPVKTLGEGLSERTLEIILSVTLCILSIILLGTAIFAFARIRQKQKEGGT
```

```
NOV15a  ------------------------------------------------------------
NOV15b  YSPQDAEIIDTKLKLDQLITVADLELKDERLTRLLSYRKSIKPISKKSFLQHVEELCTNN
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  YSPQDAEIIDTKLKLDQLITVADLELKDERLTRLLSYRKSIKPISKKSFLQHVEELCTNN

NOV15a  ------------------------------------------------------------
NOV15b  NLKFQEEFSELPKFLQDLSSTDADLPWNRAKNRFPNIKPYNNNRVKLIADASVPGSDYIN
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  NLKFQEEFSELPKFLQDLSSTDADLPWNRAKNRFPNIKPYNNNRVKLIADASVPGSDYIN

NOV15a  ------------------------------------------------------------
NOV15b  ASYISGYLCPNEFIATQGPLPGTVGDFWRMVWETRAKTLVMLTQCFEKGRIRCHQYWPED
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  ASYISGYLCPNEFIATQGPLPGTVGDFWRMVWETRAKTLVMLTQCFEKGRIRCHQYWPED

NOV15a  ------------------------------------------------------------
NOV15b  NKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDCMTVRQCNFTAWPEHGVPENSAPLI
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  NKPVTVFGDIVITKLMEDVQIDWTIRDLKIERHGDCMTVRQCNFTAWPEHGVPENSAPLI

NOV15a  ------------------------------------------------------------
NOV15b  HFVKLVRASRAHDTTPMIVHCSAGVGRTGVFIALDHLTQHINDHDFVDIYGLVAELRSER
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  HFVKLVRASRAHDTTPMIVHCSAGVGRTGVFIALDHLTQHINDHDFVDIYGLVAELRSER

NOV15a  ------------------------------------------------------------
NOV15b  MCMVQNLAQYIFLHQCILDLLSNKGSNQPICFVNYSALQKMDSLDAMEGDVELEWEETTM
NOV15c  ------------------------------------------------------------
NOV15d  ------------------------------------------------------------
NOV15e  ------------------------------------------------------------
NOV15f  ------------------------------------------------------------
NOV15g  MCMVQNLAQYIFLHQCILDLLSNKGSNQPICFVNYSALQKMDSLDAMEGDVELEWEETTM

NOV15a  (SEQ ID NO:  166)
NOV15b  (SEQ ID NO:  168)
NOV15c  (SEQ ID NO:  170)
NOV15d  (SEQ ID NO:  172)
NOV15e  (SEQ ID NO:  174)
NOV15f  (SEQ ID NO:  176)
```

NOV15g (SEQ ID NO: 178)

Further analysis of the NOV15a protein yielded the following properties shown in Table 15C.

Table 15C. Protein Sequence Properties NOV15a

SignalP analysis:	Cleavage site between residues 18 and 19

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 2;  pos.chg 0;  neg.chg 1
      H-region:  length 13;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   2.26
      possible cleavage site: between 17 and 18

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   3
      INTEGRAL    Likelihood = -2.55   Transmembrane    3 -  19
      INTEGRAL    Likelihood = -1.12   Transmembrane  605 - 621
      INTEGRAL    Likelihood =-10.46   Transmembrane 1894 -1910
      PERIPHERAL  Likelihood =  1.59 (at 2178)
      ALOM score: -10.46  (number of TMSs: 3)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 10
      Charge difference: -1.0   C(-1.0) - N( 0.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  3.60
      Hyd Moment(95):  6.71    G content:       1
      D/E content:     2       S/T content:     2
      Score: -7.20

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: KKPR (4) at  415
      pat4: KPKK (4) at 1855
      pat7: PLKPKKQ (4) at 1853
      bipartite: none
      content of basic residues:   8.8%
      NLS Score:  0.37
```

```
KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 70.6

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       56.5 %: endoplasmic reticulum
       26.1 %: mitochondrial
       17.4 %: nuclear

>> prediction for CG50718-02 is end (k=23)
```

A search of the NOV15a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 15D.

Table 15D. Geneseq Results for NOV15a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV15a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAO18736	Human NOV2a protein - Homo sapiens, 2281 aa. [WO200257450-A2, 25-JUL-2002]	1..2281 1..2281	2281/2281 (100%) 2281/2281 (100%)	0.0
AAG79724	Human KPP-2, Incyte ID No. 7480588CD1 - Homo sapiens, 2299 aa. [WO200290530-A2, 14-NOV-2002]	1..2281 1..2299	2180/2306 (94%) 2214/2306 (95%)	0.0
AAO18738	Human NOV2c protein - Homo sapiens, 2300 aa. [WO200257450-A2, 25-JUL-2002]	1..2281 1..2300	2177/2305 (94%) 2205/2305 (95%)	0.0
ABP60057	Human phosphatase protein SEQ ID #2 - Homo sapiens, 2291 aa. [WO200279452-A2, 10-OCT-2002]	1..2281 1..2291	2171/2296 (94%) 2209/2296 (95%)	0.0
ABP60058	Human phosphatase related protein #SEQ ID 4 - Homo sapiens, 2301 aa. [WO200279452-A2, 10-OCT-2002]	1..2281 1..2301	1893/2306 (82%) 2077/2306 (89%)	0.0

In a BLAST search of public sequence databases, the NOV15a protein was found to have homology to the proteins shown in the BLASTP data in Table 15E.

Table 15E. Public BLASTP Results for NOV15a				
Protein Accession Number	**Protein/Organism/Length**	**NOV15a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
O88488	Glomerular mesangial cell receptor protein-tyrosine phosphatase precursor - Rattus norvegicus (Rat), 2302 aa.	1..2281 2..2302	1893/2306 (82%) 2077/2306 (89%)	0.0
Q8BY76	Hypothetical fibronectin type III domain/fibronectin type III repeat containing protein - Mus musculus (Mouse), 1086 aa (fragment).	1..1042 1..1053	791/1058 (74%) 897/1058 (84%)	0.0
AAO42638	RE52018p - Drosophila melanogaster (Fruit fly), 1631 aa.	1142..2244 497..1531	311/1133 (27%) 505/1133 (44%)	2e-99
Q8IR87	CG1817-PC - Drosophila melanogaster (Fruit fly), 1556 aa.	1142..2244 497..1531	311/1133 (27%) 505/1133 (44%)	2e-99
Q9VYW1	CG1817 protein - Drosophila melanogaster (Fruit fly), 1962 aa.	1142..2244 497..1531	311/1133 (27%) 505/1133 (44%)	2e-99

PFam analysis indicates that the NOV15a protein contains the domains shown in the Table 15F.

Table 15F. Domain Analysis of NOV15a			
Pfam Domain	**NOV15a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
fn3	54..141	30/91 (33%) 63/91 (69%)	1.6e-14
fn3	301..383	25/88 (28%) 59/88 (67%)	2.7e-06
fn3	561..643	26/86 (30%) 56/86 (65%)	7.9e-07
fn3	657..738	28/87 (32%) 62/87 (71%)	4.2e-14
fn3	751..834	29/87 (33%) 67/87 (77%)	9.4e-12
fn3	846..926	24/87 (28%) 62/87 (71%)	7.6e-12
fn3	938..1030	25/97 (26%) 68/97 (70%)	6.8e-08
fn3	1043..1124	19/87 (22%) 56/87 (64%)	0.058
fn3	1140..1221	28/87 (32%) 57/87 (66%)	2.4e-09
fn3	1233..1318	24/87 (28%) 59/87 (68%)	8.3e-09
fn3	1330..1408	25/86 (29%) 58/86 (67%)	6.2e-11
fn3	1420..1516	29/99 (29%) 67/99 (68%)	7.7e-09
fn3	1528..1610	19/87 (22%) 60/87 (69%)	2.1e-07
fn3	1631..1724	26/97 (27%) 66/97 (68%)	1e-05
Y_phosphatase	2008..2239	116/280 (41%) 190/280 (68%)	5.1e-109

Example 16.

The NOV16 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 16A.

5

Table 16A. NOV16 Sequence Analysis

NOV16a, CG50934-03 DNA Sequence	SEQ ID NO: 179	964 bp
	ORF Start: ATG at 67	ORF Stop: TGA at 655

GACCATCTGATTGCCGGTCTCTCCTGGTGCCCCTGAGCTCTGGGAAGACCCTCGTCCGTCCCCCTCAT
GAGCCCGGCACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCC
CTGGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGAC
GACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGG
CGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCT
CGCTCCCGTCTGCCTCCCGGGACGTGCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATT
GGACGTGGAGGTCAGGAGCAGGACCACTTGGGTGGGATGTGGAGAGATGACCCGGAATGTCGGTGCAG
GCCTGGGCTCCAGACGCGTCCTGGATGGCTTCCTGCCGCTGCCGAGACGGATGGCCACAGGCCAGGTG
GCTCAGAGCAGCAGGAATGTACTATCTCACGGCTCTGCAGGCGGAACGTCCACACTCCAGGCGTGGGC
AGGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGT
AACCAGACCTGTCGCCGCCGCATTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGA
CATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCA
GGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTAT
CCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAG
ACGCTAGCTGGG

NOV16a, CG50934-03 Protein Sequence	SEQ ID NO: 180	196 aa	MW at 21256.7kD

MSPARGVSWWASLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQ
GGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWVTGWGVIGRGGQEQDHLGGMWRDDPECRC
RPGLQTRPGWLPAAAETDGHRPGGSEQQECTISRLCRRNVHTPGVGRNYCPGPSACGRRR

NOV16b, CG50934-01 DNA Sequence	SEQ ID NO: 181	1171 bp
	ORF Start: at 187	ORF Stop: TGA at 805

ATGCTGTGGCTACTGCTCCTGACCCTCCCCTGCCTGATGGGCTCTGTGCCCAGGAACCCAGGCGAGTC
CGCCCCACCCAATGCCCCTGCTGCCCAGGACCCCCTCCTTGCCCTGCCCCGGGCTCAGAGTGCCAGCC
CTGGGGTGGGTGGGGACCATCTGATTGCCGGTCTCTCCTGGTGCCCCTGAGCTCTGGGAAGACCCTCG
TCCGTCCCCCTCATGAGCCCGGCACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCG
GCCAGGAGGCACCCCTGGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGA
GGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGC
CTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCT
CATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGTGCCCTCGGGGAAGACCTGCTGGGTGACCG
GCTGGGGTGTCATTGGACGTGGAGGTCAGGAGCAGGACCACTTGGGTGGGATGTGGAGAGATGACCCG
GAATGTCGGTGCAGGCCTGGGCTCCAGACGCGTCCTGGATGGCTTCCTGCCGCTGCCGAGACGGATGG
CCACAGGCCAGGTGGCTCAGAGCAGCAGGAATGTACTATCTCACGGCTCTGCAGGCGGAACGTCCACA
CTCCAGGCGTGGGCAGGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGA
GCAACGTCCTCTGTAACCAGACCTGTCGCCGCCGCTTTCCTTCCAACCACACTGAGCGGTTTGAGCGG
CTCATCAAGGACGACATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGG
CCCCCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCG
GCCTTCGCGGCTATCCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTC
CCGCCGTTCCCCAGACGCTAGCTGGGGTGCAGTGGGGTCTGCATGATCCAGGAGGGCCCGTCTTCCTT
GTGGACACGCCTGCT

NOV16b, CG50934-01 Protein Sequence	SEQ ID NO: 182	206 aa	MW at 22191.8kD

ALGRPSSVPLMSPARGVSWWASLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRH

PQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGGQEQDHLGG
MWRDDPECRCRPGLQTRPGWLPAAAETDGHRPGGSEQQECTISRLCRRNVHTPGVGRNYCPGPSACGR
RR

NOV16c, CG50934-02 DNA Sequence	SEQ ID NO: 183	843 bp
	ORF Start: ATG at 26	ORF Stop: TAG at 836

GAGGTGGAGGTTGCAGTAAGCCAAGATGGCGCCACTGCACTCTAGCCTGTTTCTGCTGAGCGGGACCA
TGAGCCCAAAAGTGGGCATCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGGTCAGC
CTGAGGTTCTACAGCATGAAGAAGGGTCTGTGGGAGCCCATCTGTGGGGGCTCCCTCATCCACCCAGA
GTGGGTGCTGACCGCCGCCCACTGCGTCGAGCTTGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGG
TGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAG
TACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCC
GCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGTGCCCTCGGGGAAGACCT
GCTGGGTGACCGGCTGGGGTGTCATTTGGGGACACGTTTTCCTGCTCCCGCCACCCCACCTCAGGGCA
GCGGAAGGTCCAATCATGAGGACCCGAGCTTGCGAGAGGATGTATCACAAAGGCCCCACTGCCCACGT
CACCATCATCAAGGCTGCCATGCCGTGTGCAGGGGCTGAGCGCCATCTCTCCCCACAGGGCGACAACG
GGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTC
TGCGGCCTTCGCTATCCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGT
CCCGCCGTTCCCCAGACGCTAGCTGGG

NOV16c, CG50934-02 Protein Sequence	SEQ ID NO: 184	270 aa	MW at 29993.6kD

MAPLHSSLFLLSGTMSPKVGITGGCDVSARRHPWQVSLRFYSMKKGLWEPICGGSLIHPEWVLTAAHC
VELEELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPV
SLPSASLDVPSGKTCWVTGWGVIWGHVFLLPPPHLRAAEGPIMRTRACERMYHKGPTAHVTIIKAAMP
CAGAERHLSPQGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRYPGMYTRVTSYVSWIRQYVPPFPRR

NOV16d, SNP13381559 of CG50934-03, DNA Sequence	SEQ ID NO: 185	964 bp
	ORF Start: ATG at 67	ORF Stop: TGA at 655
	SNP Pos: 338	SNP Change: T to C

GACCATCTGATTGCCGGTCTCTCCTGGTGCCCCTGAGCTCTGGGAAGACCCTCGTCCGTCCCCCTCAT
GAGCCCGGCACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCC
CTGGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGAC
GACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGG
CGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGCCT
CGCTCCCGTCTGCCTCCCGGGACGTGCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATT
GGACGTGGAGGTCAGGAGCAGGACCACTTGGGTGGGATGTGGAGAGATGACCCGGAATGTCGGTGCAG
GCCTGGGCTCCAGACGCGTCCTGGATGGCTTCCTGCCGCTGCCGAGACGGATGGCCACAGGCCAGGTG
GCTCAGAGCAGCAGGAATGTACTATCTCACGGCTCTGCAGGCGGAACGTCCACACTCCAGGCGTGGGC
AGGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGT
AACCAGACCTGTCGCCGCCGCATTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGA
CATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCA
GGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTAT
CCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAG
ACGCTAGCTGGG

NOV16d, SNP13381559 of CG50934-03, Protein Sequence	SEQ ID NO: 186	196 aa	MW at 21228.6kD
	SNP Pos: 91		SNP Change: Val to Ala

MSPARGVSWWASLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQ
GGADIALLKLEAPVPLSELIHPASLPSASRDVPSGKTCWVTGWGVIGRGGQEQDHLGGMWRDDPECRC
RPGLQTRPGWLPAAAETDGHRPGGSEQQECTISRLCRRNVHTPGVGRNYCPGPSACGRRR

NOV16e, SNP13381558 of CG50934-03, DNA Sequence	SEQ ID NO: 187	964 bp
	ORF Start: ATG at 67	ORF Stop: TGA at 655
	SNP Pos: 359	SNP Change: G to T

GACCATCTGATTGCCGGTCTCTCCTGGTGCCCCTGAGCTCTGGGAAGACCCTCGTCCGTCCCCCTCAT
GAGCCCGGCACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCC

CTGGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGAC
GACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGG
CGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCT
CGCTCCCGTCTGCCTCCCTGGACGTGCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATT
GGACGTGGAGGTCAGGAGCAGGACCACTTGGGTGGGATGTGGAGAGATGACCCGGAATGTCGGTGCAG
GCCTGGGCTCCAGACGCGTCCTGGATGGCTTCCTGCCGCTGCCGAGACGGATGGCCACAGGCCAGGTG
GCTCAGAGCAGCAGGAATGTACTATCTCACGGCTCTGCAGGCGGAACGTCCACACTCCAGGCGTGGGC
AGGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGT
AACCAGACCTGTCGCCGCCGCATTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGA
CATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCA
GGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTAT
CCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAG
ACGCTAGCTGGG

NOV16e, SNP13381558 of CG50934-03, Protein Sequence	SEQ ID NO: 188	196 aa	MW at 21213.7kD
	SNP Pos: 98		SNP Change: Arg to Leu

MSPARGVSWWASLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQ
GGADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGGQEQDHLGGMWRDDPECRC
RPGLQTRPGWLPAAAETDGHRPGGSEQQECTISRLCRRNVHTPGVGRNYCPGPSACGRRR

NOV16f, SNP13376399 of CG50934-03, DNA Sequence	SEQ ID NO: 189	964 bp
	ORF Start: ATG at 67	ORF Stop: TGA at 655
	SNP Pos: 641	SNP Change: G to A

GACCATCTGATTGCCGGTCTCTCCTGGTGCCCCTGAGCTCTGGGAAGACCCTCGTCCGTCCCCCTCAT
GAGCCCGGCACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCC
CTGGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGAC
GACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGG
CGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCT
CGCTCCCGTCTGCCTCCCGGGACGTGCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATT
GGACGTGGAGGTCAGGAGCAGGACCACTTGGGTGGGATGTGGAGAGATGACCCGGAATGTCGGTGCAG
GCCTGGGCTCCAGACGCGTCCTGGATGGCTTCCTGCCGCTGCCGAGACGGATGGCCACAGGCCAGGTG
GCTCAGAGCAGCAGGAATGTACTATCTCACGGCTCTGCAGGCGGAACGTCCACACTCCAGGCGTGGGC
AGGAACTACTGCCCTGGCCCCTCAGCTTATGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGT
AACCAGACCTGTCGCCGCCGCATTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGA
CATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCA
GGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTAT
CCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAG
ACGCTAGCTGGG

NOV16f, SNP13376399 of CG50934-03, Protein Sequence	SEQ ID NO: 190	196 aa	MW at 21316.7kD
	SNP Pos: 192		SNP Change: Cys to Tyr

MSPARGVSWWASLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQ
GGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWVTGWGVIGRGGQEQDHLGGMWRDDPECRC
RPGLQTRPGWLPAAAETDGHRPGGSEQQECTISRLCRRNVHTPGVGRNYCPGPSAYGRRR

NOV16g, SNP13378301 of CG50934-03, DNA Sequence	SEQ ID NO: 191	964 bp
	ORF Start: ATG at 67	ORF Stop: TGA at 655
	SNP Pos: 868	SNP Change: G to A

GACCATCTGATTGCCGGTCTCTCCTGGTGCCCCTGAGCTCTGGGAAGACCCTCGTCCGTCCCCCTCAT
GAGCCCGGCACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCC
CTGGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGAC
GACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGG
CGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCT
CGCTCCCGTCTGCCTCCCGGGACGTGCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATT
GGACGTGGAGGTCAGGAGCAGGACCACTTGGGTGGGATGTGGAGAGATGACCCGGAATGTCGGTGCAG
GCCTGGGCTCCAGACGCGTCCTGGATGGCTTCCTGCCGCTGCCGAGACGGATGGCCACAGGCCAGGTG
GCTCAGAGCAGCAGGAATGTACTATCTCACGGCTCTGCAGGCGGAACGTCCACACTCCAGGCGTGGGC

AGGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGG**TGA**AGGTCAGGAGCAACGTCCTCTGT
AACCAGACCTGTCGCCGCCGCATTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGA
CATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCA
GGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTACGGCCTTCGCGGCTAT
CCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAG
ACGCTAGCTGGG

NOV16g, SNP13378301 of CG50934-03, Protein Sequence	SEQ ID NO: 192	196 aa	MW at 21256.7kD
			SNP Change: no change

MSPARGVSWWASLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQ
GGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWVTGWGVIGRGGQEQDHLGGMWRDDPECRC
RPGLQTRPGWLPAAAETDGHRPGGSEQQECTISRLCRRNVHTPGVGRNYCPGPSACGRRR

NOV16h, SNP13378300 of CG50934-03, DNA Sequence	SEQ ID NO: 193	964 bp
	ORF Start: ATG at 67	ORF Stop: TGA at 655
	SNP Pos: 907	SNP Change: C to T

GACCATCTGATTGCCGGTCTCTCCTGGTGCCCCTGAGCTCTGGGAAGACCCTCGTCCGTCCCCCTC**AT**
GAGCCCGGCACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCC
CTGGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGAC
GACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGG
CGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCT
CGCTCCCGTCTGCCTCCCGGGACGTGCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATT
GGACGTGGAGGTCAGGAGCAGGACCACTTGGGTGGGATGTGGAGAGATGACCCGGAATGTCGGTGCAG
GCCTGGGCTCCAGACGCGTCCTGGATGGCTTCCTGCCGCTGCCGAGACGGATGGCCACAGGCCAGGTG
GCTCAGAGCAGCAGGAATGTACTATCTCACGGCTCTGCAGGCGGAACGTCCACACTCCAGGCGTGGGC
AGGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGG**TGA**AGGTCAGGAGCAACGTCCTCTGT
AACCAGACCTGTCGCCGCCGCATTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGA
CATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCA
GGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTAT
CCCGGCATGTACACCCGCGTGATGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAG
ACGCTAGCTGGG

NOV16h, SNP13378300 of CG50934-03, Protein Sequence	SEQ ID NO: 194	196 aa	MW at 21256.7kD
			SNP Change: no change

MSPARGVSWWASLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQ
GGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWVTGWGVIGRGGQEQDHLGGMWRDDPECRC
RPGLQTRPGWLPAAAETDGHRPGGSEQQECTISRLCRRNVHTPGVGRNYCPGPSACGRRR

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 16B.

Table 16B. Comparison of the NOV16 protein sequences.

```
NOV16a  --------------MSPARGVS-----------WWASLGAATSRP-------GGTPGR--
NOV16b  ----ALGRPSSVPLMSPARGVS-----------WWASLGAATSRP-------GGTPGR--
NOV16c  MAPLHSSLFLLSGTMSPKVGITGGCDVSARRHPWQVSLRFYSMKKGLWEPICGGSLIHPE

NOV16a  -----------EELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIAL
NOV16b  -----------EELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIAL
NOV16c  WVLTAAHCVELEELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIAL

NOV16a  LKLEAPVPLSELIHPVSLPSASRDVPSGKTCWVTGWGVIG------RGGQEQDHLGGMWR
NOV16b  LKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIG------RGGQEQDHLGGMWR
NOV16c  LKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIWGHVFLLPPPHLRAAEGPIMR
```

```
NOV16a  D-DPECRCRPGLQTRPGWLPAAAETDGHRPGGSEQQECTISRLCRRNVHTPGVGRNYCPG
NOV16b  D-DPECRCRPGLQTRPGWLPAAAETDGHRPGGSEQQECTISRLCRRNVHTPGVGRNYCPG
NOV16c  TRACERMYHKGPTAHVTIIKAAMPCAGAERHLSPQGDNGGPLLCRRN--CTWVQVEVVSW

NOV16a  PSACGRRR------------------------
NOV16b  PSACGRRR------------------------
NOV16c  GKLCGLRYPGMYTRVTSYVSWIRQYVPPFPRR

NOV16a  (SEQ ID NO:  180)
NOV16b  (SEQ ID NO:  182)
NOV16c  (SEQ ID NO:  184)
```

Further analysis of the NOV16a protein yielded the following properties shown in Table 16C.

Table 16C. Protein Sequence Properties NOV16a

SignalP analysis:	Cleavage site between residues 19 and 20

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 5;  pos.chg 1;  neg.chg 0
      H-region:  length 13;  peak value   6.79
      PSG score:   2.39

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.86
      possible cleavage site: between 18 and 19

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  4.88 (at 73)
      ALOM score:   4.88  (number of TMSs: 0)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 6
      Charge difference: -3.0   C(-1.0) - N( 2.0)
      N >= C: N-terminal side will be inside

MITDISC: discrimination of mitochondrial targeting seq
      R content:       3       Hyd Moment(75):  7.01
      Hyd Moment(95):  6.08    G content:       5
      D/E content:     1       S/T content:     6
      Score: -2.71

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 29  SRP|GG

NUCDISC: discrimination of nuclear localization signals
```

```
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  12.2%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: SPAR

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 70.6

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: extracellular, including cell wall
       33.3 %: mitochondrial
```

```
      11.1 %: cytoplasmic

>> prediction for CG50934-03 is exc (k=9)
```

A search of the NOV16a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 16D.

Table 16D. Geneseq Results for NOV16a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE08591	Human NOV12 protein - Homo sapiens, 220 aa. [WO200161009-A2, 23-AUG-2001]	12..117 1..106	106/106 (100%) 106/106 (100%)	2e-56
AAU82736	Amino acid sequence of novel human protease #35 - Homo sapiens, 948 aa. [WO200200860-A2, 03-JAN-2002]	5..117 184..302	97/121 (80%) 101/121 (83%)	3e-47
AAE14347	Human protease PRTS-12 protein - Homo sapiens, 262 aa. [WO200183775-A2, 08-NOV-2001]	5..117 30..148	97/121 (80%) 101/121 (83%)	3e-47
AAE08587	Human NOV8 protein - Homo sapiens, 290 aa. [WO200161009-A2, 23-AUG-2001]	25..117 84..176	91/93 (97%) 91/93 (97%)	6e-47
AAE08590	Human NOV11 protein - Homo sapiens, 285 aa. [WO200161009-A2, 23-AUG-2001]	27..117 81..171	90/91 (98%) 90/91 (98%)	2e-46

In a BLAST search of public sequence databases, the NOV16a protein was found to have homology to the proteins shown in the BLASTP data in Table 16E.

Table 16E. Public BLASTP Results for NOV16a				
Protein Accession Number	**Protein/Organism/Length**	**NOV16a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAP21675	Mast cell protease-11 - Mus musculus (Mouse), 318 aa.	10..114 75..173	63/105 (60%) 77/105 (73%)	3e-27
AAA30855	Mastin precursor - Canis sp, 251 aa (fragment).	27..114 53..140	55/88 (62%) 61/88 (68%)	5e-22
P19236	Mastocytoma protease precursor (EC 3.4.21.-) - Canis familiaris (Dog), 269 aa.	27..114 71..158	55/88 (62%) 61/88 (68%)	5e-22
Q8SQ44	Tryptase precursor - Sus scrofa (Pig), 277 aa.	35..114 90..167	49/80 (61%) 60/80 (74%)	1e-18
I48685	mast cell proteinase 6 (EC 3.4.21.-) precursor - mouse, 230 aa.	34..114 87..164	44/81 (54%) 54/81 (66%)	2e-16

PFam analysis indicates that the NOV16a protein contains the domains shown in the Table 16F.

Table 16F. Domain Analysis of NOV16a			
Pfam Domain	**NOV16a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
trypsin	19..172	52/266 (20%) 108/266 (41%)	0.00069

Example 17.

The NOV17 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 17A.

Table 17A. NOV17 Sequence Analysis

NOV17a, CG51213-04 DNA Sequence	SEQ ID NO: 195	2804 bp
	ORF Start: ATG at 71	ORF Stop: TAG at 2636

TGGCCAGCCAGGCCTGAAGCGATCGGTCAGCCGAGAGCGCTACGTGGAGACCCTGGTGGTGGCTGACA
AG**ATG**ATGGTGGCCTATCACGGGCGCCGGGATGTGGAGCAGTATGTCCTGGCCATCATGAACATTCAG
GTTGCCAAACTTTTCCAGGACTCGAGTCTGGGAAGCACCGTTAACATCCTCGTAACTCGCCTCATCCT
GCTCACGGAGGACCAGCCCACTCTGGAGATCACCCACCATGCCGGGAAGTCCCTGGACAGCTTCTGTA
AGTGGCAGAAATCCATCGTGAACCACAGCGGCCATGGCAATGCCATTCCAGAGAACGGTGTGGCTAAC
CATGACACAGCAGTGCTCATCACACGCTATGACATCTGCATCTACAAGAACAAACCCTGCGGCACACT
AGGCCTGGCCCCGGTGGGCGGAATGTGTGAGCGCGAGAGAAGCTGCAGCGTCAATGAGGACATTGGCC
TGGCCACAGCGTTCACCATTGCCCACGAGATCGGGCACACATTCGGCATGAACCATGACGGCGTGGGA
AACAGCTGTGGGGCCCGTGGTCAGGACCCAGCCAAGCTCATGGCTGCCCACATTACCATGAAGACCAA
CCCATTCGTGTGGTCATCCTGCAGCCGTGACTACATCACCAGCTTTCTAGACTCGGGCCTGGGGCTCT
GCCTGAACAACCGGCCCCCCAGACAGGACTTTGTGTACCCGACAGTGGCACCGGGCCAAGCCTACGAT
GCAGATGAGCAATGCCGCTTTCAGCATGGAGTCAAATCGCGTCAGTGTAAATACGGGGAGGTCTGCAG
CGAGCTGTGGTGTCTGAGCAAGAGCAACCGGTGCATCACCAACAGCATCCCGGCCGCCGAGGGCACGC
TGTGCCAGACGCACACCATCGACAAGGGGTGGTGCTACAAACGGGTCTGTGTCCCCTTTGGGTCGCGC
CCAGAGGGTGTGGACGGAGCCTGGGGGCCGTGGACTCCATGGGGCGACTGCAGCCGGACCTGTGGCGG
CGGCGTGTCCTCTTCTAGCCGTCACTGCGACAGCCCCAGGCCAACCATCGGGGGCAAGTACTGTCTGA
GTGAGAGAAGGCGGCACCGCTCCTGCAACACGGATGACTGTCCCCCTGGCTCCCAGGACTTCAGAGAA
GTGCAGTGTTCTGAATTTGACAGCATCCCTTTCCGTGGGAAATTCTACAAGTGGAAAACGTACCGGGG
AGGGGGCGTGAAGGCCTGCTCGCTCACGAGCCTAGCGGAAGGCTTCAACTTCTACACGGAGAGGGCGG
CAGCCGTGGTGGACGGGACACCCTGCCGTCCAGACACGGTGGACATTTGCGTCAGTGGCGAATGCAAG
CACGTGGGCTGCGACCGAGTCCTGGGCTCCGACCTGCGGGAGGACAAGTGCCGAGTGTGTGGCGGTGA
CGGCAGTGCCTGCGAGACCATCGAGGGCGTCTTCAGCCCAGCCTCACCTGGGGCCGGGTACGAGGATG
TCGTCTGGATTCCCAAAGGCTCCGTCCACATCTTCATCCAGGATCTGAACCTCTCTCTCAGTCACTTG
GCCCTGAAGGGAGACCAGGAGTCCCTGCTGCTGGAGGGGCTGCCTGGGACCCCCCAGCCCCACCGTCT
GCCTCTAGCTGGGACCACCTTTCAACTGCGACAGGGGCCAGACCAGGTCCAGAGCCTCGAAGCCCTGG
GACCGATTAATGCATCTCTCATCGTCATGGTGCTGGCCCGGACCGAGCTGCCTGCCCTCCGCTACCGC
TTCAATGCCCCCATCGCCCGTGACTCGCTGCCCCCCTACTCCTGGCACTATGCGCCCTGGACCAAGTG
CTCGGCCCAGTGTGCAGGCGGTAGCCAGGTGCAGGCGGTGGAGTGCCGCAACCAGCTGGACAGCTCCG
CGGTCGCCCCCCACTACTGCAGTGCCCACAGCAAGCTGCCCAAAAGGCAGCGCGCCTGCAACACGGAG
CCTTGCCCTCCAGACTGGGTTGTAGGGAACTGGTCGCTCTGCAGCCGCAGCTGCGATGCAGGCGTGCG
CAGTCGCTCGGTCGTGTGCCAGCGCCGCGTCTCTGCCGCGGAGGAGAAGGCGCTGGACGACAGCGCAT
GCCCGCAGCCGCGCCCACCTGTACTGGAGGCCTGCCACGGCCCCACTTGCCCTCCGGAGTGGGCGGCC
CTCGACTGGTCTGAGTGCACCCCCAGCTGCGGGCCGGGCCTCCGCCACCGCGTGGTCCTTTGCAAGAG
CGCAGACCACCGCGCCACGCTGCCCCCGGCGCACTGCTCACCCGCCGCCAAGCCACCGGCCACCATGC
GCTGCAACTTGCGCCGCTGCCCCCCGGCCCGCTGGGTGGCTGGCGAGTGGGGTGAGTGCTCTGCACAG
TGCGGCGTCGGGCAGCGGCAGCGCTCGGTGCGCTGCACCAGCCACACGGGCCAGGCGTCGCACGAGTG
CACGGAGGCCCTGCGGCCGCCCACCACGCAGCAGTGTGAGGCCAAGTGCGACAGCCCAACCCCCGGGG
ACGGCCCTGAAGAGTGCAAGGATGTGAACAAGGTCGCCTACTGCCCCCTGGTGCTCAAATTTCAGTTC
TGCAGCCGAGCCTACTTCCGCCAGATGTGCTGCAAAACCTGCCAGGGCCAC**TAG**GGGGCGCGCGGCAC
CCGGAGCCACAGCTGGCGGGGTCTCCGCCGCCAGCCCTGCAGCGGGCCGGCCAAAGGGGGCCCCGGGG
GGGCGGGAACTGGGAGGGAAGGGTGAGACGGAGCCGGAAGTTATTTATTGGGAACCCCTGCAGGGCCC
TGGCTGGGGGGATGGA

NOV17a, CG51213-04 Protein Sequence	SEQ ID NO: 196	855 aa	MW at 93285.7kD

MMVAYHGRRDVEQYVLAIMNIQVAKLFQDSSLGSTVNILVTRLILLTEDQPTLEITHHAGKSLDSFCK
WQKSIVNHSGHGNAIPENGVANHDTAVLITRYDICIYKNKPCGTLGLAPVGGMCERERSCSVNEDIGL
ATAFTIAHEIGHTFGMNHDGVGNSCGARGQDPAKLMAAHITMKTNPFVWSSCSRDYITSFLDSGLGLC
LNNRPPRQDFVYPTVAPGQAYDADEQCRFQHGVKSRQCKYGEVCSELWCLSKSNRCITNSIPAAEGTL
CQTHTIDKGWCYKRVCVPFGSRPEGVDGAWGPWTPWGDCSRTCGGGVSSSSRHCDSPRPTIGGKYCLS
ERRRHRSCNTDDCPPGSQDFREVQCSEFDSIPFRGKFYKWKTYRGGGVKACSLTSLAEGFNFYTERAA
AVVDGTPCRPDTVDICVSGECKHVGCDRVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYEDV
VWIPKGSVHIFIQDLNLSLSHLALKGDQESLLLEGLPGTPQPHRLPLAGTTFQLRQGPDQVQSLEALG

PINASLIVMVLARTELPALRYRFNAPIARDSLPPYSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSA
VAPHYCSAHSKLPKRQRACNTEPCPPDWVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEKALDDSAC
PQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLPPAHCSPAAKPPATMR
CNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTGQASHECTEALRPPTTQQCEAKCDSPTPGD
GPEECKDVNKVAYCPLVLKFQFCSRAYFRQMCCKTCQGH

NOV17b, CG51213-01 DNA Sequence	SEQ ID NO: 197	2266 bp
	ORF Start: ATG at 589	ORF Stop: TAG at 2158

CTCGGGCCTGGGGCTCTGCCTGAACAACCGGCCCCCCAGACAGGACTTTGTGTACCCGACAGTGGCAC
CGGGCCAAGCCTACGATGCAGATGAGCAATGCCGCTTTCAGCATGGAGTCAAATCGCGTCAGTTGGTG
CTACAAACGGGTCTGTGTCCCCTTTGGGTCGCGCCCAGAGGGTGTGGACGGAGCCTGGGGGCCGTGGA
CTCCATGGGGCGACTGCAGCCGGACCTGTGGCGGCGGCGTGTCCTCTTCTAGCCGTCACTGCGACAGC
CCCAGGCCAACCATCGGGGGCAAGTACTGTCTGGGTGAGAGAAGGCGGCACCGCTCCTGCAACACGGA
TGACTGTCCCCCTGGCTCCCAGGACTTCAGAGAAGTGCAGTGTTCTGAATTTGACAGCATCCCTTTCC
GTGGGAAATTCTACAAGTGGAAAACGTACCGGGGAAGGTGAGTGTGGGACTCCAAAGGCTGTGGGGCC
GTGAAGGGCAGCCGTGGGAGTGTCCAGCAGCAGGTGGATGAATGCAGCATCCCGGGGTCTGCCATGAG
CCCTGTCCCCACCCGGGGAGACAGAGTACCTGGGATACGGTACC**ATG**GGGGTTCAACGTGACGCTGGG
AGCCCCCACTCCCTCTGCCCAAGCTGCCCTTCCTCTTGGGTCTGGGGTCTGTCCCTCTTGGCCTCACT
CCCCCAGGGAGCAAGCAAAGAGTTCCGGGGTGGCCTGGCCCGTGGTGTGACGGGGCCGTGCCCCCCAG
GGGGCGTGAAGGCCTGCTCGCTCACGTGCCTAGCGGAAGGCTTCAACTTCTACACGGAGAGGGCGGCA
GCCGTGGTGGACGGGACACCCTGCCGTCCAGACACGGTGGACATTTGCGTCAGTGGCGAATGCAAGCA
CGTGGGCTGCGACCGAGTCCTGGGCTCCGACCTGCGGGAGGACAAGTGCCGAGTGTGTGGCGGTGACG
GCAGTGCCTGCGAGACCATCGAGGGCGTCTTCAGCCCAGCCTCACCTGGGGCCGGGTACGAGGATGTC
GTCTGGATTCCCAAAGGCTCCGTCCACATCTTCATCCAGGATCTGAACCTCTCTCTCAGTCACTTGGC
CCTGAAGGGAGACCAGGAGTCCCTGCTGCTGGAGGGGCTGCCTGGGACCCCCCAGCCCCACCGTCTGC
CTCTAGCTGGGACCACCTTTCAACTGCGACAGGGGCCAGACCAGGTCCAGAGCCTCGAAGCCCTGGGA
CCGATTAATGCATCTCTCATCGTCATGGTGCTGGCCCGGACCGAGCTGCCTGCCCTCCGCTACCGCTT
CAATGCCCCCATCGCCCGTGACTCGCTGCCCCCCTACTCCTGGCACTATGCGCCCTGGACCAAGTGCT
CGGCCCAGTGTGCAGGCGGTAGCCAGGTGCAGGCGGTGGAGTGCCGCAACCAGCTGGACAGCTCCGCG
GTCGCCCCCCACTACTGCAGTGCCCACAGCAAGCTGCCCAAAAGGCAGCGCGCCTGCAACACGGAGCC
TTGCCCTCCAGACTGGGTTGTAGGGAACTGGTCGCTCTGCAGCCGCAGCTGCGATGCAGGCGTGCGCA
GCCGCTCGGTCGTGTGCCAGCGCCGCGTCTCTGCCGCGGAGGAGAAGGCGCTGGACGACAGCGCATGC
CCGCAGCCGCGCCCACCTGTACTGGAGGCCTGCCACGGCCCCACTTGCCCTCCGGAGTGGGCGGCCCT
CGACTGGTCTGAGTGCACCCCCAGCTGCGGGCCGGGCCTCCGCCACCGCGTGGTCCTTTGCAAGAGCG
CAGACCACCGCGCCACGCTGCCCCCGGCGCACTGCTCACCCGCCGCCAAGCCACCGGCCACCATGCGC
TGCAACTTGCGCCGCTGCCCCCCGGCCCGCTGGGTGGCTGGCGAGTGGGGTGAGTGCTCTGCACAGTG
CGGCGTCGGGCAGCGGCAGCGCTCGGTGCGCTGCACCAGCCACACGGGCCAGGCGTCGCACGAGTGCA
CGGAGGCCCTGCGGCCGCCCACCACGCAGCAGTGTGAGGCCAAGTGCGACAGCCCAACCCCCGGGGAC
GGCCCTGAAGAGTGCAAGGATGTGAACAAGGTCGCCTACTGCCCCCTGGTGCTCAAATTTCAGTTCTG
CAGCCGAGCCTACTTCCGCCAGATGTGCTGCAAAACCTGCCAGGGCCAC**TAG**GGGGCGCGCGGCACCC
GGAGCCACAGCTGGCGGGGTCTCCGCCGCCAGCCCTGCAGCTGGGCCGGCCAGAGGGGGCCCCGGGGG
GGCGGGAACTGGGAGGGAAGGG

NOV17b, CG51213-01 Protein Sequence	SEQ ID NO: 198	523 aa	MW at 56126.2kD

MGVQRDAGSPHSLCPSCPSSWVWGLSLLASLPQGASKEFRGGLARGVTGPCPPGGVKACSLTCLAEGF
NFYTERAAAVVDGTPCRPDTVDICVSGECKHVGCDRVLGSDLREDKCRVCGGDGSACETIEGVFSPAS
PGAGYEDVVWIPKGSVHIFIQDLNLSLSHLALKGDQESLLLEGLPGTPQPHRLPLAGTTFQLRQGPDQ
VQSLEALGPINASLIVMVLARTELPALRYRFNAPIARDSLPPYSWHYAPWTKCSAQCAGGSQVQAVEC
RNQLDSSAVAPHYCSAHSKLPKRQRACNTEPCPPDWVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEE
KALDDSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLPPAHCSPA
AKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTGQASHECTEALRPPTTQQCEAK
CDSPTPGDGPEECKDVNKVAYCPLVLKFQFCSRAYFRQMCCKTCQGH

NOV17c, 306345264 DNA Sequence	SEQ ID NO: 199	706 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCTGGCACTATGCGCCCTGGACCAAGTGCTCGGCCCAGTGTGCAGGCGGTAGCCAGGTGC
AGGCGGTGGAGTGCCGCAACCAGCTGGACAGCTCCGCGGTCGCCCCCCACTACTGCAGTGCCCACAGC

AAGCTGCCCAAAAGGCAGCGCGCCTGCAACACGGAGCCTTGCCCTCCAGACTGGGTTGTAGGGAACTG
GTCGCTCTGCAGCCGCAGCTGCGATGCAGGCGTGCGCAGCCGCTCGGTCGTGTGCCAGCGCCGCGTCT
CTGCCGCGGAGGAGAAGGCGCTGGACGACAGCGCATGCCCGCAGCCGCGCCCACCTGTACTGGAGGCC
TGCCACGGCCCCACTTGCCCTCCGGAGTGGGCGGCCCTCGACTGGTCTGAGTGCACCCCCAGCTGCGG
GCCGGGCCTCCGCCACCGCGTGGTCCTTTGCAAGAGCGCAGACCACCGCGCCACGCTGCCCCCGGCGC
ACTGCTCACCCGCCGCCAAGCCACCGGCCACCATGCGCTGCAACTTGCGCCGCTGCCCCCCGGCCCGC
TGGGTGGCTGGCGAGTGGGGTGAGTGCTCTGCACAGTGCGGCGTCGGGCAGCGGCAGCGCTCGGTGCG
CTGCACCAGCCACACGGGCCAGGCGTCGCACGAGTGCACGGAGGCCCTGCGGCCGCCCACCACGCAGC
AGTGTGAGGCCAAGTGCCTCGAGGGC

NOV17c, 306345264 Protein Sequence	SEQ ID NO: 200	235 aa	MW at 25395.4kD

TGSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSAVAPHYCSAHSKLPKRQRACNTEPCPPDWVVGNW
SLCSRSCDAGVRSRSVVCQRRVSAAEEKALDDSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCG
PGLRHRVVLCKSADHRATLPPAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVR
CTSHTGQASHECTEALRPPTTQQCEAKCLEG

NOV17d, CG51213-02 DNA Sequence	SEQ ID NO: 201	3199 bp
	ORF Start: at 1297	ORF Stop: at 3199

TAAAGGGTTCAGCCTGGTGCCTGGTCCAGAGATAGTGGTGGTCATTGTTACCCCATAATGGCATTGGT
GCAAGTCCTTTCTTATCTATCCTGTCACGTGCCTCATAGCCATTTATATAGGCAAGACAGGCATTAGG
CTGCCCATCTTGTAGATGAGTAAACTGAGGCCCAGAGAGGGGAAATATATTGCAAGTTGGTAGCAGAA
TTGAGGTCTCTGCACAACTCAAATATGCCACAGTGCCTCCTTGTGGAGAGGAGGACAAAAGCAGAGCT
GAAATCATTATCTTGAAGAGGTGTCAGAAGTGGGATTGCGACAGGACTGATGTGATATTTTTAGATAT
GGCCAAGAGGACACAGTCTGAGTTTTTAGCTGAGAAATGTCCTCTATAAGGCAGAAGGCAGAGATTCT
AGAGGACCTTTGAGGGAGAATGTATTTGAGAACAACTCTTCCAGCTTCTTACATATGTACAGGTATCT
CTCAGGGGCTGACCTAGGAAGGGTCCTTTCCTGTGGCCATTGATCGATCCAGTCCCACATCTGGAAAG
CTTACAAGAATTGGGTTCAAAGCGGGGATTACACTTGATAATTACAGAAGGACCACCTACTTCTTAGA
GGAAAGACGCTGGGAGGTTGCTTAGGATGTGGGCCAAGAGGGTCAGAGAGGACCACCTACTTTTTAGA
GGAAAGACGCTGGGAGGTTGCTTAGGATGTGGGCCAAGAGGGTCAGAGATTTTGCTTCACCTGAACTC
ACTGGGGCTTCTCCAGGGATATTAACCTGGACTTTAAGAGTCAGAGTGAGTCCCTGGGACTAGTTCAG
CCCATCCAGGATTCAGACGGGAAGAAGGTGGGGCTGATTTTTCACCTGGAGAAAGAGAGGCATGTCCC
ACACAGACCTAACTCGGCATTGTCCCCTCCCAAACTCCCACCCCTCCACATAGCTTAAAAGTGTTGGG
GGCTTCTCCAGTTTAGATGGGGGAACAAAGAGAACCAACAGCTGGAAAAAACTAGAGATGAGGCCGTT
GGCCTAGTCATCATCCAGGCCGATTTCTCAGAACCACCACTTTCTCTTCGGCTACTTTGCCCATCCCA
TAAAAGAACCCCAAATCCTTCCTGTTCATTCCTCAGCAGTTCCCACGTTTCCTTCCAGAAACTCAGAA
GGCACCAGGAACTGAATTGCAAAGTTCGTTAGAGCACAGACTCTGAATTAAAGAGCTGGGTTAAACTC
CAGGCTATTCCCTTAGTAGCTGTGTGACCTTACCTGTCTGAAGCTTGGTTTTCTCCCAGTAAGATGGG
GTAGTACTGCCTAAAGAGGTATATGGCATGTATAAAGTGCTCCATAAATGGAGCTTATTGGGAGAGTA
TAAGTCACAGGCCATGCCCCGCAAGGGGATGCACGAAGACCCACCGCGAGCCAGGAAGGGAGCACCGG
GCTCTCTGCTCTGGGACCGGCAGTGAGCCGGACATCTGGGTCCTCCCAAGCCGGGCGGGCTGCCCCAG
GGAGGAAGGGAGGGGGGCGAGCCTGAGCGGGCACCTCGGCCCGCAGGAGGTCTGCAGCGAGCTGTGGT
GTCTGAGCAAGAGCAACCGGTGCATCACCAACAGCATCCCGGCCGCCGAGGGCACGCTGTGCCAGACG
CACACCATCGACAAGGGGTGGTGCTACAAACGGGTCTGTGTCCCCTTTGGGTCGCGCCCAGAGGGTGT
GGACGGAGCCTGGGGGCCGTGGACTCCATGGGGCGACTGCAGCCGGACCTGTGGCGGCGGCGTGTCCT
CTTCTAGCCGTCACTGCGACAGCCCCAGGCCAACCATCGGGGGCAAGTACTGTCTGGGTGAGAGAAGG
CGGCACCGCTCCTGCAACACGGATGACTGTCCCCCTGGCTCCCAGGACTTCAGAGAAGTGCAGTGTTC
TGAATTTGACAGCATCCCTTTCCGTGGGAAATTCTACAAGTGGAAAACGTACCGGGGAGGGGGCGTGA
AGGCCTGCTCGCTCACGTGCCTAGCGGAAGGCTTCAACTTCTACACGGAGAGGGCGGCAGCCGTGGTG
GACGGGACACCCTGCCGTCCAGACACGGTGGACATTTGCGTCAGTGGCGAATGCAAGCACGTGGGCTG
CGACCGAGTCCTGGGCTCCGACCTGCGGGAGGACAAGTGCCGAGTGTGTGGCGGTGACGGCAGTGCCT
GCGAGACCATCGAGGGCGTCTTCAGCCCAGCCTCACCTGGGGCCGGGTACGAGGATGTCGTCTGGATT
CCCAAAGGCTCCGTCCACATCTTCATCCAGGATCTGAACCTCTCTCTCAGTCACTTGGCCCTGAAGGG
AGACCAGGAGTCCCTGCTGCTGGAGGGGCTGCCCGGGACCCCCCAGCCCCACCGTCTGCCTCTAGCTG
GGACCACCTTTCAACTGCGACAGGGGCCAGACCAGGTCCAGAGCCTCGAAGCCCTGGGACCGATTAAT
GCATCTCTCATCGTCATGGTGCTGGCCCGGACCGAGCTGCCTGCCCTCCGCTACCGCTTCAATGCCCC
CATCGCCCGTGACTCGCTGCCCCCCTACTCCTGGCACTATGCGCCCTGGACCAAGTGCTCGGCCCAGT
GTGCAGGCGGTAGCCAGGTGCAGGCGGTGGAGTGCCGCAACCAGCTGGACAGCTCCGCGGTCGCCCCC
CACTACTGCAGTGCCCACAGCAAGCTGCCCAAAAGGCAGCGCGCCTGCAACACGGAGCCTTGCCCTCC

AGACTGGGTTGTAGGGAACTGGTCGCTCTGCAGCCGCAGCTGCGATGCAGGCGTGCGCAGCCGCTCGG
TCGTGTGCCAGCGCCGCGTCTCTGCCGCGGAGGAGAAGGCGCTGGACGACAGCGCATGCCCGCAGCCG
CGCCCACCTGTACTGGAGGCCTGCCACGGCCCCACTTGCCCTCCGGAGTGGGCGGCCCTCGACTGGTC
TGAGTGCACCCCCAGCTGCGGGCCGGGCCTCCGCCACCGCGTGGTCCTTTGCAAGAGCGCAGACCACC
GCGCCACGCTGCCCCCGGCGCACTGCTCACCCGCCGCCAAGCCACCGGCCACCATGCGCTGCAACTTG
CGCCGCTGCCCCCCGGCCCGCTGGGTGGCTGGCGAGTGGGGTGAGTGCTCTGCACAGTGCGGCGTCGG
GCAGCGGCAGCGCTCGGTGCGCTGCACCAGCCACACGGGCCAGGCGTCGCACGAGTGCACGGAGGCCC
TGC

NOV17d, CG51213-02 Protein Sequence	SEQ ID NO: 202	634 aa	MW at 68853.1kD

YCLKRYMACIKCSINGAYWESISHRPCPARGCTKTHREPGREHRALCSGTGSEPDIWVLPSRAGCPRE
EGRGASLSGHLGPQEVCSELWCLSKSNRCITNSIPAAEGTLCQTHTIDKGWCYKRVCVPFGSRPEGVD
GAWGPWTPWGDCSRTCGGGVSSSSRHCDSPRPTIGGKYCLGERRRHRSCNTDDCPPGSQDFREVQCSE
FDSIPFRGKFYKWKTYRGGGVKACSLTCLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVGCD
RVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNLSLSHLALKGD
QESLLLEGLPGTPQPHRLPLAGTTFQLRQGPDQVQSLEALGPINASLIVMVLARTELPALRYRFNAPI
ARDSLPPYSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSAVAPHYCSAHSKLPKRQRACNTEPCPPD
WVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEKALDDSACPQPRPPVLEACHGPTCPPEWAALDWSE
CTPSCGPGLRHRVVLCKSADHRATLPPAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQ
RQRSVRCTSHTGQASHECTEAL

NOV17e, CG51213-03 DNA Sequence	SEQ ID NO: 203	3700 bp
	ORF Start: at 1798	ORF Stop: at 3700

CTGACATTCCACCCTTGACACCCCCCAACATCCTAACTTAGCTGGTAACTGCAGCACCCTCTAAGGAA
TTCCTAAAGAATTCTGAAGCTACTCCTCAACATCTGCTGTGACCCAGGTATCCTAACAATGATCATGG
TGTCTGACATTTACTGAGCTCTCACTATGGGCTAAGCATGTGCTGTGTGTCACCATCTAAACTCCTGA
CAATCCTGCTAGCCCCCACGTTACAGAGGAAGGGACTGAGCCATAGCATAGGGAGGATGACTTGTCCA
AGGCCACAGTTTGAGACCATGACAGAGCTGGGATTTAAATCCAGGTCTCTCATGACTCTCTAAATTTT
ACAAAGGGGCAGGGGAGGGGAGGAGCTGTCAAAATATCAAGCTTGGGCTGGCACTGGCTATATGTTGA
ATTGAGCCTTCCTTTTAGTTTTTGAAGGAACATCTTTCAGGCCATCTTGGCAAAGGGGGATTTATTTA
CTAAATGTGAACTGGTTAATATATGTAAAGGGTTCAGCCTGGTGCCTGGTCCAGAGATAGTGGTGGTC
ATTGTTACCCCATAATGGCATTGGTGCAAGTCCTTTCTTATCTATCCTGTCACGTGCCTCATAGCCAT
TTATATAGGCAAGACAGGCATTAGGCTGCCCATCTTGTAGATGAGTAAACTGAGGCCCAGAGAGGGGA
AATATATTGCAAGTTGGTAGCAGAATTGAGGTCTCTGCACAACTCAAATATGCCACAGTGCCTCCTTG
TGGAGAGGAGGACAAAAGCAGAGCTGAAATCATTATCTTGAAGAGGTGTCAGAAGTGGGATTGCGACA
GGACTGATGTGATATTTTTAGATATGGCCAAGAGGACACAGTCTGAGTTTTTAGCTGAGAAATGTCCT
CTATAAGGCAGAAGGCAGAGATTCTAGAGGACCTTTGAGGGAGAATGTATTTGAGAACAACTCTTCCA
GCTTCTTACATATGTACAGGTATCTCTCAGGGGCTGACCTAGGAAGGGTCCTTTCCTGTGGCCATTGA
TCGATCCAGTCCCACATCTGGAAAGCTTACAAGAATTGGGTTCAAAGCGGGGATTACACTTGATAATT
ACAGAAGGACCACCTACTTCTTAGAGGAAAGACGCTGGGAGGTTGCTTAGGATGTGGGCCAAGAGGGT
CAGAGAGGACCACCTACTTTTTAGAGGAAAGACGCTGGGAGGTTGCTTAGGATGTGGGCCAAGAGGGT
CAGAGATTTTGCTTCACCTGAACTCACTGGGGCTTCTCCAGGGATATTAACCTGGACTTTAAGAGTCA
GAGTGAGTCCCTGGGACTAGTTCAGCCCATCCAGGATTCAGACGGGAAGAAGGTGGGGCTGATTTTTC
ACCTGGAGAAAGAGAGGCATGTCCCACACAGACCTAACTCGGCATTGTCCCCTCCCAAACTCCCACCC
CTCCACATAGCTTAAAAGTGTTGGGGGCTTCTCCAGTTTAGATGGGGGAACAAAGAGAACCAACAGCT
GGAAAAAACTAGAGATGAGGCCGTTGGCCTAGTCATCATCCAGGCCGATTTCTCAGAACCACCACTTT
CTCTTCGGCTACTTTGCCCATCCCATAAAAGAACCCCAAATCCTTCCTGTTCATTCCTCAGCAGTTCC
CACGTTTCCTTCCAGAAACTCAGAAGGCACCAGGAACTGAATTGCAAAGTTCGTTAGAGCACAGACTC
TGAATTAAAGAGCTGGGTTAAACTCCAGGCTATTCCCTTAGTAGCTGTGTGACCTTACCTGTCTGAAG
CTTGGTTTTCTCCCAGTAAGATGGGGTAGTACTGCCTAAAGAGGTATATGGCATGTATAAAGTGCTCC
ATAAATGGAGCTTATTGGGAGAGTATAAGTCACAGGCCATGCCCCGCAAGGGGATGCACGAAGACCCA
CCGCGAGCCAGGAAGGGAGCACGGGGCTCTCTGCTCTGGGACCGGCAGTGAGCCGGACATCTGGGTCC
TCCCAAGCCGGGCGGGCTGCCCCAGGGAGGAAGGGAGGGGGGCGAGCCTGAGCGGGCACCTCGGCCCG
CAGGAGGTCTGCAGCGAGCTGTGGTGTCTGAGCAAGAGCAACCGGTGCATCACCAACAGCATCCCGGC
CGCCGAGGGCACGCTGTGCCAGACGCACACCATCGACAAGGGGTGGTGCTACAAACGGGTCTGTGTCC
CCTTTGGGTCGCGCCCAGAGGGTGTGGACGGAGCCTGGGGGCCGTGGACTCCATGGGGCGACTGCAGC
CGGACCTGTGGCGGCGGCGTGTCCTCTTCTAGCCGTCACTGCGACAGCCCCAGGCCAACCATCGGGGG
CAAGTACTGTCTGGGTGAGAGAAGGCGGCACCGCTCCTGCAACACGGATGACTGTCCCCCTGGCTCCC

AGGACTTCAGAGAAGTGCAGTGTTCTGAATTTGACAGCATCCCTTTCCGTGGGAAATTCTACAAGTGG
AAAACGTACCGGGGAGGGGGCGTGAAGGCCTGCTCGCTCACGTGCCTAGCGGAAGGCTTCAACTTCTA
CACGGAGAGGGCGGCAGCCGTGGTGGACGGGACACCCTGCCGTCCAGACACGGTGGACATTTGCGTCA
GTGGCGAATGCAAGCACGTGGGCTGCGACCGAGTCCTGGGCTCCGACCTGCGGGAGGACAAGTGCCGA
GTGTGTGGCGGTGACGGCAGTGCCTGCGAGACCATCGAGGGCGTCTTCAGCCCAGCCTCACCTGGGGC
CGGGTACGAGGATGTCGTCTGGATTCCCAAAGGCTCCGTCCACATCTTCATCCAGGATCTGAACCTCT
CTCTCAGTCACTTGGCCCTGAAGGGAGACCAGGAGTCCCTGCTGCTGGAGGGGCTGCCCGGGACCCCC
CAGCCCCACCGTCTGCCTCTAGCTGGGACCACCTTTCAACTGCGACAGGGGCCAGACCAGGTCCAGAG
CCTCGAAGCCCTGGGACCGATTAATGCATCTCTCATCGTCATGGTGCTGGCCCGGACCGAGCTGCCTG
CCCTCCGCTACCGCTTCAATGCCCCCATCGCCCGTGACTCGCTGCCCCCCTACTCCTGGCACTATGCG
CCCTGGACCAAGTGCTCGGCCCAGTGTGCAGGCGGTAGCCAGGTGCAGGCGGTGGAGTGCCGCAACCA
GCTGGACAGCTCCGCGGTCGCCCCCCACTACTGCAGTGCCCACAGCAAGCTGCCCAAAAGGCAGCGCG
CCTGCAACACGGAGCCTTGCCCTCCAGACTGGGTTGTAGGGAACTGGTCGCTCTGCAGCCGCAGCTGC
GATGCAGGCGTGCGCAGCCGCTCGGTCGTGTGCCAGCGCCGCGTCTCTGCCGCGGAGGAGAAGGCGCT

GGACGACAGCGCATGCCCGCAGCCGCGCCCACCTGTACTGGAGGCCTGCCACGGCCCCACTTGCCCTC
CGGAGTGGGCGGCCCTCGACTGGTCTGAGTGCACCCCCAGCTGCGGGCCGGGCCTCCGCCACCGCGTG
GTCCTTTGCAAGAGCGCAGACCACCGCGCCACGCTGCCCCCGGCGCACTGCTCACCCGCCGCCAAGCC
ACCGGCCACCATGCGCTGCAACTTGCGCCGCTGCCCCCCGGCCCGCTGGGTGGCTGGCGAGTGGGGTG
AGTGCTCTGCACAGTGCGGCGTCGGGCAGCGGCAGCGCTCGGTGCGCTGCACCAGCCACACGGGCCAG
GCGTCGCACGAGTGCACGGAGGCCCTGC

NOV17e, CG51213-03 Protein Sequence	SEQ ID NO: 204	634 aa	MW at 68754.0kD

YCLKRYMACIKCSINGAYWESISHRPCPARGCTKTHREPGREHGALCSGTGSEPDIWVLPSRAGCPRE
EGRGASLSGHLGPQEVCSELWCLSKSNRCITNSIPAAEGTLCQTHTIDKGWCYKRVCVPFGSRPEGVD
GAWGPWTPWGDCSRTCGGGVSSSSRHCDSPRPTIGGKYCLGERRRHRSCNTDDCPPGSQDFREVQCSE
FDSIPFRGKFYKWKTYRGGGVKACSLTCLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVGCD
RVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNLSLSHLALKGD
QESLLLEGLPGTPQPHRLPLAGTTFQLRQGPDQVQSLEALGPINASLIVMVLARTELPALRYRFNAPI
ARDSLPPYSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSAVAPHYCSAHSKLPKRQRACNTEPCPPD
WVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEKALDDSACPQPRPPVLEACHGPTCPPEWAALDWSE
CTPSCGPGLRHRVVLCKSADHRATLPPAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQ
RQRSVRCTSHTGQASHECTEAL

NOV17f, CG51213-05 DNA Sequence	SEQ ID NO: 205	3400 bp
	ORF Start: at 1	ORF Stop: TAG at 3232

CGGTCTCAAGATGAGTTCCTGTCCAGTCTGGAGAGCTATGAGATCGCCTTCCCCACCCGCGTGGACCA
CAACGGGGCACTGCTGGCCTTCTCGCCACCTCCTCCCCGGAGGCAGCGCCGCGGCACGGGGGCCACAG
CCGAGTCCCGCCTCTTCTACAAAGTAGCCTCGCCCAGCACCCACTTCCTGCTGAACCTGACCCGCAGC
TCCCGTCTACTGGCAGGGCACGTCTCCGTGGAGTACTGGACACGGGAGGGCCTGGCCTGGCAGAGGGC
GGCCCGGCCCCACTGCCTCTACGCTGGTCACCTGCAGGGCCAGGCCAGCAGCTCCCATGTGGCCATCA
GCACCTGTGGAGGCCTGCACGGCCTGATCGTGGCAGACGAGGAAGAGTACCTGATTGAGCCCCTGCAC
GGTGGGCCCAAGGGTTCTCGGAGCCCGGAGGAAAGTGGACCACATGTGGTGTACAAGCGTTCCTCTCT
GCGTCACCCCCACCTGGACACAGCCTGTGGAGTGAGAGATGAGAAACCGTGGAAAGGGCGGCCATGGT
GGCTGCGGACCTTGAAGCCACCGCCTGCCAGACCCCTGGGGAATGAAACAGAGCGTGGCCAGCCAGGC
CTGAAGCGATCGGTCAGCCGAGAGCGCTACGTGGAGACCCTGGTGGTGGCTGACAAGATGATGGTGGC
CTATCACGGGCGCCGGGATGTGGAGCAGTATGTCCTGGCCATCATGAACATTGTTGCCAAACTTTTCC
AGGACTCGAGTCTGGGAAGCACCGTTAACATCCTCGTAACTCGCCTCATCCTGCTCACGGAGGACCAG
CCCACTCTGGAGATCACCCACCATGCCGGGAAGTCCCTAGACAGCTTCTGTAAGTGGCAGAAATCCAT
CGTGAACCACAGCGGCCATGGCAATGCCATTCCAGAGAACGGTGTGGCTAACCATGACACAGCAGTGC
TCATCACACGCTATGACATCTGCATCTACAAGAACAAACCCTGCGGCACACTAGGCCTGGCCCCGGTG
GGCGGAATGTGTGAGCGCGAGAGAAGCTGCAGCGTCAATGAGGACATTGGCCTGCCACAAGCGTTCAC
CATTGCCCACGAGATCGGGCACACATTCGGCATGAACCATGACGGCGTGGGAAACAGCTGTGGGGCCC
GTGGTCAGGACCCAGCCAAGCTCATGGCTGCCCACATTACCATGAAGACCAACCCATTCGTGTGGTCA
TCCTGCAACCGTGACTACATCACCAGCTTTCTAGACTCGGGCCTGGGGCTCTGCCTGAACAACCGGCC
CCCCAGACAGGACTTTGTGTACCCGACAGTGGCACCGGGCCAAGCCTACGATGCAGATGAGCAATGCC
GCTTTCAGCATGGAGTCAAATCGCGTCAGTGTAAATACGGGGAGGTCTGCAGCGAGCTGTGGTGTCTG
AGCAAGAGCAACCGGTGCATCACCAACAGCATCCCGGCCGCCGAGGGCACGCTGTGCCAGACGCACAC

CATCGACAAGGGGTGGTGCTACAAACGGGTCTGTGTCCCCTTTGGGTCGCGCCCAGAGGGTGTGGACG
GAGCCTGGGGGCCGTGGACTCCATGGGGCGACTGCAGCCGGACCTGTGGCGGCGGCGTGTCCTCTTCT
AGTCGTCACTGCGACAGCCCCAGGCCAACCATCGGGGGCAAGTACTGTCTGGGTGAGAGAAGGCGGCA
CCGCTCCTGCAACACGGATGACTGTCCCCCTGGCTCCCAGGACTTCAGAGAAGTGCAGTGTTCTGAAT
TTGACAGCATCCCTTTCCGTGGGAAATTCTACAAGTGGAAAACGTACCGGGGAGGGGGCGTGAAGGCC
TGCTCGCTCACGAGCCTAGCGGAAGGCTTCAACTTCTACACGGAGAGGGCGGCAGCCGTGGTGGACGG
GACACCCTGCCGTCCAGACACGGTGGACATTTGCGTCAGTGGCGAATGCAAGCACGTGGGCTGCGACC
GAGTCCTGGGCTCCGACCTGCGGGAGGACAAGTGCCGAGTGTGTGGCGGTGACGGCAGTGCCTGCGAG
ACCATCGAGGGCGTCTTCAGCCCAGCCTCACCTGGGGCCGGGTACGAGGATGTCGTCTGGATTCCCAA
AGGCTCCGTCCACATCTTCATCCAGGATCTGAACCTCTCTCTCAGTCACTTGGCCCTGAAGGGAGACC
AGGAGTCCCTGCTGCTGGAGGGGCTGCCTGGGACCCCCCAGCCCCACCGTCTGCCTCTAGCTGGGACC
ACCTTTCAACTGCGACAGGGGCCAGACCAGGTCCAGAGCCTCGAAGCCCTGGGACCGATTAATGCATC
TCTCATCGTCATGGTGCTGGCCCGGACCGAGCTGCCTGCCCTCCGCTACCGCTTCAATGCCCCCATCG
CCCGTGACTCGCTGCCCCCCTACTCCTGGCACTATGCGCCCTGGACCAAGTGCTCGGCCCAGTGTGCA
GGCGGTAGCCAGGTGCAGGCGGTGGAGTGCCGCAACCAGCTGGACAGCTCCGCGGTCGCCCCCCACTA
CTGCAGTGCCCACAGCAAGCTGCCCAAAAGGCAGCGCGCCTGCAACACGGAGCCTTGCCCTCCAGACT
GGGTTGTAGGGAACTGGTCGCTCTGCAGCCGCAGCTGCGATGCAGGCGTGCGCAGTCGCTCGGTCGTG
TGCCAGCGCCGCGTCTCTGCCGCGGAGGAGAAGGCGCTGGACGACAGCGCATGCCCGCAGCCGCGCCC
ACCTGTACTGGAGGCCTGCCACGGCCCCACTTGCCCTCCGGAGTGGGCGGCCCTCGACTGGTCTGAGT
GCACCCCCAGCTGCGGGCCGGGCCTCCGCCACCGCGTGGTCCTTTGCAAGAGCGCAGACCACCGCGCC
ACGCTGCCCCCGGCGCACTGCTCACCCGCCGCCAAGCCACCGGCCACCATGCGCTGCAACTTGCGCCG
CTGCCCCCCGGCCCGCTGGGTGGCTGGCGAGTGGGGTGAGTGCTCTGCACAGTGCGGCGTCGGGCAGC
GGCAGCGCTCGGTGCGCTGCACCAGCCACACGGGCCAGGCGTCGCACGAGTGCACGGAGGCCCTGCGG
CCGCCCACCACGCAGCAGTGTGAGGCCAAGTGCGACAGCCCAACCCCCGGGGACGGCCCTGAAGAGTG
CAAGGATGTGAACAAGGTCGCCTACTGCCCCCTGGTGCTCAAATTTCAGTTCTGCAGCCGAGCCTACT
TCCGCCAGATGTGCTGCAAAACCTGCCAGGGCCAC**TAG**GGGGCGCGCGGCACCCGGAGCCACAGCTGG
CGGGGTCTCCGCCGCCAGCCCTGCAGCGGGCCGGCCAAAGGGGGCCCCGGGGGGGCGGGAACTGGGAG

GGAAGGGTGAGACGGAGCCGGAAGTTATTTATTGGGAACCCCTGCAGGGCCCTGGCTGGGGGGATGGA

NOV17f, CG51213-05 Protein Sequence	SEQ ID NO: 206	1077 aa	MW at 118071.4kD

RSQDEFLSSLESYEIAFPTRVDHNGALLAFSPPPPRRQRRGTGATAESRLFYKVASPSTHFLLNLTRS
SRLLAGHVSVEYWTREGLAWQRAARPHCLYAGHLQGQASSSHVAISTCGGLHGLIVADEEEYLIEPLH
GGPKGSRSPEESGPHVVYKRSSLRHPHLDTACGVRDEKPWKGRPWWLRTLKPPPARPLGNETERGQPG
LKRSVSRERYVETLVVADKMMVAYHGRRDVEQYVLAIMNIVAKLFQDSSLGSTVNILVTRLILLTEDQ
PTLEITHHAGKSLDSFCKWQKSIVNHSGHGNAIPENGVANHDTAVLITRYDICIYKNKPCGTLGLAPV
GGMCERERSCSVNEDIGLPQAFTIAHEIGHTFGMNHDGVGNSCGARGQDPAKLMAAHITMKTNPFVWS
SCNRDYITSFLDSGLGLCLNNRPPRQDFVYPTVAPGQAYDADEQCRFQHGVKSRQCKYGEVCSELWCL
SKSNRCITNSIPAAEGTLCQTHTIDKGWCYKRVCVPFGSRPEGVDGAWGPWTPWGDCSRTCGGGVSSS
SRHCDSPRPTIGGKYCLGERRRHRSCNTDDCPPGSQDFREVQCSEFDSIPFRGKFYKWKTYRGGGVKA
CSLTSLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVGCDRVLGSDLREDKCRVCGGDGSACE
TIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNLSLSHLALKGDQESLLLEGLPGTPQPHRLPLAGT
TFQLRQGPDQVQSLEALGPINASLIVMVLARTELPALRYRFNAPIARDSLPPYSWHYAPWTKCSAQCA
GGSQVQAVECRNQLDSSAVAPHYCSAHSKLPKRQRACNTEPCPPDWVVGNWSLCSRSCDAGVRSRSVV
CQRRVSAAEEKALDDSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRA
TLPPAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTGQASHECTEALR
PPTTQQCEAKCDSPTPGDGPEECKDVNKVAYCPLVLKFQFCSRAYFRQMCCKTCQGH

NOV17g, CG51213-06 DNA Sequence	SEQ ID NO: 207	978 bp
	ORF Start: at 1	ORF Stop: end of sequence

TCCATAAATGGAGCTTATTGGGAGAGTATAAGTCACAGGCCATGCCCCGCAAGGGGATGCACGAAGAC
CCACCGCGAGCCAGGAAGGGAGCACCGGGCTCTCTGCTCTGGGACCGGCAGTGAGCCGGACATCTGGG
TCCTCCCAAGCCGGGCGGGCTGCCCCAGGGAGGAAGGGAGGGGGGCGAGCCTGAGCGGGCACCTCGGC
CCGCAGGAGGTCTGCAGCGAGCTGTGGTGTCTGAGCAAGAGCAACCGGTGCATCACCAACAGCATCCC
GGCCGCCGAGGGCACGCTGTGCCAGACGCACACCATCGACAAGGGGTGGTGCTACAAACGGGTCTGTG
TCCCCTTTGGGTCGCGCCCAGAGGGTGTGGACGGAGCCTGGGGGCCGTGGACTCCATGGGGCGACTGC
AGCCGGACCTGTGGCGGCGGCGTGTCCTCTTCTAGCCGTCACTGCGACAGCCCCAGGCCAACCATCGG
GGGCAAGTACTGTCTGGGTGAGAGAAGGCGGCACCGCTCCTGCAACACGGATGACTGTCCCCCTGGCT

CCCAGGACTTCAGAGAAGTGCAGTGTTCTGAATTTGACAGCATCCCTTTCCGTGGGAAATTCTACAAG
TGGAAAACGTACCGGGGAGGGGGCGTGAAGGCCTGCTCGCTCACGTGCCTAGCGGAAGGCTTCAACTT
CTACACGGAGAGGGCGGCAGCCGTGGTGGACGGGACACCCTGCCGTCCAGACACGGTGGACATTTGCG
TCAGTGGCGAATGCAAGCACGTGGGCTGCGACCGAGTCCTGGGCTCCGACCTGCGGGAGGACAAGTGC
CGAGTGTGTGGCGGTGACGGCAGTGCCTGCGAGACCATCGAGGGCGTCTTCAGCCCAGCCTCACCTGG
GGCCGGGTACGAGGATGTCGTCTGGATTCCCAAAGGCTCCGTCCACATCTTCATCCAGGATCTGAACC
TCTCTCTCAGTCACTTGGCCCTGAAG

NOV17g, CG51213-06 Protein Sequence	SEQ ID NO: 208	326 aa	MW at 35330.2kD

SINGAYWESISHRPCPARGCTKTHREPGREHRALCSGTGSEPDIWVLPSRAGCPREEGRGASLSGHLG
PQEVCSELWCLSKSNRCITNSIPAAEGTLCQTHTIDKGWCYKRVCVPFGSRPEGVDGAWGPWTPWGDC
SRTCGGGVSSSSRHCDSPRPTIGGKYCLGERRRHRSCNTDDCPPGSQDFREVQCSEFDSIPFRGKFYK
WKTYRGGGVKACSLTCLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVGCDRVLGSDLREDKC
RVCGGDGSACETIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNLSLSHLALK

NOV17h, CG51213-07 DNA Sequence	SEQ ID NO: 209	1866 bp
	ORF Start: at 1	ORF Stop: end of sequence

TCCATAAATGGAGCTTATTGGGAGAGTATAAGTCACAGGCCATGCCCCGCAAGGGGATGCACGAAGAC
CCACCGCGAGCCAGGAAGGGAGCACCGGGCTCTCTGCTCTGGGACCGGCAGTGAGCCGGACATCTGGG
TCCTCCCAAGCCGGGCGGGCTGCCCCAGGGAGGAAGGGAGGGGGGCGAGCCTGAGCGGGCACCTCGGC
CCGCAGGAGGTCTGCAGCGAGCTGTGGTGTCTGAGCAAGAGCAACCGGTGCATCACCAACAGCATCCC
GGCCGCCGAGGGCACGCTGTGCCAGACGCACACCATCGACAAGGGGTGGTGCTACAAACGGGTCTGTG
TCCCCTTTGGGTCGCGCCCAGAGGGTGTGGACGGAGCCTGGGGGCCGTGGACTCCATGGGGCGACTGC
AGCCGGACCTGTGGCGGCGGCGTGTCCTCTTCTAGCCGTCACTGCGACAGCCCCAGGCCAACCATCGG
GGGCAAGTACTGTCTGGGTGAGAGAAGGCGGCACCGCTCCTGCAACACGGATGACTGTCCCCCTGGCT
CCCAGGACTTCAGAGAAGTGCAGTGTTCTGAATTTGACAGCATCCCTTTCCGTGGGAAATTCTACAAG
TGGAAAACGTACCGGGGAGGGGGCGTGAAGGCCTGCTCGCTCACGTGCCTAGCGGAAGGCTTCAACTT
CTACACGGAGAGGGCGGCAGCCGTGGTGGACGGGACACCCTGCCGTCCAGACACGGTGGACATTTGCG
TCAGTGGCGAATGCAAGCACGTGGGCTGCGACCGAGTCCTGGGCTCCGACCTGCGGGAGGACAAGTGC
CGAGTGTGTGGCGGTGACGGCAGTGCCTGCGAGACCATCGAGGGCGTCTTCAGCCCAGCCTCACCTGG
GGCCGGGTACGAGGATGTCGTCTGGATTCCCAAAGGCTCCGTCCACATCTTCATCCAGGATCTGAACC
TCTCTCTCAGTCACTTGGCCCTGAAGGGAGACCAGGAGTCCCTGCTGCTGGAGGGGCTGCCCGGGACC
CCCCAGCCCCACCGTCTGCCTCTAGCTGGGACCACCTTTCAACTGCGACAGGGGCCAGACCAGGTCCA
GAGCCTCGAAGCCCTGGGACCGATTAATGCATCTCTCATCGTCATGGTGCTGGCCCGGACCGAGCTGC
CTGCCCTCCGCTACCGCTTCAATGCCCCCATCGCCCGTGACTCGCTGCCCCCCTACTCCTGGCACTAT
GCGCCCTGGACCAAGTGCTCGGCCCAGTGTGCAGGCGGTAGCCAGGTGCAGGCGGTGGAGTGCCGCAA
CCAGCTGGACAGCTCCGCGGTCGCCCCCCACTACTGCAGTGCCCACAGCAAGCTGCCCAAAAGGCAGC
GCGCCTGCAACACGGAGCCTTGCCCTCCAGACTGGGTTGTAGGGAACTGGTCGCTCTGCAGCCGCAGC
TGCGATGCAGGCGTGCGCAGCCGCTCGGTCGTGTGCCAGCGCCGCGTCTCTGCCGCGGAGGAGAAGGC
GCTGGACGACAGCGCATGCCCGCAGCCGCGCCCACCTGTACTGGAGGCCTGCCACGGCCCCACTTGCC
CTCCGGAGTGGGCGGCCCTCGACTGGTCTGAGTGCACCCCCAGCTGCGGGCCGGGCCTCCGCCACCGC
GTGGTCCTTTGCAAGAGCGCAGACCACCGCGCCACGCTGCCCCCGGCGCACTGCTCACCCGCCGCCAA
GCCACCGGCCACCATGCGCTGCAACTTGCGCCGCTGCCCCCCGGCCCGCTGGGTGGCTGGCGAGTGGG
GTGAGTGCTCTGCACAGTGCGGCGTCGGGCAGCGGCAGCGCTCGGTGCGCTGCACCAGCCACACGGGC
CAGGCGTCGCACGAGTGCACGGAGGCCCTG

NOV17h, CG51213-07 Protein Sequence	SEQ ID NO: 210	622 aa	MW at 67376.2kD

SINGAYWESISHRPCPARGCTKTHREPGREHRALCSGTGSEPDIWVLPSRAGCPREEGRGASLSGHLG
PQEVCSELWCLSKSNRCITNSIPAAEGTLCQTHTIDKGWCYKRVCVPFGSRPEGVDGAWGPWTPWGDC
SRTCGGGVSSSSRHCDSPRPTIGGKYCLGERRRHRSCNTDDCPPGSQDFREVQCSEFDSIPFRGKFYK
WKTYRGGGVKACSLTCLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVGCDRVLGSDLREDKC
RVCGGDGSACETIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNLSLSHLALKGDQESLLLEGLPGT
PQPHRLPLAGTTFQLRQGPDQVQSLEALGPINASLIVMVLARTELPALRYRFNAPIARDSLPPYSWHY
APWTKCSAQCAGGSQVQAVECRNQLDSSAVAPHYCSAHSKLPKRQRACNTEPCPPDWVVGNWSLCSRS
CDAGVRSRSVVCQRRVSAAEEKALDDSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHR
VVLCKSADHRATLPPAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTG
QASHECTEAL

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 17B.

Table 17B. Comparison of the NOV17 protein sequences.

```
NOV17a  ------------------------------------------------------------
NOV17b  ------------------------------------------------------------
NOV17c  ------------------------------------------------------------
NOV17d  ------------------------------------------------------------
NOV17e  ------------------------------------------------------------
NOV17f  RSQDEFLSSLESYEIAFPTRVDHNGALLAFSPPPPRRQRRGTGATAESRLFYKVASPSTH
NOV17g  ------------------------------------------------------------
NOV17h  ------------------------------------------------------------

NOV17a  ------------------------------------------------------------
NOV17b  ------------------------------------------------------------
NOV17c  ------------------------------------------------------------
NOV17d  ------------------------------------------------------------
NOV17e  ------------------------------------------------------------
NOV17f  FLLNLTRSSRLLAGHVSVEYWTREGLAWQRAARPHCLYAGHLQGQASSSHVAISTCGGLH
NOV17g  ------------------------------------------------------------
NOV17h  ------------------------------------------------------------

NOV17a  ------------------------------------------------------------
NOV17b  ------------------------------------------------------------
NOV17c  ------------------------------------------------------------
NOV17d  ------------------------------------------------------------
NOV17e  ------------------------------------------------------------
NOV17f  GLIVADEEEYLIEPLHGGPKGSRSPEESGPHVVYKRSSLRHPHLDTACGVRDEKPWKGRP
NOV17g  ------------------------------------------------------------
NOV17h  ------------------------------------------------------------

NOV17a  ------------------------------------------MMVAYHGRRDVEQYVLAI
NOV17b  ------------------------------------------------------------
NOV17c  ------------------------------------------------------------
NOV17d  ------------------------------------------------------------
NOV17e  ------------------------------------------------------------
NOV17f  WWLRTLKPPPARPLGNETERGQPGLKRSVSRERYVETLVVADKMMVAYHGRRDVEQYVLA
NOV17g  ------------------------------------------------------------
NOV17h  ------------------------------------------------------------

NOV17a  MNIQVAKLFQDSSLGSTVNILVTRLILLTEDQPTLEITHHAGKSLDSFCKWQKSIVNHSG
NOV17b  ------------------------------------------------------------
NOV17c  ------------------------------------------------------------
NOV17d  ------------------------------------------------------------
NOV17e  ------------------------------------------------------------
NOV17f  IMNIVAKLFQDSSLGSTVNILVTRLILLTEDQPTLEITHHAGKSLDSFCKWQKSIVNHSG
NOV17g  ------------------------------------------------------------
NOV17h  ------------------------------------------------------------

NOV17a  HGNAIPENGVANHDTAVLITRYDICIYKNKPCGTLGLAPVGGMCERERSCSVNEDIGLAT
NOV17b  ------------------------------------------------------------
NOV17c  ------------------------------------------------------------
```

```
NOV17d  ------------------------------------------------------------
NOV17e  ------------------------------------------------------------
NOV17f  HGNAIPENGVANHDTAVLITRYDICIYKNKPCGTLGLAPVGGMCERERSCSVNEDIGLPQ
NOV17g  ------------------------------------------------------------
NOV17h  ------------------------------------------------------------

NOV17a  AFTIAHEIGHTFGMNHDGVGNSCGARGQDPAKLMAAHITMKTNPFVWSSCSRDYITSFLD
NOV17b  ------------------------------------------------------------
NOV17c  ------------------------------------------------------------
NOV17d  ------------------------------------------------------------
NOV17e  ------------------------------------------------------------
NOV17f  AFTIAHEIGHTFGMNHDGVGNSCGARGQDPAKLMAAHITMKTNPFVWSSCNRDYITSFLD
NOV17g  ------------------------------------------------------------
NOV17h  ------------------------------------------------------------

NOV17a  SGLGLCLNNRPPRQDFVYPTVAPGQAYDADEQCRFQHGVKSRQCKYGEVCSELWCLSKSN
NOV17b  ------------------------------------------------------------
NOV17c  ------------------------------------------------------------
NOV17d  ----------------------------YCLKRYMACIKCSINGAYWESISHRPCPAR
NOV17e  ----------------------------YCLKRYMACIKCSINGAYWESISHRPCPAR
NOV17f  SGLGLCLNNRPPRQDFVYPTVAPGQAYDADEQCRFQHGVKSRQCKYGEVCSELWCLSKSN
NOV17g  ------------------------------------------SINGAYWESISHRPCPAR
NOV17h  ------------------------------------------SINGAYWESISHRPCPAR

NOV17a  RCITNSIPAAEGTLCQTHTIDKGWCYKRVCVPFGSRPEGVDGAWGPWTPWGDCSRTCGGG
NOV17b  -------------------------------MGVQRD----AGSPHSLCPSCPSSWVWG
NOV17c  ------------------------------------------------------------
NOV17d  GCTKTHREPGREHRALCSGTGSEPDIWVLPSRAGCPREEGRGASLSGHLGPQEVCSELWC
NOV17e  GCTKTHREPGREHGALCSGTGSEPDIWVLPSRAGCPREEGRGASLSGHLGPQEVCSELWC
NOV17f  RCITNSIPAAEGTLCQTHTIDKGWCYKRVCVPFGSRPEGVDGAWGPWTPWGDCSRTCGGG
NOV17g  GCTKTHREPGREHRALCSGTGSEPDIWVLPSRAGCPREEGRGASLSGHLGPQEVCSELWC
NOV17h  GCTKTHREPGREHRALCSGTGSEPDIWVLPSRAGCPREEGRGASLSGHLGPQEVCSELWC

NOV17a  VS----------------------S--SS---------RHCD-----------------
NOV17b  LS----------------------L--LA---------SLPQ-----------------
NOV17c  ------------------------------------------------------------
NOV17d  LSKSNRCITNSIPAAEGTLCQTHTIDKGWCYKRVCVPFGSRPEGVDGAWGPWTPWGDCSR
NOV17e  LSKSNRCITNSIPAAEGTLCQTHTIDKGWCYKRVCVPFGSRPEGVDGAWGPWTPWGDCSR
NOV17f  VS----------------------S--SS---------RHCD-----------------
NOV17g  LSKSNRCITNSIPAAEGTLCQTHTIDKGWCYKRVCVPFGSRPEGVDGAWGPWTPWGDCSR
NOV17h  LSKSNRCITNSIPAAEGTLCQTHTIDKGWCYKRVCVPFGSRPEGVDGAWGPWTPWGDCSR

NOV17a  ---S-PR--P--------TIGGKYCLSERRRHRSCNTDDCPPGSQDFREVQCSEFDSIPF
NOV17b  ---G-AS--K--------EFRGGLARG--------VTGPCPPG-----------------
NOV17c  ------------------------------------------------------------
NOV17d  TCGGGVSSSSRHCDSPRPTIGGKYCLGERRRHRSCNTDDCPPGSQDFREVQCSEFDSIPF
NOV17e  TCGGGVSSSSRHCDSPRPTIGGKYCLGERRRHRSCNTDDCPPGSQDFREVQCSEFDSIPF
NOV17f  ---S-PR--P--------TIGGKYCLGERRRHRSCNTDDCPPGSQDFREVQCSEFDSIPF
NOV17g  TCGGGVSSSSRHCDSPRPTIGGKYCLGERRRHRSCNTDDCPPGSQDFREVQCSEFDSIPF
NOV17h  TCGGGVSSSSRHCDSPRPTIGGKYCLGERRRHRSCNTDDCPPGSQDFREVQCSEFDSIPF

NOV17a  RGKFYKWKTYRGGGVKACSLTSLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVG
NOV17b  ------------GVKACSLTCLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVG
NOV17c  ------------------------------------------------------------
```

```
NOV17d  RGKFYKWKTYRGGGVKACSLTCLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVG
NOV17e  RGKFYKWKTYRGGGVKACSLTCLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVG
NOV17f  RGKFYKWKTYRGGGVKACSLTSLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVG
NOV17g  RGKFYKWKTYRGGGVKACSLTCLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVG
NOV17h  RGKFYKWKTYRGGGVKACSLTCLAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVG

NOV17a  CDRVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNL
NOV17b  CDRVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNL
NOV17c  ------------------------------------------------------------
NOV17d  CDRVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNL
NOV17e  CDRVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNL
NOV17f  CDRVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNL
NOV17g  CDRVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNL
NOV17h  CDRVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYEDVVWIPKGSVHIFIQDLNL

NOV17a  SLSHLALKGDQESLLLEGLPGTPQPHRLPLAGTTFQLRQGPDQVQSLEALGPINASLIVM
NOV17b  SLSHLALKGDQESLLLEGLPGTPQPHRLPLAGTTFQLRQGPDQVQSLEALGPINASLIVM
NOV17c  ------------------------------------------------------------
NOV17d  SLSHLALKGDQESLLLEGLPGTPQPHRLPLAGTTFQLRQGPDQVQSLEALGPINASLIVM
NOV17e  SLSHLALKGDQESLLLEGLPGTPQPHRLPLAGTTFQLRQGPDQVQSLEALGPINASLIVM
NOV17f  SLSHLALKGDQESLLLEGLPGTPQPHRLPLAGTTFQLRQGPDQVQSLEALGPINASLIVM
NOV17g  SLSHLALK------------------------------------------------------
NOV17h  SLSHLALKGDQESLLLEGLPGTPQPHRLPLAGTTFQLRQGPDQVQSLEALGPINASLIVM

NOV17a  VLARTELPALRYRFNAPIARDSLPPYSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSAV
NOV17b  VLARTELPALRYRFNAPIARDSLPPYSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSAV
NOV17c  ----------------------TGSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSAV
NOV17d  VLARTELPALRYRFNAPIARDSLPPYSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSAV
NOV17e  VLARTELPALRYRFNAPIARDSLPPYSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSAV
NOV17f  VLARTELPALRYRFNAPIARDSLPPYSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSAV
NOV17g  ------------------------------------------------------------
NOV17h  VLARTELPALRYRFNAPIARDSLPPYSWHYAPWTKCSAQCAGGSQVQAVECRNQLDSSAV

NOV17a  APHYCSAHSKLPKRQRACNTEPCPPDWVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEK
NOV17b  APHYCSAHSKLPKRQRACNTEPCPPDWVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEK
NOV17c  APHYCSAHSKLPKRQRACNTEPCPPDWVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEK
NOV17d  APHYCSAHSKLPKRQRACNTEPCPPDWVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEK
NOV17e  APHYCSAHSKLPKRQRACNTEPCPPDWVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEK
NOV17f  APHYCSAHSKLPKRQRACNTEPCPPDWVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEK
NOV17g  ------------------------------------------------------------
NOV17h  APHYCSAHSKLPKRQRACNTEPCPPDWVVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEK

NOV17a  ALDDSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLP
NOV17b  ALDDSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLP
NOV17c  ALDDSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLP
NOV17d  ALDDSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLP
NOV17e  ALDDSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLP
NOV17f  ALDDSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLP
NOV17g  ------------------------------------------------------------
NOV17h  ALDDSACPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLP

NOV17a  PAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTGQASHEC
NOV17b  PAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTGQASHEC
NOV17c  PAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTGQASHEC
```

```
NOV17d  PAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTGQASHEC
NOV17e  PAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTGQASHEC
NOV17f  PAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTGQASHEC
NOV17g  ------------------------------------------------------------
NOV17h  PAHCSPAAKPPATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQRSVRCTSHTGQASHEC

NOV17a  TEALRPPTTQQCEAKCDSPTPGDGPEECKDVNKVAYCPLVLKFQFCSRAYFRQMCCKTCQ
NOV17b  TEALRPPTTQQCEAKCDSPTPGDGPEECKDVNKVAYCPLVLKFQFCSRAYFRQMCCKTCQ
NOV17c  TEALRPPTTQQCEAKCLEG----------------------------------------
NOV17d  TEAL------------------------------------------------------------
NOV17e  TEAL------------------------------------------------------------
NOV17f  TEALRPPTTQQCEAKCDSPTPGDGPEECKDVNKVAYCPLVLKFQFCSRAYFRQMCCKTCQ
NOV17g  ------------------------------------------------------------
NOV17h  TEAL------------------------------------------------------------

NOV17a  GH
NOV17b  GH
NOV17c  --
NOV17d  --
NOV17e  --
NOV17f  GH
NOV17g  --
NOV17h  --

NOV17a  (SEQ ID NO:  196)
NOV17b  (SEQ ID NO:  198)
NOV17c  (SEQ ID NO:  200)
NOV17d  (SEQ ID NO:  202)
NOV17e  (SEQ ID NO:  204)
NOV17f  (SEQ ID NO:  206)
NOV17g  (SEQ ID NO:  208)
NOV17h  (SEQ ID NO:  210)
```

Further analysis of the NOV17a protein yielded the following properties shown in Table 17C.

Table 17C. Protein Sequence Properties NOV17a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 10;  pos.chg 2;  neg.chg 1
      H-region:  length 1;  peak value  -0.15
      PSG score:  -4.55

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -8.03
      possible cleavage site: between 33 and 34

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
```

```
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   0
      PERIPHERAL  Likelihood =  1.38 (at 31)
      ALOM score:  -1.70  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       2       Hyd Moment(75):  5.01
      Hyd Moment(95):  5.34    G content:       1
      D/E content:     2       S/T content:     0
      Score: -5.94

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 18  GRR|DV

NUCDISC: discrimination of nuclear localization signals
      pat4: RRRH (3) at  342
      pat4: RRHR (3) at  343
      pat7: PKRQRAC (5) at  625
      bipartite: none
      content of basic residues:  10.6%
      NLS Score:  0.33

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: found
      KLPK at 623

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:
      Leucine zipper pattern (PS00029):  *** found ***
         LGSTVNILVTRLILLTEDQPTL at 32
 none
```

```
checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 76.7

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       95.7 %: nuclear
        4.3 %: cytoplasmic

>> prediction for CG51213-04 is nuc (k=23)
```

A search of the NOV17a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 17D.

Table 17D. Geneseq Results for NOV17a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV17a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU97888	Human aggrecanase protein #2 - Homo sapiens, 1104 aa. [WO200234895-A2, 02-MAY-2002]	1..855 250..1103	852/855 (99%) 852/855 (99%)	0.0
AAB74945	Human ADAM type metal protease MDTS2 protein SEQ ID NO:10 - Homo sapiens, 1103 aa. [JP2001008687-A, 16-JAN-2001]	1..855 250..1103	852/855 (99%) 852/855 (99%)	0.0
AAU72890	Human metalloprotease partial protein sequence #2 - Homo sapiens, 1103 aa. [WO200183782-A2, 08-NOV-2001]	1..855 250..1103	851/855 (99%) 852/855 (99%)	0.0
ABG76505	DNA encoding protein modification and maintenance molecule #9 - Homo sapiens, 1103 aa. [WO200260942-A2, 08-AUG-2002]	1..855 250..1103	851/855 (99%) 852/855 (99%)	0.0
AAB47719	ADAMTS-E - Homo sapiens, 1104 aa. [EP1149903-A1, 31-OCT-2001]	1..855 250..1104	852/856 (99%) 852/856 (99%)	0.0

In a BLAST search of public sequence databases, the NOV17a protein was found to have homology to the proteins shown in the BLASTP data in Table 17E.

Table 17E. Public BLASTP Results for NOV17a				
Protein Accession Number	**Protein/Organism/Length**	**NOV17a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9H324	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Homo sapiens (Human), 1077 aa (fragment).	1..855 224..1077	850/855 (99%) 851/855 (99%)	0.0
Q8CG28	Zinc metalloendopeptidase - Mus musculus (Mouse), 1070 aa.	1..854 216..1069	809/855 (94%) 824/855 (95%)	0.0
CAD20434	Sequence 8 from Patent WO0188156 - Homo sapiens (Human), 1044 aa (fragment).	1..796 250..1044	793/796 (99%) 793/796 (99%)	0.0
CAC37777	Sequence 1 from Patent WO0123561 - Homo sapiens (Human), 634 aa (fragment).	246..797 83..634	550/552 (99%) 550/552 (99%)	0.0
CAD20435	Sequence 11 from Patent WO0188156 - Homo sapiens (Human), 814 aa.	1..557 250..805	552/557 (99%) 552/557 (99%)	0.0

PFam analysis indicates that the NOV17a protein contains the domains shown in the Table 17F.

Table 17F. Domain Analysis of NOV17a			
Pfam Domain	**NOV17a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Reprolysin	2..209	59/235 (25%) 151/235 (64%)	1.4e-09
tsp_1	303..353	21/53 (40%) 37/53 (70%)	2.8e-07
tsp_1	581..636	11/59 (19%) 38/59 (64%)	0.056
tsp_1	640..696	14/63 (22%) 38/63 (60%)	0.13
tsp_1	698..754	16/57 (28%) 34/57 (60%)	0.43
tsp_1	759..809	17/55 (31%) 34/55 (62%)	0.2

Example 18.

The NOV18 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 18A.

Table 18A. NOV18 Sequence Analysis		
NOV18a, CG51448-05 DNA Sequence	SEQ ID NO: 211	1762 bp
	ORF Start: ATG at 197	ORF Stop: TGA at 1760

TTTGAGTTAGACAAGCAGCAGCACACGCCTCCCTACCTCATGGCGACAGAAAATGGAGCAGTTGAGCT
GGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGGTCCCACAGGTGAAAGACCCCTGGCTGCAG
GGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGGATCCACCCACCCTGAAGAAAGATGCCAAA
GCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAACCCTCAACTAGCAGCCCTGCCCCAGCAGA
CTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTGAGCAGGGAGCCTCAGGCAGCCAGGATCCT
GGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAAGCAGCAGCCAGGAGGGGCTCACCTGCCTT
TCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTCTGAGAAGCTGCTGGCCAAGAAGCCCCCAA
GCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGACCCACAGCCCCACGGATCCCAGGCCAGCC
AAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAGAAGGAAGTGGGAGAGAAAACCCCAGGCCA
GGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGGGATTGAGTTCCAGGCTGTTCCCTCAGAGA
AATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGGAGGAGGACTGCTTCCAGATTTTGGATGAT
TGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTGGAGCTGAGGACCGGGAATGTCAGCAGTGA
ATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGTGGCAAGTTTGGGGCAGTCTGTACCTGCATGGAGA
AAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGAAACAGACTCCCAAAGACAAGGAAATGGTG
TTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGCAATCTGATCCAGCTGTATGCAGCCATCGA
GACTCCGCATGAGATCGTCCTGTTCATGGAGTACATCGAGGGCGGAGAGCTCTTCGAGAGGATTGTGG
ATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTGTCAGGCAGATCTGTGACGGGATCCTC
TTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCAGAGAACATCCTGTGTGTCAACACCAC
CGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAGGTATAACCCCAACGAGAAGCTGAAGG

TGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGTGGTGAATTATGACCAAATCTCCGATAAGACA
GACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTGAGCGGCCTCTCCCCCTTCCTGGGAGATGA
TGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTGGTACTTTGATGAAGAGACCTTTGAGGCCG
TATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCGTCAAGGACCAGAGGGCCCGGATGAACGCT
GCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAGAAAGCCAAACGCTGTAACCGACGCCT
TAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCGCTGGAAGAAAAACTTCATTGCTGTCA
GCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCACTGATGGCTCTGGGGGTC**TGA**

NOV18a, CG51448-05 Protein Sequence	SEQ ID NO: 212	521 aa	MW at 57751.9kD

MPKPLPQRKGMVPWPNPQLAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQAAARRGS
PAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQKEVGEKT
PGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMVELRTGNV
SSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHRNLIQLYA
AIETPHEIVLFMEYIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKPENILCV
NTTGHLVKIIDFGLARRYNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYMLLSGLSPFL
GDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQCLAHPWLNNLAEKAKRCN
RRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18b, CG51448-01 DNA Sequence	SEQ ID NO: 213	1788 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1786

ATGGCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGG
TCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGG
ATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAA
CCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAGGGGCGGGGGGCCCGCGGAGGGCAGTGC
TGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTG
AGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAA
GCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTC
TGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGA
CCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAG
AAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGG
GATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGG
AGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTG
GAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGGGGCAA
GTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGA
AACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGC
AATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCCTGTTCATGGAGATCGAGGG
CGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTG
TCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCA
GAGAACATCCTGTGTGTCAACACCACCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAG
GTACCACAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGTGG
TGAATTATGACCAAATCTCCGATAAGACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTG
AGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTG
GTACTTTGATGAAGAGACCTTTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCG
TCAAGGACCAGGCCCGGATGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAG
AAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCG
CTGGAAGAAAAACTTCATTGCTGTCAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCAC
TGATGGCTCTGGGGGTC**TGA**

NOV18b, CG51448-01 Protein Sequence	SEQ ID NO: 214	595 aa	MW at 64501.9kD

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKP
ENILCVNTTGHLVKIIDFGLARRYHNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQARMNAAQCLAHPWLNNLAE

KAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18c, 274051198 DNA Sequence	SEQ ID NO: 215	787 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCAAGCTTATGAACTCCAAGGAGGCGCTCGGAGGGGGCAAGTTTGGGGCAGTCTGTACCTGCATGG
AGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGAAACAGACTCCCAAAGACAAGGAAATG
GTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGCAATCTGATCCAGCTGTATGCAGCCAT
CGAGACTCCGCATGAGATCGTCCTGTTCATGGAGTACATCGAGGGCGGAGAGCTCTTCGAGAGGATTG
TGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTGTCAGGCAGATCTGTGACGGGATC
CTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCAGAGAACATCCTGTGTGTCAACAC
CACCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAGGTATAACCCCAACGAGAAGCTGA
AGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGTGGTGAATTATGACCAAATCTCCGATAAG
ACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTGAGCGGCCTCTCCCCCTTCCTGGGAGA
TGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTGGTACTTTGATGAAGAGACCTTTGAGG
CCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCGTCAAGGACCAGAGGGCCCGGATGAAC
GCTGCCCAGTGTCTCGCCCATCCCTGGCTCGTCGACGGC

NOV18c, 274051198 Protein Sequence	SEQ ID NO: 216	262 aa	MW at 29771.1kD

TKLMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHRNLIQLYAAI
ETPHEIVLFMEYIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKPENILCVNT
TGHLVKIIDFGLARRYNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYMLLSGLSPFLGD
DDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQCLAHPWLVDG

NOV18d, 274051170 DNA Sequence	SEQ ID NO: 217	1813 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCAAGCTTCCCACCATGGCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAG
ACAAGGCACCTAAAGGTCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAAGACCCTGGCCCCCCAGAC
CCAAAGAAAGCTCCGGATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCAGAGAAAGGGGA
TGGTACCCTGGCCCAACCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAGGGGCGGGGGGC
CCGCGGAGGGCAGTGCTGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCGACACCTGAGACCAGCGTC
AAGAAGCCCAAGGCTGAGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCCAGGGTGGGCAAGAA
GGCAGCAGAGGGCCAAGCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCCCAGCTGTCCTG
CCATCATCTCCAGTTCTGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGCTCACCTTT
GAAGGGGTGCCCATGACCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAGAACAT
CCTGGCAGAGAGCCAGAAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGCTAAGATGCAAG
GGGACACCTCGAGGGGGATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTC
TGTCTCACAGCCAGGGAGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTT
CCCTCACCGCATGGTGGAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAACTCCAAGGAGG
CGCTCGGAGGTGGCAAGTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCA
GCCAAGGTCATCAAGAAACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATTGAGGTCATGAA
CCAGCTGAACCACCGCAATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCCTGT
TCATGGAGTACATCGAGGGCGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAG
GTGGACACCATGGTGTTTGTCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTT
GCACCTGGACCTCAAGCCAGAGAACATCCTGTGTGTCAACACCACCGGGCATTTGGTGAAGATCATTG
ACTTTGGCCTGGCACGGAGGTATAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCCCAGAGTTC
CTGTCACCTGAGGTGGTGAATTATGACCAAATCTCCGATAAGACAGACATGTGGAGTATGGGGGTGAT
CACCTACATGCTGCTGAGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCTAAACAACG
TTCTATCTGGCAACTGGTACTTTGATGAAGAGACCTTTGAGGCCGTATCAGACGAGGCCAAAGACTTT
GTCTCCAACCTCATCGTCAAGGACCAGAGGGCCCGGATGAACGCTGCCCAGTGTCTCGCCCATCCCTG
GCTCAACAACCTGGCGGAGAAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTTGCTTAAGA
AATACCTCATGAAGAGGCGCTGGAAGAAAAACTTCATTGCTGTCAGCGCTGCCAACCGCTTCAAGAAG
ATCAGCAGCTCGGGGGCACTGATGGCTCTGGGGGTCGTCGACGGC

NOV18d, 274051170 Protein Sequence	SEQ ID NO: 218	604 aa	MW at 65496.1kD

TKLPTMATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGD
GTLAQPSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKK

AAEGQAAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNI
LAESQKEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPF
PHRMVELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMN
QLNHRNLIQLYAAIETPHEIVLFMEYIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVL
HLDLKPENILCVNTTGHLVKIIDFGLARRYNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVI
TYMLLSGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQCLAHPW
LNNLAEKAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGVVDG

NOV18e, CG51448-02 DNA Sequence	SEQ ID NO: 219	2010 bp
	ORF Start: ATG at 40	ORF Stop: TGA at 1828

TTTGAGTTAGACAAGCAGCAGCACACGCCTCCCTACCTC**ATG**GCGACAGAAAATGGAGCAGTTGAGCT
GGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGGTCCCACAGGTGAAAGACCCCTGGCTGCAG
GGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGGATCCACCCACCCTGAAGAAAGATGCCAAA
GCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAACCCTCAACTAGCAGCCAAGGCCCCAAAGG
AGAGGGTGACAGGGGCGGGGGGCCCCGCGGAGGGCAGTGCTGGGCCCCCGGCAGCCCTGCCCCAGCAGA
CTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTGAGCAGGGAGCCTCAGGCAGCCAGGATCCT
GGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAAGCAGCAGCCAGGAGGGGCTCACCTGCCTT
TCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTCTGAGAAGCTGCTGGCCAAGAAGCCCCCAA
GCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGACCCACAGCCCCACGGATCCCAGGCCAGCC
AAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAGAAGGAAGTGGGAGAGAAAACCCCAGGCCA
GGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGGGATTGAGTTCCAGGCTGTTCCCTCAGAGA
AATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGGAGGAGGACTGCTTCCAGATTTTGGATGAT
TGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTGGAGCTGAGGACCGGGAATGTCAGCAGTGA
ATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGTGGCAAGTTTGGGGCAGTCTGTACCTGCATGGAGA
AAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGAAACAGACTCCCAAAGACAAGGAAATGGTG
TTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGCAATCTGATCCAGCTGTATGCAGCCATCGA
GACTCCGCATGAGATCGTCCTGTTCATGGAGTACATCGAGGGCGGAGAGCTCTTCGAGAGGATTGTGG
ATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTGTCAGGCAGATCTGTGACGGGATCCTC
TTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCAGAGAACATCCTGTGTGTCAACACCAC
CGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAGGTATAACCCCAACGAGAAGCTGAAGG
TGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGCGGTGAATTATGACCAAATCTCCGATAAGACA
GACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTGAGCGGCCTCTCCCCCTTCCTGGGAGATGA
TGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTGGTACTTTGATGAAGAGACCTTTGAGGCCG
TATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCGTCAAGGACCAGAGGGCCCGGATGAACGCT
GCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAGAAAGCCAAACGCTGTAACCGACGCCT
TAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCGCTGGAAGAAAAACTTCATTGCTGTCA
GCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCACTGATGGCTCTGGGGGT**CTGA**GCCCTG
GGCGCAGCTGAAGCCTGGACGCAGCCACACAGTGGCCGGGGCTGAAGCCACACAGCCCAGAAGGCCAG
AAAAGGCAGCCAGATCCCCAGGGCAGCCTCGTTAGGACAAGGCTGTGCCAGGCTGGGAGGCTCGGGGC
TCCCCACGCCCCCATGCAGTGACCGCTTCCCCGATGTG

NOV18e, CG51448-02 Protein Sequence	SEQ ID NO: 220	596 aa	MW at 64656.1kD

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEYIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLK
PENILCVNTTGHLVKIIDFGLARRYNPNEKLKVNFGTPEFLSPEAVNYDQISDKTDMWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQCLAHPWLNNLA
EKAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18f, CG51448-03 DNA Sequence	SEQ ID NO: 221	1839 bp
	ORF Start: ATG at 49	ORF Stop: TGA at 1837

CTAGAAAGACTTGAGTTAGACAAGCAGCAGCACACGCCTCCCTACCTC**A**TGGCGACAGAAAATGGAGC
AGTTGAGCTGGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGGTCCCACAGGTGAAAGACCCC
TGGCTGCAGGGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGGATCCACCCACCCTGAAGAAA

GATGCCAAAGCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAACCCTCAACTAGCAGCCAAGG
CCCCAAAGGAGAGGGTGACAGGGGCGGGGGGCCCGCGGAGGGCAGTGCTGGGCCCCCGGCAGCCCTGC
CCCAGCAGACTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTGAGCAGGGAGCCTCAGGCAGC
CAGGATCCTGGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAAGCAGCAGCCAGGAGGGGCTC
ACCTGCCTTTCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTCTGAGAAGCTGCTGGCCAAGA
AGCCCCCAAGCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGACCCACAGCCCCACGGATCCC
AGGCCAGCCAAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAGAAGGAAGTGGGAGAGAAAAC
CCCAGGCCAGGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGGGATTGAGTTCCAGGCTGTTC
CCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGGAGGAGGACTGCTTCCAGATT
TTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTGGAGCTGAGGACCGGGAATGT
CAGCAGTGAATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGTGGCAAGTTTGGGGCAGTCTGTACCT
GCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGAAACAGACTCCCAAAGACAAG
GAAATGGTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGCAATCTGATCCAGCTGTATGC
AGCCATCGAGACTCCGCATGAGATCGTCCTGTTCATGGAGTACATCGAGGGCGGAGAGCTCTTCGAGA
GGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTGTCAGGCAGATCTGTGAC
GGGATCCTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCAGAGAACATCCTGTGTGT
CAACACCACCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAGGTATAACCCCAACGAGA
AGCTGAAGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGCGGTGAATTATGACCAAATCTCC
GATAAGACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTGAGCGGCCTCTCCCCCTTCCT
GGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTGGTACTTTGATGAAGAGACCT
TTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCGTCAAGGACCAGAGGGCCCGG
ATGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAGAAAGCCAAACGCTGTAA
CCGACGCCTTAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCGCTGGAAGAAAAACTTCA
TTGCTGTCAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCACTGATGGCTCTGGGGGTC
TGA

NOV18f, CG51448-03 Protein Sequence	SEQ ID NO: 222	596 aa	MW at 64656.1kD

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEYIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLK
PENILCVNTTGHLVKIIDFGLARRYNPNEKLKVNFGTPEFLSPEAVNYDQISDKTDMWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQCLAHPWLNNLA
EKAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18g, CG51448-04 DNA Sequence	SEQ ID NO: 223	2558 bp
	ORF Start: ATG at 164	ORF Stop: TGA at 1949

CTTTGCTCCAGGTACCTCTCTCCCCTCAGTTAGCAGGCCTCGGCTTCCTGTCTCACTGCAGCCAGACG
AGAGGGGAAATTGGACAGCCTGACACACTCCACTCTTGTTTCTGCAGCTAGAAAGACTTGAGTTAGAC
AAGCAGCAGCACACGCCTCCCTACCTC**ATG**GCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAA
CCCATCAACAGACAAGGCACCTAAAGGTCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAAGACCCTG
GCCCCCCAGACCCAAAGAAAGCTCCGGATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCA
GAGAAAGGGGATGGTACCCTGGCCCAACCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAG
GGGCGGGGGGCCCGCGGAGGGCAGTGCTGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCGACACCTG
AGACCAGCGTCAAGAAGCCCAAGGCTGAGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCCAGG
GTGGGCAAGAAGGCAGCAGAGGGCCAAGCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCC
CAGCTGTCCCGCCATCATCTCCAGTTCTGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAG
AGCTCACCTTTGAAGGGGTGCCCATGACCCACAGCCCCACGGATCCCAGGTCGGCCAAGGCAGAAGAA
GGAAAGAACATCCTGGCAGAGAGCCAGAAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGC
TAAGATGCAAGGGGACACCTCGAGGGGGATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGG
GGCAGGCCCTCTGTCTCACAGCCAGGGAGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCT
CCGGCCCCCTTCCCTCACCGCATGGTGGAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAA
CTCCAAGGAGGCGCTCGGAGGGGGCAAGTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCC
TCAAGCTGGCAGCCAAGGTCATCAAGAAACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATT
GAGGTCATGAACCAGCTGAACCACCGCAATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGA
GATCGTCCTGTTCATGGAGATCGAGGGCGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATC

TGACCGAGGTGGACACCATGGTGTTTGTCAGGCAGATCTGTGACGGGATCCTCTTGATGCACAAGATG
AGGGTTTTGCACCTGGACCTCAAGCCAGAGAACATCCTGTGTGTCAACACCACCGGGCATTTGGTGAA
GATCATTGACTTTGGCCTGGCACGGAGGTATAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCC
CAGAGTTCCTGTCACCTGAGGTGGTGAATTATGACCAAATCTCCGATAAGACAGACATGTGGAGTATG
GGGGTGATCACCTACATGCTGCTGAGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCT
AAACAACGTTCTATCTGGCAACTGGTACTTTGATGAAGAGACCTTTGAGGCCGTATCAGACGAGGCCA
AAGACTTTGTCTCCAACCTCATCGTCAAGGACCAGAGGGCCCGGATGAACGCTGCCCAGTGTCTCGCC
CATCCCTGGCTCAACAACCTGGCGGAGAAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTT
GCTTAAGAAATACCTCATGAAGAGGCGCTGGAAGAAAAACTTCATTGCTGTCAGCGCTGCCAACCGCT
TCAAGAAGATCAGCAGCTCGGGGGCACTGATGGCTCTGGGGGTCTGAGCCCTGGGCGCAGCTGAAGCC
TGGACGCAGCCACACAGTGGCCGGGGCTGAAGCCACACAGCCCAGAAGGCCAGAAAAGGCAGCCAGAT
CCCCAGGGCAGCCTCGTTAGGACAAGGCTGTGCCAGGCTGGGAGGCTCGGGGCTCCCCACGCCCCCAT
GCAGTGACCGCTTCCCCGATGTGAGCCGCCTCGGAGTGTGGCCTGGATCCATCCTGCTAGCACCTCCC
CAGACAGGGCTCCAGCCTGTCGGCCACACCCCAGACTCCAGGCCCCCGTTGAAGCCGCTCCCGGTTCC
CTCCCCAGCTCCTCGTCTTTGAACTGCCGCCGCCGTGGTGACCCCTGCTTTGCCCCACTGGGAGAGTC
CTTAGCCTGGGCCTCCTCCTAGCTGGAGTGCCATGGCTGGGGGGTCTCAGCATGTAGGGCTTCTGTGG
TTGTGGATGGGAGGCTCCTGGTGGGGCAGAAAGGCTGCAACGCTGATTCCTAAGGCCCAGCTGCCAGG
GAAGACAGAGCAGGCTTTGTGAGAGAGGACCTCCATGCCCCCGCCACCTCCCCACTCCAGCAGATAAG
GCCGAGCCCACACCATCTGGCCCAGGCTGGCCCCCACCACCT

NOV18g, CG51448-04 Protein Sequence	SEQ ID NO: 224	595 aa	MW at 64476.9kD

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRSAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILLMHKMRVLHLDLKP
ENILCVNTTGHLVKIIDFGLARRYNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYMLLS
GLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQCLAHPWLNNLAE
KAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18h, SNP13375535 of CG51448-01, DNA Sequence	SEQ ID NO: 225	1788 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1786
	SNP Pos: 102	SNP Change: A to G

ATGGCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGG
TCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAGGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGG
ATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAA
CCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAGGGGCGGGGGGCCCGCGGAGGGCAGTGC
TGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTG
AGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAA
GCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTC
TGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGA
CCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAG
AAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGG
GATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGG
AGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTG
GAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGGGGCAA
GTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGA
AACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGC
AATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCCTGTTCATGGAGATCGAGGG
CGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTG
TCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCA
GAGAACATCCTGTGTGTCAACACCACCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAG
GTACCACAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGTGG
TGAATTATGACCAAATCTCCGATAAGACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTG
AGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTG
GTACTTTGATGAAGAGACCTTTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCG

TCAAGGACCAGGCCCGGATGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAG
AAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCG
CTGGAAGAAAAACTTCATTGCTGTCAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCAC
TGATGGCTCTGGGGGTC**TGA**

NOV18h, SNP13375535 of CG51448-01, Protein Sequence	SEQ ID NO: 226	595 aa	MW at 64501.9kD
	SNP Pos: 34		SNP Change: Lys to Lys

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKP
ENILCVNTTGHLVKIIDFGLARRYHNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQARMNAAQCLAHPWLNNLAE
KAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18i, SNP13375536 of CG51448-01, DNA Sequence	SEQ ID NO: 227	1788 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1786
	SNP Pos: 287	SNP Change: C to T

ATGGCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGG
TCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGG
ATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAA
CCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAGGGGCGGGGGGCCCGCGGAGGGCAGTGC
TGGGCCCCCGGCAGTCCTGCCCCAGCAGACTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTG
AGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAA
GCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTC
TGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGA
CCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAG
AAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGG
GATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGG
AGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTG
GAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGGGGCAA
GTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGA
AACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGC
AATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCCTGTTCATGGAGATCGAGGG
CGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTG
TCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCA
GAGAACATCCTGTGTGTCAACACCACCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAG
GTACCACAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGTGG
TGAATTATGACCAAATCTCCGATAAGACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTG
AGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTG
GTACTTTGATGAAGAGACCTTTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCG
TCAAGGACCAGGCCCGGATGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAG
AAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCG
CTGGAAGAAAAACTTCATTGCTGTCAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCAC
TGATGGCTCTGGGGGTC**TGA**

NOV18i, SNP13375536 of CG51448-01, Protein Sequence	SEQ ID NO: 228	595 aa	MW at 64530.0kD
	SNP Pos: 96		SNP Change: Ala to Val

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAVLPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKP
ENILCVNTTGHLVKIIDFGLARRYHNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQARMNAAQCLAHPWLNNLAE

KAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18j, SNP13375537 of CG51448-01, DNA Sequence	SEQ ID NO: 229	1788 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1786
	SNP Pos: 350	SNP Change: C to T

ATGGCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGG
TCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGG
ATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAA
CCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAGGGGCGGGGGGCCCGCGGAGGGCAGTGC
TGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTG
AGCAGGGAGTCTCAGGCAGCCAGGATCCTGGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAA
GCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTC
TGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGA
CCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAG
AAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGG
GATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGG
AGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTG
GAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGGGGCAA
GTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGA
AACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGC
AATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCCTGTTCATGGAGATCGAGGG
CGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTG
TCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCA
GAGAACATCCTGTGTGTCAACACCACCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAG
GTACCACAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGTGG
TGAATTATGACCAAATCTCCGATAAGACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTG
AGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTG
GTACTTTGATGAAGAGACCTTTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCG
TCAAGGACCAGGCCCGGATGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAG
AAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCG
CTGGAAGAAAAACTTCATTGCTGTCAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCAC
TGATGGCTCTGGGGGTCTGA

NOV18j, SNP13375537 of CG51448-01, Protein Sequence	SEQ ID NO: 230	595 aa	MW at 64530.0kD
	SNP Pos: 117		SNP Change: Ala to Val

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGVSGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKP
ENILCVNTTGHLVKIIDFGLARRYHNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQARMNAAQCLAHPWLNNLAE
KAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18k, SNP13375538 of CG51448-01, DNA Sequence	SEQ ID NO: 231	1788 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1786
	SNP Pos: 1249	SNP Change: A to T

ATGGCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGG
TCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGG
ATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAA
CCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAGGGGCGGGGGGCCCGCGGAGGGCAGTGC
TGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTG
AGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAA
GCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTC
TGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGA
CCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAG

AAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGG
GATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGG
AGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTG
GAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGGGGCAA
GTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGA
AACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGC
AATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCCTGTTCATGGAGATCGAGGG
CGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTG
TCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCA
GAGAACATCCTGTGTGTCAACACC**T**CCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAG
GTACCACAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGTGG
TGAATTATGACCAAATCTCCGATAAGACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTG
AGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTG
GTACTTTGATGAAGAGACCTTTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCG
TCAAGGACCAGGCCCGGATGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAG
AAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCG
CTGGAAGAAAAACTTCATTGCTGTCAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCAC
TGATGGCTCTGGGGGTC**TGA**

NOV18k, SNP13375538 of CG51448-01, Protein Sequence	SEQ ID NO: 232	595 aa	MW at 64487.9kD
	SNP Pos: 417		SNP Change: Thr to Ser

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKP
ENILCVNT**S**GHLVKIIDFGLARRYHNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQARMNAAQCLAHPWLNNLAE
KAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18l, SNP13375539 of CG51448-01, DNA Sequence	SEQ ID NO: 233	1788 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1786
	SNP Pos: 1358	SNP Change: T to C

ATGGCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGG
TCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGG
ATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAA
CCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAGGGGCGGGGGGCCCGCGGAGGGCAGTGC
TGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTG
AGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAA
GCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTC
TGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGA
CCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAG
AAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGG
GATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGG
AGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTG
GAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGGGGCAA
GTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGA
AACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGC
AATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCCTGTTCATGGAGATCGAGGG
CGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTG
TCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCA
GAGAACATCCTGTGTGTCAACACCACCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAG
GTACCACAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGG**C**GG
TGAATTATGACCAAATCTCCGATAAGACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTG
AGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTG
GTACTTTGATGAAGAGACCTTTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCG
TCAAGGACCAGGCCCGGATGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAG

AAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCG
CTGGAAGAAAAACTTCATTGCTGTCAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCAC
TGATGGCTCTGGGGGTC**TGA**

NOV18l, SNP13375539 of CG51448-01, Protein Sequence	SEQ ID NO: 234	595 aa	MW at 64473.8kD
	SNP Pos: 453		SNP Change: Val to Ala

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKP
ENILCVNTTGHLVKIIDFGLARRYHNPNEKLKVNFGTPEFLSPE**A**VNYDQISDKTDMWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNL**I**VKDQARMNAAQCLAHPWLNNLAE
KAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18m, SNP13375540 of CG51448-01, DNA Sequence	SEQ ID NO: 235	1788 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1786
	SNP Pos: 1391	SNP Change: A to G

ATGGCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGG
TCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGG
ATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAA
CCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAGGGGCGGGGGGCCCGCGGAGGGCAGTGC
TGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTG
AGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAA
GCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTC
TGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGA
CCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAG
AAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGG
GATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGG
AGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTG
GAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGGGGCAA
GTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGA
AACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGC
AATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCCTGTTCATGGAGATCGAGGG
CGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTG
TCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCA
GAGAACATCCTGTGTGTCAACACCACCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAG
GTACCACAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGTGG
TGAATTATGACCAAATCTCCGATAAGACAGGCATGTGGAGTATGGGGGTGATCACCTACATGCTGCTG
AGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTG
GTACTTTGATGAAGAGACCTTTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCG
TCAAGGACCAGGCCCGGATGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAG
AAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCG
CTGGAAGAAAAACTTCATTGCTGTCAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCAC
TGATGGCTCTGGGGGTC**TGA**

NOV18m, SNP13375540 of CG51448-01, Protein Sequence	SEQ ID NO: 236	595 aa	MW at 64443.9kD
	SNP Pos: 464		SNP Change: Asp to Gly

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKP
ENILCVNTTGHLVKIIDFGLARRYHNPNEKLKVNFGTPEFLSPEVVNYDQISDKT**G**MWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQARMNAAQCLAHPWLNNLAE
KAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18n, SNP13375541 of CG51448-01, DNA Sequence	SEQ ID NO: 237	1788 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1786
	SNP Pos: 1511	SNP Change: A to G

ATGGCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGG
TCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGG
ATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAA
CCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAGGGGCGGGGGGCCCGCGGAGGGCAGTGC
TGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTG
AGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAA
GCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTC
TGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGA
CCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAG
AAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGG
GATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGG
AGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTG
GAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGGGGCAA
GTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGA
AACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGC
AATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCCTGTTCATGGAGATCGAGGG
CGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTG
TCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCA
GAGAACATCCTGTGTGTCAACACCACCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAG
GTACCACAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGTGG
TGAATTATGACCAAATCTCCGATAAGACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTG
AGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTG
GTACTTTGATGAAGGGACCTTTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCG
TCAAGGACCAGGCCCGGATGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAG
AAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCG
CTGGAAGAAAAACTTCATTGCTGTCAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCAC
TGATGGCTCTGGGGGTC**TGA**

NOV18n, SNP13375541 of CG51448-01, Protein Sequence	SEQ ID NO: 238	595 aa	MW at 64429.8kD
	SNP Pos: 504		SNP Change: Glu to Gly

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKP
ENILCVNTTGHLVKIIDFGLARRYHNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEGTFEAVSDEAKDFVSNLIVKDQARMNAAQCLAHPWLNNLAE
KAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18o, SNP13375542 of CG51448-01, DNA Sequence	SEQ ID NO: 239	1788 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1786
	SNP Pos: 1675	SNP Change: C to T

ATGGCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGG
TCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGG
ATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAA
CCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAGGGGCGGGGGGCCCGCGGAGGGCAGTGC
TGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTG
AGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAA
GCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTC
TGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGA
CCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAG
AAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGG
GATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGG

AGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTG
GAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGGGGCAA
GTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGA
AACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGC
AATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCCTGTTCATGGAGATCGAGGG
CGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTG
TCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCA
GAGAACATCCTGTGTGTCAACACCACCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAG
GTACCACAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGTGG
TGAATTATGACCAAATCTCCGATAAGACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTG
AGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTG
GTACTTTGATGAAGAGACCTTTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCG
TCAAGGACCAGGCCCGGATGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAG
AAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTTGTTTAAGAAATACCTCATGAAGAGGCG
CTGGAAGAAAAACTTCATTGCTGTCAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCAC
TGATGGCTCTGGGGGTCTGA

NOV18o, SNP13375542 of CG51448-01, Protein Sequence	SEQ ID NO: 240	595 aa	MW at 64535.9kD
	SNP Pos: 559		SNP Change: Leu to Phe

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKP
ENILCVNTTGHLVKIIDFGLARRYHNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQARMNAAQCLAHPWLNNLAE
KAKRCNRRLKSQILFKKYLMKRRWKKNFIAVSAANRFKKISSSGALMALGV

NOV18p, SNP13375543 of CG51448-01, DNA Sequence	SEQ ID NO: 241	1788 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1786
	SNP Pos: 1724	SNP Change: T to G

ATGGCGACAGAAAATGGAGCAGTTGAGCTGGGAATTCAGAACCCATCAACAGACAAGGCACCTAAAGG
TCCCACAGGTGAAAGACCCCTGGCTGCAGGGAAAGACCCTGGCCCCCCAGACCCAAAGAAAGCTCCGG
ATCCACCCACCCTGAAGAAAGATGCCAAAGCCCCTGCCTCAGAGAAAGGGGATGGTACCCTGGCCCAA
CCCTCAACTAGCAGCCAAGGCCCCAAAGGAGAGGGTGACAGGGGCGGGGGGCCCGCGGAGGGCAGTGC
TGGGCCCCCGGCAGCCCTGCCCCAGCAGACTGCGACACCTGAGACCAGCGTCAAGAAGCCCAAGGCTG
AGCAGGGAGCCTCAGGCAGCCAGGATCCTGGAAAGCCCAGGGTGGGCAAGAAGGCAGCAGAGGGCCAA
GCAGCAGCCAGGAGGGGCTCACCTGCCTTTCTGCATAGCCCCAGCTGTCCTGCCATCATCTCCAGTTC
TGAGAAGCTGCTGGCCAAGAAGCCCCCAAGCGAGGCATCAGAGCTCACCTTTGAAGGGGTGCCCATGA
CCCACAGCCCCACGGATCCCAGGCCAGCCAAGGCAGAAGAAGGAAAGAACATCCTGGCAGAGAGCCAG
AAGGAAGTGGGAGAGAAAACCCCAGGCCAGGCTGGCCAGGCTAAGATGCAAGGGGACACCTCGAGGGG
GATTGAGTTCCAGGCTGTTCCCTCAGAGAAATCCGAGGTGGGGCAGGCCCTCTGTCTCACAGCCAGGG
AGGAGGACTGCTTCCAGATTTTGGATGATTGCCCGCCACCTCCGGCCCCCTTCCCTCACCGCATGGTG
GAGCTGAGGACCGGGAATGTCAGCAGTGAATTCAGTATGAACTCCAAGGAGGCGCTCGGAGGGGGCAA
GTTTGGGGCAGTCTGTACCTGCATGGAGAAAGCCACAGGCCTCAAGCTGGCAGCCAAGGTCATCAAGA
AACAGACTCCCAAAGACAAGGAAATGGTGTTGCTGGAGATTGAGGTCATGAACCAGCTGAACCACCGC
AATCTGATCCAGCTGTATGCAGCCATCGAGACTCCGCATGAGATCGTCCTGTTCATGGAGATCGAGGG
CGGAGAGCTCTTCGAGAGGATTGTGGATGAGGACTACCATCTGACCGAGGTGGACACCATGGTGTTTG
TCAGGCAGATCTGTGACGGGATCCTCTTCATGCACAAGATGAGGGTTTTGCACCTGGACCTCAAGCCA
GAGAACATCCTGTGTGTCAACACCACCGGGCATTTGGTGAAGATCATTGACTTTGGCCTGGCACGGAG
GTACCACAACCCCAACGAGAAGCTGAAGGTGAACTTTGGGACCCCAGAGTTCCTGTCACCTGAGGTGG
TGAATTATGACCAAATCTCCGATAAGACAGACATGTGGAGTATGGGGGTGATCACCTACATGCTGCTG
AGCGGCCTCTCCCCCTTCCTGGGAGATGATGACACAGAGACCCTAAACAACGTTCTATCTGGCAACTG
GTACTTTGATGAAGAGACCTTTGAGGCCGTATCAGACGAGGCCAAAGACTTTGTCTCCAACCTCATCG
TCAAGGACCAGGCCCGGATGAACGCTGCCCAGTGTCTCGCCCATCCCTGGCTCAACAACCTGGCGGAG
AAAGCCAAACGCTGTAACCGACGCCTTAAGTCCCAGATCTTGCTTAAGAAATACCTCATGAAGAGGCG
CTGGAAGAAAAACTTCATTGCTGGCAGCGCTGCCAACCGCTTCAAGAAGATCAGCAGCTCGGGGGCAC

TGATGGCTCTGGGGGTCTGA			
NOV18p, SNP13375543 of CG51448-01, Protein Sequence	SEQ ID NO: 242	595 aa	MW at 64459.8kD
	SNP Pos: 575		SNP Change: Val to Gly

MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAKAPASEKGDGTLAQ
PSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQGASGSQDPGKPRVGKKAAEGQ
AAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELTFEGVPMTHSPTDPRPAKAEEGKNILAESQ
KEVGEKTPGQAGQAKMQGDTSRGIEFQAVPSEKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMV
ELRTGNVSSEFSMNSKEALGGGKFGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHR
NLIQLYAAIETPHEIVLFMEIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKP
ENILCVNTTGHLVKIIDFGLARRYHNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYMLL
SGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQARMNAAQCLAHPWLNNLAE
KAKRCNRRLKSQILLKKYLMKRRWKKNFIAGSAANRFKKISSSGALMALGV

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 18B.

Table 18B. Comparison of the NOV18 protein sequences.

```
NOV18a  ------------------------------------------------------------
NOV18b  -----MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAK
NOV18c  ------------------------------------------------------------
NOV18d  TKLPTMATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAK
NOV18e  -----MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAK
NOV18f  -----MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAK
NOV18g  -----MATENGAVELGIQNPSTDKAPKGPTGERPLAAGKDPGPPDPKKAPDPPTLKKDAK

NOV18a  -------------------MPKPLPQRKG-MVPWPNPQLAALPQQTATPETSVKKPKAEQ
NOV18b  APASEKGDGTLAQPSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQ
NOV18c  ------------------------------------------------------------
NOV18d  APASEKGDGTLAQPSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQ
NOV18e  APASEKGDGTLAQPSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQ
NOV18f  APASEKGDGTLAQPSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQ
NOV18g  APASEKGDGTLAQPSTSSQGPKGEGDRGGGPAEGSAGPPAALPQQTATPETSVKKPKAEQ

NOV18a  GASGSQDPGKPRVGKKAAEGQAAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELT
NOV18b  GASGSQDPGKPRVGKKAAEGQAAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELT
NOV18c  ------------------------------------------------------------
NOV18d  GASGSQDPGKPRVGKKAAEGQAAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELT
NOV18e  GASGSQDPGKPRVGKKAAEGQAAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELT
NOV18f  GASGSQDPGKPRVGKKAAEGQAAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELT
NOV18g  GASGSQDPGKPRVGKKAAEGQAAARRGSPAFLHSPSCPAIISSSEKLLAKKPPSEASELT

NOV18a  FEGVPMTHSPTDPRPAKAEEGKNILAESQKEVGEKTPGQAGQAKMQGDTSRGIEFQAVPS
NOV18b  FEGVPMTHSPTDPRPAKAEEGKNILAESQKEVGEKTPGQAGQAKMQGDTSRGIEFQAVPS
NOV18c  ------------------------------------------------------------
NOV18d  FEGVPMTHSPTDPRPAKAEEGKNILAESQKEVGEKTPGQAGQAKMQGDTSRGIEFQAVPS
NOV18e  FEGVPMTHSPTDPRPAKAEEGKNILAESQKEVGEKTPGQAGQAKMQGDTSRGIEFQAVPS
NOV18f  FEGVPMTHSPTDPRPAKAEEGKNILAESQKEVGEKTPGQAGQAKMQGDTSRGIEFQAVPS
NOV18g  FEGVPMTHSPTDPRSAKAEEGKNILAESQKEVGEKTPGQAGQAKMQGDTSRGIEFQAVPS

NOV18a  EKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMVELRTGNVSSEFSMNSKEALGGGK
NOV18b  EKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMVELRTGNVSSEFSMNSKEALGGGK
NOV18c  ----------------------------------------------TKLMNSKEALGGGK
```

```
NOV18d  EKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMVELRTGNVSSEFSMNSKEALGGGK
NOV18e  EKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMVELRTGNVSSEFSMNSKEALGGGK
NOV18f  EKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMVELRTGNVSSEFSMNSKEALGGGK
NOV18g  EKSEVGQALCLTAREEDCFQILDDCPPPPAPFPHRMVELRTGNVSSEFSMNSKEALGGGK

NOV18a  FGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHRNLIQLYAAIETPHEI
NOV18b  FGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHRNLIQLYAAIETPHEI
NOV18c  FGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHRNLIQLYAAIETPHEI
NOV18d  FGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHRNLIQLYAAIETPHEI
NOV18e  FGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHRNLIQLYAAIETPHEI
NOV18f  FGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHRNLIQLYAAIETPHEI
NOV18g  FGAVCTCMEKATGLKLAAKVIKKQTPKDKEMVLLEIEVMNQLNHRNLIQLYAAIETPHEI

NOV18a  VLFMEYIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKPENILCV
NOV18b  VLFMEIE-GGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKPENILCV
NOV18c  VLFMEYIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKPENILCV
NOV18d  VLFMEYIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKPENILCV
NOV18e  VLFMEYIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKPENILCV
NOV18f  VLFMEYIEGGELFERIVDEDYHLTEVDTMVFVRQICDGILFMHKMRVLHLDLKPENILCV
NOV18g  VLFMEIE-GGELFERIVDEDYHLTEVDTMVFVRQICDGILLMHKMRVLHLDLKPENILCV

NOV18a  NTTGHLVKIIDFGLARRYN-PNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYM
NOV18b  NTTGHLVKIIDFGLARRYHNPNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYM
NOV18c  NTTGHLVKIIDFGLARRYN-PNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYM
NOV18d  NTTGHLVKIIDFGLARRYN-PNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYM
NOV18e  NTTGHLVKIIDFGLARRYN-PNEKLKVNFGTPEFLSPEAVNYDQISDKTDMWSMGVITYM
NOV18f  NTTGHLVKIIDFGLARRYN-PNEKLKVNFGTPEFLSPEAVNYDQISDKTDMWSMGVITYM
NOV18g  NTTGHLVKIIDFGLARRYN-PNEKLKVNFGTPEFLSPEVVNYDQISDKTDMWSMGVITYM

NOV18a  LLSGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQC
NOV18b  LLSGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQ-ARMNAAQC
NOV18c  LLSGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQC
NOV18d  LLSGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQC
NOV18e  LLSGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQC
NOV18f  LLSGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQC
NOV18g  LLSGLSPFLGDDDTETLNNVLSGNWYFDEETFEAVSDEAKDFVSNLIVKDQRARMNAAQC

NOV18a  LAHPWLNNLAEKAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMAL
NOV18b  LAHPWLNNLAEKAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMAL
NOV18c  LAHPWLVDG----------------------------------------------------
NOV18d  LAHPWLNNLAEKAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMAL
NOV18e  LAHPWLNNLAEKAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMAL
NOV18f  LAHPWLNNLAEKAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMAL
NOV18g  LAHPWLNNLAEKAKRCNRRLKSQILLKKYLMKRRWKKNFIAVSAANRFKKISSSGALMAL

NOV18a  GV---
NOV18b  GV---
NOV18c  -----
NOV18d  GVVDG
NOV18e  GV---
NOV18f  GV---
NOV18g  GV---

NOV18a  (SEQ ID NO:  212)
```

NOV18b (SEQ ID NO: 214) NOV18c (SEQ ID NO: 216) NOV18d (SEQ ID NO: 218) NOV18e (SEQ ID NO: 220) NOV18f (SEQ ID NO: 222) NOV18g (SEQ ID NO: 224)

Further analysis of the NOV18a protein yielded the following properties shown in Table 18C.

Table 18C. Protein Sequence Properties NOV18a

SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	

```
PSG:  a new signal peptide prediction method
      N-region:  length 9;  pos.chg 3;  neg.chg 0
      H-region:  length 20;  peak value   1.84
      PSG score:  -2.56

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1): -10.69
      possible cleavage site: between 29 and 30

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:    0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  1.48 (at 393)
      ALOM score:   1.48  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75): 14.68
      Hyd Moment(95):  9.57    G content:       1
      D/E content:     1       S/T content:     2
      Score: -2.37

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 18  QRK|GM

NUCDISC: discrimination of nuclear localization signals
      pat4: KKPK (4) at   34
      pat7: none
      bipartite: RRLKSQILLKKYLMKRR at  477
      content of basic residues:  12.9%
      NLS Score:  0.27

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none
```

```
SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  found
      RIVDEDYHL at 295

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 55.5

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       69.6 %: nuclear
       17.4 %: mitochondrial
       13.0 %: cytoplasmic

>> prediction for CG51448-05 is nuc (k=23)
```

A search of the NOV18a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 18D.

Table 18D. Geneseq Results for NOV18a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV18a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAE22849	Human cardiac myosin light chain kinase (cMLCK) mutant, G89D - Homo sapiens, 596 aa. [WO200224889-A2, 28-MAR-2002]	20..521 95..596	502/502 (100%) 502/502 (100%)	0.0
AAE22847	Human cardiac myosin light chain kinase (cMLCK) mutant, A87V - Homo sapiens, 596 aa. [WO200224889-A2, 28-MAR-2002]	20..521 95..596	502/502 (100%) 502/502 (100%)	0.0
AAE22723	Human cardiac myosin light chain kinase (cMLCK) protein - Homo sapiens, 596 aa. [WO200224889-A2, 28-MAR-2002]	20..521 95..596	502/502 (100%) 502/502 (100%)	0.0
AAU03521	Human protein kinase #21 - Homo sapiens, 612 aa. [WO200138503-A2, 31-MAY-2001]	20..521 111..612	502/502 (100%) 502/502 (100%)	0.0
AAE16340	Human POLY4 protein - Homo sapiens, 596 aa. [WO200185767-A2, 15-NOV-2001]	20..521 95..596	501/502 (99%) 501/502 (99%)	0.0

In a BLAST search of public sequence databases, the NOV18a protein was found to have homology to the proteins shown in the BLASTP data in Table 18E.

Table 18E. Public BLASTP Results for NOV18a				
Protein Accession Number	**Protein/Organism/Length**	**NOV18a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
CAC88318	Sequence 7 from Patent WO0164905 - Homo sapiens (Human), 596 aa.	20..521 95..596	502/502 (100%) 502/502 (100%)	0.0
Q9H1R3	Myosin light chain kinase 2, skeletal/cardiac muscle (EC 2.7.1.117) (MLCK2) - Homo sapiens (Human), 595 aa.	20..521 94..595	502/502 (100%) 502/502 (100%)	0.0
A35021	myosin-light-chain kinase (EC 2.7.1.117), skeletal muscle - rabbit, 608 aa.	20..521 105..608	462/504 (91%) 479/504 (94%)	0.0
P07313	Myosin light chain kinase 2, skeletal/cardiac muscle (EC 2.7.1.117) (MLCK2) - Oryctolagus cuniculus (Rabbit), 607 aa.	20..521 104..607	462/504 (91%) 479/504 (94%)	0.0
A28798	myosin-light-chain kinase (EC 2.7.1.117), skeletal muscle - rat, 610 aa.	20..521 112..610	429/502 (85%) 453/502 (89%)	0.0

PFam analysis inidcates that the NOV18a protein contains the domains shown in the Table 18F.

Table 18F. Domain Analysis of NOV18a			
Pfam Domain	**NOV18a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
pkinase	210..465	93/297 (31%) 198/297 (67%)	9.3e-76

Example 19.

The NOV19 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 19A.

Table 19A. NOV19 Sequence Analysis

NOV19a, CG51752-02 DNA Sequence	SEQ ID NO: 243	1042 bp
	ORF Start: ATG at 207	ORF Stop: TGA at 1008

AGAGTGCTCTAAACCCAGCTCGGCCTTTGCTGTATTAGACAGAAGCACCTCATTCATATCCCTGGGGC
CCCTGATGGTGCAGTGGTCTGGCTGTGGTCTGCACACCAGCTATTCTGTTTTGTTTTGTTTTGTTTTT
TCCTACCTTTTTCCAATCCTCACACCTTCTGATCAACAGCCCCAGTAGGGTTTAAAGGTCCTAGAGCT
AC**ATG**GGATTTAGGTTTCTGGGCACAGCCAATTCTGCCACTTTTGAGACTTCCCTTCCCCTTCCACTT
GCCCCTCTCTGGTTCTCTGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACGACTTAAC
TAGCCCATCCATGGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGAGCCAACA
TGGACAATGACATTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGTGCCCATC
TGCCTCCCCACGCAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCCAGACCAA
TGCTGCTGACAAAAACTCTGTGAAAACGGATCTGATGAAAGTGCCAATGGTCATCATGGACTGGGAGG
AGTGTTCAAAGATGTTTCCAAAACTTACCAAAAATATGCTGTGTGCCGGATACAAGAATGAGAGCTAT
GATGCCTGCAAGGGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGTGGTACCA
GGTGGGCATCATCAGCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACACCTCGTTGG
TGAACTACAACCTCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGAGAAAAGG
AGGACTTCTGTCAAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCAGCCCCAG
ATCCTGGCTCCTGCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTAC**TGA**TAATAAAATA
GAGGCTATTCTTTCAACCGAAA

NOV19a, CG51752-02 Protein Sequence	SEQ ID NO: 244	267 aa	MW at 29498.9kD

MGFRFLGTANSATFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKRANM
DNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADKNSVKTDLMKVPMVIMDWEE
CSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYTSLV
NYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILY

NOV19b, CG51752-03 DNA Sequence	SEQ ID NO: 245	888 bp	
	ORF Start: ATG at 79	ORF Stop: TGA at 880	

TTGTTTTTTCCTACCTTTTTCCAATCCTCACACCTTCTGATCAACAGCCCCAGTAGGGTTTAAAGGTC
CTAGAGCTAC**ATG**GGATTTAGGTTTCTGGGCACAGCCAATTCTGCCACTTTTGAGACTTCCCTTCCCC
TTCCACTTGCCCCTCTCTGGTTCTCTGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAAC
GACTTAACTAGCCCATCCATGGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAG
AGCCAACATGGACAATGACATTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGG
TGCCCATCTGCCTCCCCACGCAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGC
CAGACCAATGCTGCTGACAAAAACTCTGTGAAAACGGATCTGATGAAAGCGCCAATGGTCATCATGGA
CTGGGAGGAGTGTTCAAAGATGTTTCCAAAACTTACCAAAAATATGCTGTGTGCCGGATACAAGAATG
AGAGCTATGATGCCTGCAAGGGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAG
TGGTACCAGGTGGGCATCATAACCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACAC
CTCGTTGGTGAACTACAACCTCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAG
AGAAAAGGAGGACTTCTGTCAAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGC
AGCCCCAGATCCTGGCTCCCGCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTAC**TGA**TT
ATAA

NOV19b, CG51752-03 Protein Sequence	SEQ ID NO: 246	267 aa	MW at 29468.8kD

MGFRFLGTANSATFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKRANM
DNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADKNSVKTDLMKAPMVIMDWEE
CSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGEKWYQVGIITWGKSCGEKNTPGIYTSLV
NYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGSPVSGVPEPGSPRSWLPLCPLSHVLFRAILY

NOV19c, 175069825 DNA Sequence	SEQ ID NO: 247	816 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCACCATGGGATTTAGGTTTCTGGGCACAGCCAATTCTGCCGCTTTTGAGACTTCCCTTCCCCT
TCCACTTGCCCCTCTCTGGTTCTCTGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACG
ACTTAACTAGCCCATCCATGGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGA

GCCAACATGGACAATGACATTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGT
GCCCATCTGCCTCCCCACGCAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCC
AGACCAATGCTGCTGACAAAAACTCTGTGAAAACGGATCTGATGAAAGCGCCAATGGTCATCATGGAC
TGGGAGGAGTGTTCAAAGATGTTTCCAAAACTTACCAAAAATATGCTGTGTGCCGGATACAAGAATGA
GAGCTATGATGCCTGCAAGGGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGT
GGTACCAGGTGGGCATCATAAGCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACACC
TCGTTGGTGAACTACAACCTCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGA
GAAAAGGAGGACTTCTGTCAAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCA
GCCCCAGATCCTGGCTCCTGCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTACCTCGAG

NOV19c, 175069825 Protein Sequence	SEQ ID NO: 248	272 aa	MW at 29928.3kD

GSTMGFRFLGTANSAAFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKR
ANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADKNSVKTDLMKAPMVIMD
WEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYT
SLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILYLE

NOV19d, 175069842 DNA Sequence	SEQ ID NO: 249	816 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCACCATGGGATTTAGGTTTCTGGGCACAGCCAATTCTGCCGCTTTTGAGACTTCCCTTCCCCT
TCCACTTGCCCCTCTCTGGTTCTCTGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACG
ACTTAACTAGCCCATCCATGGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGA
GCCAACATGGACAATGACATTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGT
GCCCATCTGCCTCCCCACGCAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCC
AGACCAATGCTGCTGACAAAAACTCTGTGAAAACGGATCTGATGAAAGCGCCAATGGTCATCATGGAC
TGGGAGGAGTGTTCAAAGATGTTTCCAAAACTTACCAAAAATATGCTGCGTGCCGGATACAAGAATGA
GAGCTATGATGCCTGCAAGGGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGT
GGTACCAGGTGGGCATCATAAGCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACACC
TCGTTGGTGAACTACAACCTCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGA
GAAAAGGAGGACTTCTGTCAAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCA
GCCCCAGATCCTGGCTCCTGCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTACCTCGAG

NOV19d, 175069842 Protein Sequence	SEQ ID NO: 250	272 aa	MW at 29981.3kD

GSTMGFRFLGTANSAAFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKR
ANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADKNSVKTDLMKAPMVIMD
WEECSKMFPKLTKNMLRAGYKNESYDACKGDSGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYT
SLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILYLE

NOV19e, 258076315 DNA Sequence	SEQ ID NO: 251	729 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACGACTTAACTAGCCCATCCAT
GGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGAGCCAACATGGACAATGACA
TTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGTGCCCATCTGCCTCCCCACG
CAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCCAGACCAATGCTGCTGACAA
AAACTCTGTGAAAACGGATCTGATGAAAGCGCCAATGGTCATCATGGACTGGGAGGAGTGTTCAAAGA
TGTTTCCAAAACTTACCAAAAATATGCTGTGTGCCGGATACAAGAATGAGAGCTATGATGCCTGCAAG
GGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGTGGTACCAGGTGGGCATCAT
CAGCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACACCTCGTTGGTGAACTACAACC
TCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGAGAAAAGGAGGACTTCTGTC
AAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCAGCCCCAGATCCTGGCTCCT
GCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTACCTCGAG

NOV19e, 258076315 Protein Sequence	SEQ ID NO: 252	243 aa	MW at 26802.7kD

GSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPT
QPGPATWRECWVAGWGQTNAADKNSVKTDLMKAPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACK
GDSGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSV

KQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILYLE		
NOV19f, 258076366 DNA Sequence	SEQ ID NO: 253	729 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACGACTTAACTAGCCCATCCAT
GGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGAGCCAACATGGACAATGACA
TTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGTGCCCATCTGCCTCCCCACG
CAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCCAGACCAATGCTGCTGACAA
AAACTCTGTGAAAACGGATCTGATGAAAGTGCCAATGGTCATCATGGACTGGGAGGAGTGTTCAAAGA
TGTTTCCAAAACTTACCAAAAATATGCTGTGTGCCGGATACAAGAATGAGAGCTATGATGCCTGCAAG
GGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGTGGTACCAGGTGGGCATCAT
CAGCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACACCTCGTTGGTGAACTACAACC
TCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGAGAAAAGGAGGACTTCTGTC
AAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCAGCCCCAGATCCTGGCTCCT
GCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTACCTCGAG

NOV19f, 258076366 Protein Sequence	SEQ ID NO: 254	243 aa	MW at 26830.8kD

GSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPT
QPGPATWRECWVAGWGQTNAADKNSVKTDLMKVPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACK
GDSGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSV
KQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILYLE

NOV19g, CG51752-04 DNA Sequence	SEQ ID NO: 255		852 bp
	ORF Start: ATG at 46		ORF Stop: TGA at 847

CTTCTGATCAACAGCCCCAGTAGGGTTTAAAGGTCCTAGAGCTAC**ATG**GGATTTAGGTTTCTGGGCAC
AGCCAATTCTGCCACTTTTGAGGCTTCCCTTCCCCTTCCACTTGCCCCTCTCTGGTTCTCTGCCCCCA
GTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACGACTTAACTAGCCCATCCATGGAAATAAAGGAG
GTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGAGCCAACATGGACAATGACATTGCCTTGCTGCT
GCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGTGCCCATCTGCCTCCCCACGCAGCCCGGCCCTG
CCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCCAGACCAATGCTGCTGACAAAAACTCTGTGAAA
ACGGATCTGATGAAAGCGCCAATGGTCATCATGGACTGGGAGGAGTGTTCAAAGATGTTTCCAAAACT
TACCAAAAATATGCTGTGTGCCGGATACAAGAATGAGAGCTATGATGCCTGCAAGGGTGACAGTGGGG
GGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGTGGTACCAGGTGGGCATCATCAGCTGGGGAAAG
AGCTGTGGAGAGAAGAACACCCCAGGGATATACACCTCGTTGGTGAATTACAACCTCTGGATCGAGAA
AGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGAGAAAAGGAGGACTTCTGTCAAACAGAAACCTA
TGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCAGCCCCAGATCCTGGCTCCTGCTCTGTCCCCTG
TCCCATGTGTTGTTCAGAGCTATTTTGTAC**TGA**TAA

NOV19g, CG51752-04 Protein Sequence	SEQ ID NO: 256	267 aa	MW at 29436.8kD

MGFRFLGTANSATFEASLPLPLAPLWFSAPSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKRANM
DNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADKNSVKTDLMKAPMVIMDWEE
CSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYTSLV
NYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILY

NOV19h, 191887409 DNA Sequence	SEQ ID NO: 257	816 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCACCATGGGATTTAGGTTTCTGGGCACAGCCAATTCTGCCACTTTTGAGACTTCCCTTCCCCT
TCCACTTGCCCCTCTCTGGTTCTCTGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACG
ACTTAACTAGCCCATCCATGGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGA
GCCAACATGGACAATGACATTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGT
GCCCATCTGCCTCCCCACGCAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCC
AGACCAATGCTGCTGACAAAAACTCTGTGAAAACGGATCTGATGAAAGTGCCAATGGTCATCATGGAC
TGGGAGGAGTGTTCAAAGATGTTTCCAAAACTTACCAAAAATATGCTGTGTGCCGGATACAAGAATGA
GAGCTATGATGCCTGCAAGGGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGT
GGTACCAGGTGGGCATCATCAGCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACACC
TCGTTGGTGAACTACAACCTCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGA

GAAAAGGAGGACCTCTGTCAAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCA
GCCCCAGATCCTGGCTCCTGCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTACCTCGAG

NOV19h, 191887409 Protein Sequence	SEQ ID NO: 258	272 aa	MW at 29986.4kD

GSTMGFRFLGTANSATFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKR
ANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADKNSVKTDLMKVPMVIMD
WEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYT
SLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILYLE

NOV19i, CG51752-01 DNA Sequence	SEQ ID NO: 259	1078 bp
	ORF Start: ATG at 243	ORF Stop: TGA at 1044

TTGATCCGTGCCAAGTGGCTTTTTGTGGGCTCTGTAGAGTGCTCTAAACCCAGCTCGGCCTTTGCTGT
ATTAGACAGAAGCACCTCATTCATATCCCTGGGGCCCCTGATGGTGCAGTGGTCTGGCTGTGGTCTGC
ACACCAGCTATTCTGTTTTGTTTTGTTTTGTTTTTTTCCTACCTTTTTCCAATCCTCACACCTTCTGA
TCAACAGCCCCAGTAGGGTTTAAAGGTCCTAGAGCTAC**ATG**GGATTTAGGTTTCTGGGCACAGCCAAT
TCTGCCACTTTTGAGACTTCCCTTCCCCTTCCACTTGCCCCTCTCTGGTTCTCTGCCACCAGTCCAGA
AGAACTGAGTGTCGTGCTGGGGACCAACGACTTAACTAGCCCATCCATGGAAATAAAGGAGGTCGCCA
GCATCATTCTTCACAAAGACTTTAAGAGAGCCAACATGGACAATGACATTGCCTTGCTGCTGCTGGCT
TCGCCCATCAAGCTCGATGACCTGAAGGTGCCCATCTGCCTCCCCACGCAGCCCGGCCCTGCCACATG
GCGCGAATGCTGGGTGGCAGGTTGGGGCCAGACCAATGCTGCTGACAAAAACTCTGTGAAAACGGATC
TGATGAAAGTGCCAATGGTCATCATGGACTGGGAGGAGTGTTCAAAGATGTTTCCAAAACTTACCAAA
AATATGCTGTGTGCCGGATACAAGAATGAGAGCTATGATGCCTGCAAGGGTGACAGTGGGGGGCCTCT
GGTCTGCACCCCAGAGCCTGGTGAGAAGTGGTACCAGGTGGGCATCATCAGCTGGGGAAAGAGCTGTG
GAGATAAGAACACCCCAGGGATATACACCTCGTTGGTGAACTACAACCTCTGGATCGAGAAAGTGACC
CAGCTAGGAGGCAGGCCCTTCAATGCAGAGAAAAGGAGGACTTCTGTCAAACAGAAACCTATGGGCTC
CCCAGTCTCGGGAGTCCCAGAGCCAGGCAGCCCCAGATCCTGGCTCCTGCTCTGTCCCCTGTCCCATG
TGTTGTTCAGAGCTATTTTGTAC**TGA**TAATAAAATAGAGGCTATTCTTTCAACCGAAA

NOV19i, CG51752-01 Protein Sequence	SEQ ID NO: 260	267 aa	MW at 29412.8kD

MGFRFLGTANSATFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKRANM
DNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADKNSVKTDLMKVPMVIMDWEE
CSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGEKWYQVGIISWGKSCGDKNTPGIYTSLV
NYNLWIEKVTQLGGRPFNAEKRRTSVKQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILY

NOV19j, CG51752-05 DNA Sequence	SEQ ID NO: 261	816 bp
	ORF Start: ATG at 10	ORF Stop: at 811

GGATCCACC**ATG**GGATTTAGGTTTCTGGGCACAGCCAATTCTGCCGCTTTTGAGACTTCCCTTCCCCT
TCCACTTGCCCCTCTCTGGTTCTCTGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACG
ACTTAACTAGCCCATCCATGGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGA
GCCAACATGGACAATGACATTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGT
GCCCATCTGCCTCCCCACGCAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCC
AGACCAATGCTGCTGACAAAAACTCTGTGAAAACGGATCTGATGAAAGCGCCAATGGTCATCATGGAC
TGGGAGGAGTGTTCAAAGATGTTTCCAAAACTTACCAAAAATATGCTGTGTGCCGGATACAAGAATGA
GAGCTATGATGCCTGCAAGGGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGT
GGTACCAGGTGGGCATCATAAGCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACACC
TCGTTGGTGAACTACAACCTCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGA
GAAAAGGAGGACTTCTGTCAAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCA
GCCCCAGATCCTGGCTCCTGCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTACCTCGAG

NOV19j, CG51752-05 Protein Sequence	SEQ ID NO: 262	267 aa	MW at 29440.8kD

MGFRFLGTANSAAFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKRANM
DNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADKNSVKTDLMKAPMVIMDWEE
CSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYTSLV
NYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILY

NOV19k, CG51752-06 DNA Sequence	SEQ ID NO: 263	729 bp
	ORF Start: at 7	ORF Stop: at 724

GGATCCGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACGACTTAACTAGCCCATCCAT
GGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGAGCCAACATGGACAATGACA
TTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGTGCCCATCTGCCTCCCCACG
CAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCCAGACCAATGCTGCTGACAA
AAACTCTGTGAAAACGGATCTGATGAAAGTGCCAATGGTCATCATGGACTGGGAGGAGTGTTCAAAGA
TGTTTCCAAAACTTACCAAAAATATGCTGTGTGCCGGATACAAGAATGAGAGCTATGATGCCTGCAAG
GGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGTGGTACCAGGTGGGCATCAT
CAGCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACACCTCGTTGGTGAACTACAACC
TCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGAGAAAAGGAGGACCTCTGTC
AAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCAGCCCCAGATCCTGGCTCCT
GCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTACCTCGAG

NOV19k, CG51752-06 Protein Sequence	SEQ ID NO: 264	239 aa	MW at 26444.4kD

ATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQP
GPATWRECWVAGWGQTNAADKNSVKTDLMKVPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGD
SGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQ
KPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILY

NOV19l, CG51752-07 DNA Sequence	SEQ ID NO: 265	816 bp
	ORF Start: ATG at 10	ORF Stop: at 811

GGATCCACCATGGGATTTAGGTTTCTGGGCACAGCCAATTCTGCCACTTTTGAGACTTCCCTTCCCCT
TCCACTTGCCCCTCTCTGGTTCTCTGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACG
ACTTAACTAGCCCATCCATGGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGA
GCCAACATGGACAATGACATTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGT
GCCCATCTGCCTCCCCACGCAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCC
AGACCAATGCTGCTGACAAAAACTCTGTGAAAACGGATCTGATGAAAGTGCCAATGGTCATCATGGAC
TGGGAGGAGTGTTCAAAGATGTTTCCAAAACTTACCAAAAATATGCTGTGTGCCGGATACAAGAATGA
GAGCTATGATGCCTGCAAGGGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGT
GGTACCAGGTGGGCATCATCAGCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACACC
TCGTTGGTGAACTACAACCTCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGA
GAAAAGGAGGACCTCTGTCAAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCA
GCCCCAGATCCTGGCTCCTGCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTACCTCGAG

NOV19l, CG51752-07 Protein Sequence	SEQ ID NO: 266	267 aa	MW at 29498.9kD

MGFRFLGTANSATFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKRANM
DNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADKNSVKTDLMKVPMVIMDWEE
CSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYTSLV
NYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILY

NOV19m, SNP13374584 of CG51752-02, DNA Sequence	SEQ ID NO: 267	1042 bp
	ORF Start: ATG at 207	ORF Stop: TGA at 1008
	SNP Pos: 243	SNP Change: A to G

AGAGTGCTCTAAACCCAGCTCGGCCTTTGCTGTATTAGACAGAAGCACCTCATTCATATCCCTGGGGC
CCCTGATGGTGCAGTGGTCTGGCTGTGGTCTGCACACCAGCTATTCTGTTTTGTTTTGTTTTGTTTTT
TCCTACCTTTTTCCAATCCTCACACCTTCTGATCAACAGCCCCAGTAGGGTTTAAAGGTCCTAGAGCT
ACATGGGATTTAGGTTTCTGGGCACAGCCAATTCTGCCGCTTTTGAGACTTCCCTTCCCCTTCCACTT
GCCCCTCTCTGGTTCTCTGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACGACTTAAC
TAGCCCATCCATGGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGAGCCAACA
TGGACAATGACATTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGTGCCCATC
TGCCTCCCCACGCAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCCAGACCAA
TGCTGCTGACAAAAACTCTGTGAAAACGGATCTGATGAAAGTGCCAATGGTCATCATGGACTGGGAGG

AGTGTTCAAAGATGTTTCCAAAACTTACCAAAAATATGCTGTGTGCCGGATACAAGAATGAGAGCTAT
GATGCCTGCAAGGGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGTGGTACCA
GGTGGGCATCATCAGCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACACCTCGTTGG
TGAACTACAACCTCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGAGAAAAGG
AGGACTTCTGTCAAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCAGCCCCAG
ATCCTGGCTCCTGCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTACTGATAATAAAATA
GAGGCTATTCTTTCAACCGAAA

NOV19m, SNP13374584 of CG51752-02, Protein Sequence	SEQ ID NO: 268	267 aa	MW at 29468.8kD
	SNP Pos: 13		SNP Change: Thr to Ala

MGFRFLGTANSAAFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKRANM
DNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADKNSVKTDLMKVPMVIMDWEE
CSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYTSLV
NYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILY

NOV19n, SNP13374585 of CG51752-02, DNA Sequence	SEQ ID NO: 269	1042 bp
	ORF Start: ATG at 207	ORF Stop: TGA at 1008
	SNP Pos: 586	SNP Change: T to C

AGAGTGCTCTAAACCCAGCTCGGCCTTTGCTGTATTAGACAGAAGCACCTCATTCATATCCCTGGGGC
CCCTGATGGTGCAGTGGTCTGGCTGTGGTCTGCACACCAGCTATTCTGTTTTGTTTTGTTTTGTTTTT
TCCTACCTTTTTCCAATCCTCACACCTTCTGATCAACAGCCCCAGTAGGGTTTAAAGGTCCTAGAGCT
ACATGGGATTTAGGTTTCTGGGCACAGCCAATTCTGCCACTTTTGAGACTTCCCTTCCCCTTCCACTT
GCCCCTCTCTGGTTCTCTGCCACCAGTCCAGAAGAACTGAGTGTCGTGCTGGGGACCAACGACTTAAC
TAGCCCATCCATGGAAATAAAGGAGGTCGCCAGCATCATTCTTCACAAAGACTTTAAGAGAGCCAACA
TGGACAATGACATTGCCTTGCTGCTGCTGGCTTCGCCCATCAAGCTCGATGACCTGAAGGTGCCCATC
TGCCTCCCCACGCAGCCCGGCCCTGCCACATGGCGCGAATGCTGGGTGGCAGGTTGGGGCCAGACCAA
TGCTGCTGACAAAAACTCTGTGAAAACGGATCTGATGAAAGCGCCAATGGTCATCATGGACTGGGAGG
AGTGTTCAAAGATGTTTCCAAAACTTACCAAAAATATGCTGTGTGCCGGATACAAGAATGAGAGCTAT
GATGCCTGCAAGGGTGACAGTGGGGGGCCTCTGGTCTGCACCCCAGAGCCTGGTGAGAAGTGGTACCA
GGTGGGCATCATCAGCTGGGGAAAGAGCTGTGGAGAGAAGAACACCCCAGGGATATACACCTCGTTGG
TGAACTACAACCTCTGGATCGAGAAAGTGACCCAGCTAGAGGGCAGGCCCTTCAATGCAGAGAAAAGG
AGGACTTCTGTCAAACAGAAACCTATGGGCTCCCCAGTCTCGGGAGTCCCAGAGCCAGGCAGCCCCAG
ATCCTGGCTCCTGCTCTGTCCCCTGTCCCATGTGTTGTTCAGAGCTATTTTGTACTGATAATAAAATA
GAGGCTATTCTTTCAACCGAAA

NOV19n, SNP13374585 of CG51752-02, Protein Sequence	SEQ ID NO: 270	267 aa	MW at 29470.8kD
	SNP Pos: 127		SNP Change: Val to Ala

MGFRFLGTANSATFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASIILHKDFKRANM
DNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADKNSVKTDLMKAPMVIMDWEE
CSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGEKWYQVGIISWGKSCGEKNTPGIYTSLV
NYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGSPVSGVPEPGSPRSWLLLCPLSHVLFRAILY

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 19B.

Table 19B. Comparison of the NOV19 protein sequences.

```
NOV19a   ---MGFRFLGTANSATFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASI
NOV19b   ---MGFRFLGTANSATFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASI
NOV19c   GSTMGFRFLGTANSAAFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASI
NOV19d   GSTMGFRFLGTANSAAFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASI
NOV19e   ----------------------------GSATSPEELSVVLGTNDLTSPSMEIKEVASI
NOV19f   ----------------------------GSATSPEELSVVLGTNDLTSPSMEIKEVASI
NOV19g   ---MGFRFLGTANSATFEASLPLPLAPLWFSAPSPEELSVVLGTNDLTSPSMEIKEVASI
NOV19h   GSTMGFRFLGTANSATFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASI
```

```
NOV19i  ---MGFRFLGTANSATFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASI
NOV19j  ---MGFRFLGTANSAAFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASI
NOV19k  ------------------------------ATSPEELSVVLGTNDLTSPSMEIKEVASI
NOV19l  ---MGFRFLGTANSATFETSLPLPLAPLWFSATSPEELSVVLGTNDLTSPSMEIKEVASI

NOV19a  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK
NOV19b  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK
NOV19c  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK
NOV19d  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK
NOV19e  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK
NOV19f  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK
NOV19g  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK
NOV19h  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK
NOV19i  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK
NOV19j  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK
NOV19k  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK
NOV19l  ILHKDFKRANMDNDIALLLLASPIKLDDLKVPICLPTQPGPATWRECWVAGWGQTNAADK

NOV19a  NSVKTDLMKVPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGE
NOV19b  NSVKTDLMKAPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGE
NOV19c  NSVKTDLMKAPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGE
NOV19d  NSVKTDLMKAPMVIMDWEECSKMFPKLTKNMLRAGYKNESYDACKGDSGGPLVCTPEPGE
NOV19e  NSVKTDLMKAPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGE
NOV19f  NSVKTDLMKVPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGE
NOV19g  NSVKTDLMKAPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGE
NOV19h  NSVKTDLMKVPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGE
NOV19i  NSVKTDLMKVPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGE
NOV19j  NSVKTDLMKAPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGE
NOV19k  NSVKTDLMKVPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGE
NOV19l  NSVKTDLMKVPMVIMDWEECSKMFPKLTKNMLCAGYKNESYDACKGDSGGPLVCTPEPGE

NOV19a  KWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGS
NOV19b  KWYQVGIITWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGS
NOV19c  KWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGS
NOV19d  KWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGS
NOV19e  KWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGS
NOV19f  KWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGS
NOV19g  KWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGS
NOV19h  KWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGS
NOV19i  KWYQVGIISWGKSCGDKNTPGIYTSLVNYNLWIEKVTQLGGRPFNAEKRRTSVKQKPMGS
NOV19j  KWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGS
NOV19k  KWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGS
NOV19l  KWYQVGIISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRRTSVKQKPMGS

NOV19a  PVSGVPEPGSPRSWLLLCPLSHVLFRAILY--
NOV19b  PVSGVPEPGSPRSWLPLCPLSHVLFRAILY--
NOV19c  PVSGVPEPGSPRSWLLLCPLSHVLFRAILYLE
NOV19d  PVSGVPEPGSPRSWLLLCPLSHVLFRAILYLE
NOV19e  PVSGVPEPGSPRSWLLLCPLSHVLFRAILYLE
NOV19f  PVSGVPEPGSPRSWLLLCPLSHVLFRAILYLE
NOV19g  PVSGVPEPGSPRSWLLLCPLSHVLFRAILY--
NOV19h  PVSGVPEPGSPRSWLLLCPLSHVLFRAILYLE
NOV19i  PVSGVPEPGSPRSWLLLCPLSHVLFRAILY--
NOV19j  PVSGVPEPGSPRSWLLLCPLSHVLFRAILY--
```

```
NOV19k  PVSGVPEPGSPRSWLLLCPLSHVLFRAILY--
NOV19l  PVSGVPEPGSPRSWLLLCPLSHVLFRAILY--

NOV19a  (SEQ ID NO:  244)
NOV19b  (SEQ ID NO:  246)
NOV19c  (SEQ ID NO:  248)
NOV19d  (SEQ ID NO:  250)
NOV19e  (SEQ ID NO:  252)
NOV19f  (SEQ ID NO:  254)
NOV19g  (SEQ ID NO:  256)
NOV19h  (SEQ ID NO:  258)
NOV19i  (SEQ ID NO:  260)
NOV19j  (SEQ ID NO:  262)
NOV19k  (SEQ ID NO:  264)
NOV19l  (SEQ ID NO:  266)
```

Further analysis of the NOV19a protein yielded the following properties shown in Table 19C.

Table 19C. Protein Sequence Properties NOV19a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 4;  pos.chg 1;  neg.chg 0
      H-region:  length 10;  peak value   4.03
      PSG score:  -0.38

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.06
      possible cleavage site: between 32 and 33

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   0
      PERIPHERAL  Likelihood =  2.38 (at 72)
      ALOM score:   0.05  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  8.51
      Hyd Moment(95):  7.65    G content:       2
      D/E content:     2       S/T content:     8
      Score: -4.16

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 14  FRF|LG
```

```
NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  10.5%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: GFRF

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:
      Bacterial regulatory proteins, gntR family signature (PS00043):
*** found ***
         VASIILHKDFKRANMDNDIALL at 54

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 70.6

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------
```

```
Final Results (k = 9/23):

       43.5 %: mitochondrial
       34.8 %: cytoplasmic
        8.7 %: peroxisomal
        4.3 %: vacuolar
        4.3 %: nuclear
        4.3 %: endoplasmic reticulum

>> prediction for CG51752-02 is mit (k=23)
```

A search of the NOV19a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 19D.

Table 19D. Geneseq Results for NOV19a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU08686	Human FCTR6b polypeptide sequence - Homo sapiens, 267 aa. [WO200166747-A2, 13-SEP-2001]	1..267 1..267	267/267 (100%) 267/267 (100%)	e-160
AAU08685	Human FCTR6a polypeptide sequence - Homo sapiens, 267 aa. [WO200166747-A2, 13-SEP-2001]	1..267 1..267	265/267 (99%) 266/267 (99%)	e-159
AAU82737	Amino acid sequence of novel human protease #36 - Homo sapiens, 352 aa. [WO200200860-A2, 03-JAN-2002]	32..267 117..352	235/236 (99%) 235/236 (99%)	e-141
AAB03160	Human trypsin family serine protease Tespec PRO-3 - Homo sapiens, 352 aa. [WO200026352-A1, 11-MAY-2000]	32..267 117..352	233/236 (98%) 234/236 (98%)	e-140
AAU76529	Human LP polypeptide #2 - Homo sapiens, 356 aa. [WO200216578-A2, 28-FEB-2002]	32..249 117..334	214/218 (98%) 215/218 (98%)	e-127

In a BLAST search of public sequence databases, the NOV19a protein was found to have homology to the proteins shown in the BLASTP data in Table 19E.

Table 19E. Public BLASTP Results for NOV19a

Protein Accession Number	Protein/Organism/Length	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC88680	Sequence 22 from Patent WO0166747 - Homo sapiens (Human), 267 aa.	1..267 1..267	267/267 (100%) 267/267 (100%)	e-160
CAC88679	Sequence 20 from Patent WO0166747 - Homo sapiens (Human), 267 aa.	1..267 1..267	265/267 (99%) 266/267 (99%)	e-159
Q9MZZ6	Hypothetical 29.5 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 267 aa.	1..267 1..267	248/267 (92%) 251/267 (93%)	e-146
CAD28997	Sequence 3 from Patent WO0216578 - Homo sapiens (Human), 356 aa.	32..249 117..334	214/218 (98%) 215/218 (98%)	e-127
AAH49714	1700049K14Rik protein - Mus musculus (Mouse), 321 aa.	27..225 98..297	87/200 (43%) 119/200 (59%)	5e-45

PFam analysis indicates that the NOV19a protein contains the domains shown in the Table 19F.

Table 19F. Domain Analysis of NOV19a

Pfam Domain	NOV19a Match Region	Identities/ Similarities for the Matched Region	Expect Value
trypsin	5..210	75/267 (28%) 153/267 (57%)	8.6e-34

Example 20.

The NOV20 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 20A.

Table 20A. NOV20 Sequence Analysis		
NOV20a, CG51914-02 DNA Sequence	SEQ ID NO: 271	3261 bp
	ORF Start: ATG at 92	ORF Stop: TGA at 3146

AGCGGGGCTCTGGGGTCTGGGGGCATTGCTCAGCGGTGCTAGGCTGGCGCGGCTTGAGCCGCCGCCGG
ACTGACAGCTCGGTCTGCGGACC**ATG**GAGACCTGCGCCGGTCCACACCCGCTGCGCCTCTTCCTCTGC
CGGATGCAGCTCTGTCTCGCGCTGCTTTTGGGACCCTGGCGGCCTGGGACCGCCGAGGAAGTTATCCT
CCTGGATTCCAAAGCCTCCCAGGCCGAGCTGGGCTGGACTGCACTGCCAAGTAATGGGTGGGAGGAGA
TCAGCGGCGTGGATGAACACGACCGTCCCATCCGCACGTACCAAGTGTGCAATGTGCTGGAGCCCAAC
CAGGACAACTGGCTGCAGACTGGCTGGATAAGCCGTGGCCGCGGGCAGCGCATCTTCGTGGAACTGCA
GTTCACACTCCGTGACTGCAGCAGCATCCCTGGCGCCGCGGGTACCTGCAAGGAGACCTTCAACGTCT
ACTACCTGGAAACTGAGGCCGACCTGGGCCGTGGGCGTCCCCGCCTAGGCGGCAGCCGGCCCCGCAAA
ATCGACACGATCGCGGCGGACGAGAGCTTCACGCAGGGCGACCTGGGTGAGCGCAAGATGAAGCTGAA
CACAGAGGTGCGCGAGATCGGACCGCTCAGCCGGCGGGGTTTCCACCTGGCCTTTCAGGACGTGGGCG
CATGCGTGGCGCTTGTCTCGGTGCGCGTCTACTACAAGCAGTGCCGCGCCACCGTGCGGGGCCTGGCC
ACGTTCCCAGCCACCGCAGCCGAGAGCGCCTTCTCCACACTGGTGGAAGTGGCCGGAACGTGCGTGGC
GCACTCGGAAGGGGAGCCTGGCAGCCCCCCACGCATGCACTGCGGCGCCGACGGCGAGTGGCTGGTGC
CTGTGGGCCGCTGCAGCTGCAGCGCGGGATTCCAGGAGCGTGGTGACTTCTGCGAATGTCCCCCAGGG
TTTTACAAGGTGTCCCCGCGGCGGCCCCTCTGCTCACCGTGCCCAGAGCACAGCCGGGCCCTGGAAAA
CGCCTCCACCTTCTGCGTGTGCCAGGACAGCTATGCGCGCTCACCCACCGACCCGCCCTCGGCTTCCT
GCACCCGTCCGCCGTCGGCGCCGCGGGACCTGCAGTACAGCCTGAGCCGCTCGCCGCTGGTGCTGCGA
CTGCGCTGGCTGCCGCCGGCCGACTCGGGAGGCCGCTCGGACGTCACCTACTCGCTGCTGTGCCTGCG
CTGCGGCCGCGAGGGCCCGGCGGGCGCCTGCGAGCCGTGCGGGCCGCGCGTGGCCTTCCTACCGCGCC
AGGCAGGGCTGCGGGAGCGAGCCGCCACGCTGCTGCACCTGCGGCCCGGGGCGCGCTACACCGTGCGC
GTGGCCGTGCTCAACGGCGTCTCGTCAGCGCCCTGGGAGGAGGATGAGATCCGCAGGGACCGAGTGGA
ACCCCAGAGCGTGTCCCTGTCGTGGCGGGAGCCCATCCCTGCCGGAGCCCCTGGGGCCAATGACACGG
AGTACGAGATCCGATACTACGAGAAGGGTCAGAGTGAGCAGACTTACTCCATGGTGAAGACGGGGGCG
CCCACAGTCACCGTCACCAACCTGAAGCCGGCTACCCGCTACGTCTTTCAGATCCGGGCCGCTTCCCC
GGGGCCATCCTGGGAGGCCCAGAGTTTTAACCCCAGCATTGAAGTACAGACCCTGGGGGAGGCTGCCT
CAGGGTCCAGGGACCAGAGCCCCGCCATTGTCGTCACCGTAGTGACCATCTCGGCCCTCCTCGTCCTG
GGCTCCGTGATGAGTGTGCTGGCCATTTGGAGGAGGCCCTGCAGCTATGGCAAAGGAGGAGGGGATGC
CCATGATGAAGAGGAGCTGTATTTCCACTTCAAAGTCCCAACACGTCGCACATTCCTGGACCCCCAGA
GCTGTGGGGACCTGCTGCAGGCTGTGCATCTGTTCGCCAAGGAACTGGATGCGAAAAGCGTCACGCTG
GAGAGGAGCCTTGGAGGAGGCAAGCTGGGCGCCCAGGAAGCCTTGTCCCCATCTGGAAGCCTCACCCA
CTCTATAGGCCCCGCCCCTACTCTGTCCACACCTCTATCCGGGCGGTTTGGGGAGCTGTGCTGTGGCT
GCTTGCAGCTCCCCGGTCGCCAGGAGCTGCTCGTAGCCGTGCATATGCTGAGGGACAGCGCCTCCGAC
TCACAGAGGCTCGGCTTCCTGGCCGAGGCCCTCACGCTGGGCCAGTTTGACCATAGCCACATCGTGCG
GCTGGAGGGCGTTGTTACCCGAGGTAGGGGAAGCACCTTGATGATTGTCACCGAGTACATGAGCCATG
GGGCCCTGGACGGCTTCCTCAGGCGGCACGAGGGGCAGCTGGTGGCTGGGCAACTGATGGGGTTGCTG
CCTGGGCTGGCATCAGCCATGAAGTATCTGTCAGAGATGGGCTACGTTCACCGGGGCCTGGCAGCTCG
CCATGTGCTGGTCAGCAGCGACCTTGTCTGCAAGATCTCTGGCTTCGGGCGGGGCCCCCGGGACCGAT
CAGAGGCTGTCTACACCACTGGCCGGAGCCCAGCGCTATGGGCCGCTCCCGAGACACTTCAGTTTGGC
CACTTCAGCTCTGCCAGTGACGTGTGGAGCTTCGGCATCATCATGTGGGAGGTGATGGCCTTTGGGGA
GCGGCCTTACTGGGACATGTCTGGCCAAGACGTGATCAAGGCTGTGGAGGATGGCTTCCGGCTGCCAC
CCCCCAGGAACTGTCCTAACCTTCTGCACCGACTAATGCTCGACTGCTGGCAGAAGGACCCAGGTGAG
CGGCCCAGGTTCTCCCAGATCCACAGCATCCTGAGCAAGATGGTGCAGGACCCAGAGCCCCCCAAGTG
TGCCCTGACTACCTGTCCCAGGCCTCCCACCCCACTAGCGGACCGTGCCTTCTCCACCTTCCCCTCCT
TTGGCTCTGTGGGCGCGTGGCTGGAGGCCCTGGACCTGTGCCGCTACAAGGACAGCTTCGCGGCTGCT
GGCTATGGGAGCCTGGAGGCCGTGGCCGAGATGACTGCCCAGGACCTGGTGAGCCTAGGCATCTCTTT
GGCTGAACATCGAGAGGCCCTCCTCAGCGGGATCAGCGCCCTGCAGGCACGAGTGCTCCAGCTGCAGG
GCCAGGGGGTGCAGGTG**TGA**GTGGACCCCATTCTTCCAAGGCAGGACTCCGGTGGGGGTCCAGTCCCC
CAGCCCTGCCCAAGGACCGTGGCAAGCTGCGCTCCAGCAGTGTGGGAGGGAGCGCTCTCTTCCTC

NOV20a, CG51914-02 Protein Sequence	SEQ ID NO: 272	1018 aa	MW at 110851.8kD

METCAGPHPLRLFLCRMQLCLALLLGPWRPGTAEEVILLDSKASQAELGWTALPSNGWEEISGVDEHD
RPIRTYQVCNVLEPNQDNWLQTGWISRGRGQRIFVELQFTLRDCSSIPGAAGTCKETFNVYYLETEAD

LGRGRPRLGGSRPRKIDTIAADESFTQGDLGERKMKLNTEVREIGPLSRRGFHLAFQDVGACVALVSV
RVYYKQCRATVRGLATFPATAAESAFSTLVEVAGTCVAHSEGEPGSPPRMHCGADGEWLVPVGRCSCS
AGFQERGDFCECPPGFYKVSPRRPLCSPCPEHSRALENASTFCVCQDSYARSPTDPPSASCTRPPSAP
RDLQYSLSRSPLVLRLRWLPPADSGGRSDVTYSLLCLRCGREGPAGACEPCGPRVAFLPRQAGLRERA
ATLLHLRPGARYTVRVAVLNGVSSAPWEEDEIRRDRVEPQSVSLSWREPIPAGAPGANDTEYEIRYYE
KGQSEQTYSMVKTGAPTVTVTNLKPATRYVFQIRAASPGPSWEAQSFNPSIEVQTLGEAASGSRDQSP
AIVVTVVTISALLVLGSVMSVLAIWRRPCSYGKGGGDAHDEEELYFHFKVPTRRTFLDPQSCGDLLQA
VHLFAKELDAKSVTLERSLGGGKLGAQEALSPSGSLTHSIGPAPTLSTPLSGRFGELCCGCLQLPGRQ
ELLVAVHMLRDSASDSQRLGFLAEALTLGQFDHSHIVRLEGVVTRGRGSTLMIVTEYMSHGALDGFLR
RHEGQLVAGQLMGLLPGLASAMKYLSEMGYVHRGLAARHVLVSSDLVCKISGFGRGPRDRSEAVYTTG
RSPALWAAPETLQFGHFSSASDVWSFGIIMWEVMAFGERPYWDMSGQDVIKAVEDGFRLPPPRNCPNL
LHRLMLDCWQKDPGERPRFSQIHSILSKMVQDPEPPKCALTTCPRPPTPLADRAFSTFPSFGSVGAWL
EALDLCRYKDSFAAAGYGSLEAVAEMTAQDLVSLGISLAEHREALLSGISALQARVLQLQGQGVQV

NOV20b, CG51914-01 DNA Sequence	SEQ ID NO: 273	3003 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 3001

ATGGTATTGACAACTGCTATACCAGCCTGGCTTCTTAGCTGTTCCCTCCCACTCTCATCCTGGGCCCA
CCATGCGACACCGCCCCTCCGTCTAGTAGTTATCCTCCTGGATTCCAAAGCCTCCCAGGCCGAGCTGG
GCTGGACTGCACTGCCAAGTAATGGGTGGGAGGAGATCAGCGGCGTGGATGAACACGACCGTCCCATC
CGCACGTACCAAGTGTGCAATGTGCTGGAGCCCAACCAGGACAACTGGCTGCAGACTGGCTGGATAAG
CCGTGGCCGCGGGCAGCGCATCTTCGTGGAACTGCAGTTCACACTCCGTGACTGCAGCAGCATCCCTG
GCGCCGCGGGTACCTGCAAGGAGACCTTCAACGTCTACTACCTGGAAACTGAGGCCGACCTGGGCCGT
GGGCGTCCCCGCCTAGGCGGCAGCCGGCCCCGCAAAATCGACACGATCGCGGCGGACGAGAGCTTCAC
GCAGGGCGACCTGGGTGAGCGCAAGATGAAGCTGAACACAGAGGTGCGCGAGATCGGACCGCTCAGCC
GGCGGGGTTTCCACCTGGCCTTTCAGGACGTGGGCGCATGCGTGGCGCTTGTCTCGGTGCGCGTCTAC
TACAAGCAGTGCCGCGCCACCGTGCGGGGCCTGGCCACGTTCCCAGCCACCGCAGCCGAGAGCGCCTT
CTCCACACTGGTGGAAGTGGCCGGAACGTGCGTGGCGCACTCGGAAGGGGAGCCTGGCAGCCCCCCAC
GCATGCACTGCGGCGCCGACGGCGAGTGGCTGGTGCCTGTGGGCCGCTGCAGCTGCAGCGCGGGATTC
CAGGAGCGTGGTGACTTCTGCGAATGTCCCCCAGGGTTTTACAAGGTGTCCCCGCGGCGGCCCCTCTG
CTCACCGTGCCCAGAGCACAGCCGGGCCCTGGAAAACGCCTCCACCTTCTGCGTGTGCCAGGACAGCT
ATGCGCGCTCACCCACCGACCCGCCCTCGGCTTCCTGCACCCGTCCGCCGTCGGCGCCGCGGGACCTG
CAGTACAGCCTGAGCCGCTCGCCGCTGGTGCTGCGACTGCGCTGGCTGCCGCCGGCCGACTCGGGAGG
CCGCTCGGACGTCACCTACTCGCTGCTGTGCCTGCGCTGCGGCCGCGAGGGCCCGGCGGGCGCCTGCG
AGCCGTGCGGGCCGCGCGTGGCCTTCCTACCGCGCCAGGCAGGGCTGCGGGAGCGAGCCGCCACGCTG
CTGCACCTGCGGCCCGGCGCGCGCTACACCGTGCGCGTGGCCGCGCTCAACGGCGTCTCGGGCCCGGC
GGCCGCCGCGGGAACCACCTACGCGCAGGTCACCGTCTCCACCGGGCCCTCAGCGCCCTGGGAGGAGG
ATGAGATCCGCAGGGACCGAGTGGAACCCCAGAGCGTGTCCCTGTCGTGGCGGGAGCCCATCCCTGCC
GGAGCCCCTGGGGCCAATGACACGGAGTACGAGATCCGATACTACGAGAAGCAGAGTGAGCAGACTTA
CTCCATGGTGAAGACAGGGGCGCCCACAGTCACCGTCACCAACCTGAAGCCGGCTACCCGCTACGTCT
TTCAGATCCGGGCCGCTTCCCCGGGGCCATCCTGGGAGGCCCAGAGTTTTAACCCCAGCATTGAAGTA
CAGACCCTGGGGGAGGCTGCCTCAGGGTCCAGGGACCAGAGCCCCGCCATTGTCGTCACCGTAGTGAC
CATCTCGGCCCTCCTCGTCCTGGGCTCCGTGATGAGTGTGCTGGCCATTTGGAGGAGGAGGCCCTGCA
GCTATGGCAAAGGAGGAGGGGATGCCCATGATGAAGAGGAGCTGTATTTCCACTGTAAAGTCCCAACA
CGTCGCACATTCCTGGACCCCCAGAGCTGTGGGGACCTGCTGCAGGCTGTGCATCTGTTCGCCAAGGA
ACTGGATGCGAAAAGCGTCACGCTGGAGAGGAGCCTTGGAGGAGGCAAGTTTGGGGAGCTGTGCTGTG
GCTGCTTGCAGCTCCCCGGTCGCCAGGAGCTGCTCGTAGCCGTGCACATGCTGAGGGACAGCGCCTCC
GACTCACAGAGGCTCGGCTTCCTGGCCGAGGCCCTCACGCTGGGCCAGTTTGACCATAGCCACATCGT
GCGGCTGGAGGGCGTTGTTACCCGAGGTAGGACCTTGATGATTGTCACCGAGTACATGAGCCATGGGG
CCCTGGACGGCTTCCTCAGGCACGAGGGGCAGCTGGTGGCTGGGCAACTGATGGGGTTGCTGCCTGGG
CTGGCATCAGCCATGAAGTATCTGTCAGAGATGGGCTACGTTCACCGGGGCCTGGCAGCTCGCCATGT
GCTGGTCAGCAGCGACCTTGTCTGCAAGATCTCTGGCTTCGGGCGGGGCCCCCGGGACCGATCAGAGG
CTGTCTACACCACTGGCCGGAGCCCAGCGCTATGGGCCGCTCCCGAGACACTTCAGTTTGGCCACTTC
AGCTCTGCCAGTGACGTGTGGAGCTTCGGCATCATCATGTGGGAGGTGATGGCCTTTGGGGAGCGGCC
TTACTGGGACATGTCTGGCCAAGACGTGAAGGCTGTGGAGGATGGCTTCCGGCTGCCACCCCCCAGGA
ACTGTCCTAACCTTCTGCACCGACTAATGCTCGACTGCTGGCAGAAGGACCCAGGTGAGCGGCCCAGG
TTCTCCCAGATCCACAGCATCCTGAGCAAGATGGTGCAGGACCCAGAGCCCCCCAAGTGTGCCCTGAC
TACCTGTCCCAGGCCTCCCACTCCACTAGCCGACCGTGCCTTCTCCACCTTCCCCTCCTTTGGCTCTG
TGGGCGCGTGGCTGGAGGCCCTGGACCTGTGCCGCTACAAGGACAGCTTCGCGGCTGCTGGCTATGGG
AGCCTGGAGGCCGTGGCCGAGATGACTGCCCAGGACCTGGTGAGCCTAGGCATCTCTTTGGCTGAACA

TCGAGAGGCCCTCCTCAGCGGGATCAGCGCCCTGCAGGCACGAGTGCTCCAGCTGCAGGGCCAGGGGG TGCAGGTG**TGA**			
NOV20b, CG51914-01 Protein Sequence	SEQ ID NO: 274	1000 aa	MW at 108840.5kD

MVLTTAIPAWLLSCSLPLSSWAHHATPPLRLVVILLDSKASQAELGWTALPSNGWEEISGVDEHDRPI
RTYQVCNVLEPNQDNWLQTGWISRGRGQRIFVELQFTLRDCSSIPGAAGTCKETFNVYYLETEADLGR
GRPRLGGSRPRKIDTIAADESFTQGDLGERKMKLNTEVREIGPLSRRGFHLAFQDVGACVALVSVRVY
YKQCRATVRGLATFPATAAESAFSTLVEVAGTCVAHSEGEPGSPPRMHCGADGEWLVPVGRCSCSAGF
QERGDFCECPPGFYKVSPRRPLCSPCPEHSRALENASTFCVCQDSYARSPTDPPSASCTRPPSAPRDL
QYSLSRSPLVLRLRWLPPADSGGRSDVTYSLLCLRCGREGPAGACEPCGPRVAFLPRQAGLRERAATL
LHLRPGARYTVRVAALNGVSGPAAAAGTTYAQVTVSTGPSAPWEEDEIRRDRVEPQSVSLSWREPIPA
GAPGANDTEYEIRYYEKQSEQTYSMVKTGAPTVTVTNLKPATRYVFQIRAASPGPSWEAQSFNPSIEV
QTLGEAASGSRDQSPAIVVTVVTISALLVLGSVMSVLAIWRRRPCSYGKGGGDAHDEEELYFHCKVPT
RRTFLDPQSCGDLLQAVHLFAKELDAKSVTLERSLGGGKFGELCCGCLQLPGRQELLVAVHMLRDSAS
DSQRLGFLAEALTLGQFDHSHIVRLEGVVTRGRTLMIVTEYMSHGALDGFLRHEGQLVAGQLMGLLPG
LASAMKYLSEMGYVHRGLAARHVLVSSDLVCKISGFGRGPRDRSEAVYTTGRSPALWAAPETLQFGHF
SSASDVWSFGIIMWEVMAFGERPYWDMSGQDVKAVEDGFRLPPPRNCPNLLHRLMLDCWQKDPGERPR
FSQIHSILSKMVQDPEPPKCALTTCPRPPTPLADRAFSTFPSFGSVGAWLEALDLCRYKDSFAAAGYG
SLEAVAEMTAQDLVSLGISLAEHREALLSGISALQARVLQLQGQGVQV

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 20B.

Table 20B. Comparison of the NOV20 protein sequences.

```
NOV20a  METCAGPHPLRLFLCRMQLCLALLLGPWRPGTAEEVILLDSKASQAELGWTALPSNGWEE
NOV20b  -MVLTTAIPAWLLSC--SLPLSSWAHHATPPLRLVVILLDSKASQAELGWTALPSNGWEE

NOV20a  ISGVDEHDRPIRTYQVCNVLEPNQDNWLQTGWISRGRGQRIFVELQFTLRDCSSIPGAAG
NOV20b  ISGVDEHDRPIRTYQVCNVLEPNQDNWLQTGWISRGRGQRIFVELQFTLRDCSSIPGAAG

NOV20a  TCKETFNVYYLETEADLGRGRPRLGGSRPRKIDTIAADESFTQGDLGERKMKLNTEVREI
NOV20b  TCKETFNVYYLETEADLGRGRPRLGGSRPRKIDTIAADESFTQGDLGERKMKLNTEVREI

NOV20a  GPLSRRGFHLAFQDVGACVALVSVRVYYKQCRATVRGLATFPATAAESAFSTLVEVAGTC
NOV20b  GPLSRRGFHLAFQDVGACVALVSVRVYYKQCRATVRGLATFPATAAESAFSTLVEVAGTC

NOV20a  VAHSEGEPGSPPRMHCGADGEWLVPVGRCSCSAGFQERGDFCECPPGFYKVSPRRPLCSP
NOV20b  VAHSEGEPGSPPRMHCGADGEWLVPVGRCSCSAGFQERGDFCECPPGFYKVSPRRPLCSP

NOV20a  CPEHSRALENASTFCVCQDSYARSPTDPPSASCTRPPSAPRDLQYSLSRSPLVLRLRWLP
NOV20b  CPEHSRALENASTFCVCQDSYARSPTDPPSASCTRPPSAPRDLQYSLSRSPLVLRLRWLP

NOV20a  PADSGGRSDVTYSLLCLRCGREGPAGACEPCGPRVAFLPRQAGLRERAATLLHLRPGARY
NOV20b  PADSGGRSDVTYSLLCLRCGREGPAGACEPCGPRVAFLPRQAGLRERAATLLHLRPGARY

NOV20a  TVRVAVLNGVS----------SAPWEEDEIRRDRVEPQSVSLSWREPIPAG
NOV20b  TVRVAALNGVSGPAAAAGTTYAQVTVSTGPSAPWEEDEIRRDRVEPQSVSLSWREPIPAG

NOV20a  APGANDTEYEIRYYEKGQSEQTYSMVKTGAPTVTVTNLKPATRYVFQIRAASPGPSWEAQ
NOV20b  APGANDTEYEIRYYEK-QSEQTYSMVKTGAPTVTVTNLKPATRYVFQIRAASPGPSWEAQ

NOV20a  SFNPSIEVQTLGEAASGSRDQSPAIVVTVVTISALLVLGSVMSVLAIWRR-PCSYGKGGG
NOV20b  SFNPSIEVQTLGEAASGSRDQSPAIVVTVVTISALLVLGSVMSVLAIWRRRPCSYGKGGG
```

```
NOV20a  DAHDEEELYFHFKVPTRRTFLDPQSCGDLLQAVHLFAKELDAKSVTLERSLGGGKLGAQE
NOV20b  DAHDEEELYFHCKVPTRRTFLDPQSCGDLLQAVHLFAKELDAKSVTLERSLGGG------

NOV20a  ALSPSGSLTHSIGPAPTLSTPLSGRFGELCCGCLQLPGRQELLVAVHMLRDSASDSQRLG
NOV20b  ----------------------KFGELCCGCLQLPGRQELLVAVHMLRDSASDSQRLG

NOV20a  FLAEALTLGQFDHSHIVRLEGVVTRGRGSTLMIVTEYMSHGALDGFLRRHEGQLVAGQLM
NOV20b  FLAEALTLGQFDHSHIVRLEGVVTRGR--TLMIVTEYMSHGALDGFLR-HEGQLVAGQLM

NOV20a  GLLPGLASAMKYLSEMGYVHRGLAARHVLVSSDLVCKISGFGRGPRDRSEAVYTTGRSPA
NOV20b  GLLPGLASAMKYLSEMGYVHRGLAARHVLVSSDLVCKISGFGRGPRDRSEAVYTTGRSPA

NOV20a  LWAAPETLQFGHFSSASDVWSFGIIMWEVMAFGERPYWDMSGQDVIKAVEDGFRLPPPRN
NOV20b  LWAAPETLQFGHFSSASDVWSFGIIMWEVMAFGERPYWDMSGQD-VKAVEDGFRLPPPRN

NOV20a  CPNLLHRLMLDCWQKDPGERPRFSQIHSILSKMVQDPEPPKCALTTCPRPPTPLADRAFS
NOV20b  CPNLLHRLMLDCWQKDPGERPRFSQIHSILSKMVQDPEPPKCALTTCPRPPTPLADRAFS

NOV20a  TFPSFGSVGAWLEALDLCRYKDSFAAAGYGSLEAVAEMTAQDLVSLGISLAEHREALLSG
NOV20b  TFPSFGSVGAWLEALDLCRYKDSFAAAGYGSLEAVAEMTAQDLVSLGISLAEHREALLSG

NOV20a  ISALQARVLQLQGQGVQV
NOV20b  ISALQARVLQLQGQGVQV

NOV20a  (SEQ ID NO:  272)
NOV20b  (SEQ ID NO:  274)
```

Further analysis of the NOV20a protein yielded the following properties shown in Table 20C.

Table 20C. Protein Sequence Properties NOV20a

SignalP analysis:	Cleavage site between residues 34 and 35

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 1;  neg.chg 1
      H-region:  length 4;  peak value   0.56
      PSG score:  -3.84

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   0.06
      possible cleavage site: between 33 and 34

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   3
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -9.08   Transmembrane  546 - 562
      PERIPHERAL  Likelihood =  0.85 (at 754)
```

```
      ALOM score:  -9.08  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 553
      Charge difference:  3.5   C( 1.5) - N(-2.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 1b  (cytoplasmic tail 546 to 1018)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       3       Hyd Moment(75):  5.96
      Hyd Moment(95):  3.64    G content:       3
      D/E content:     2       S/T content:     2
      Score: -5.44

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 39  WRP|GT

NUCDISC: discrimination of nuclear localization signals
      pat4: RPRK (4) at  148
      pat7: none
      bipartite: none
      content of basic residues:  10.2%
      NLS Score: -0.22

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  found
      RLFLCRMQL at 11

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: too long tail

Dileucine motif in the tail: found
      LL at 556
      LL at 609
      LL at 682
```

```
      LL at 762
      LL at 884
      LL at 997

checking 63 PROSITE DNA binding motifs:
      Leucine zipper pattern (PS00029):  *** found ***
         LSGRFGELCCGCLQLPGRQELL at 662
         LVSLGISLAEHREALLSGISAL at 983
 none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 70.6

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       39.1 %: nuclear
       21.7 %: cytoplasmic
       17.4 %: mitochondrial
        8.7 %: vesicles of secretory system
        4.3 %: vacuolar
        4.3 %: peroxisomal
        4.3 %: endoplasmic reticulum

>> prediction for CG51914-02 is nuc (k=23)
```

A search of the NOV20a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 20D.

Table 20D. Geneseq Results for NOV20a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV20a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABB98843	Human NEPHA - Homo sapiens, 1008 aa. [WO200283735-A1, 24-OCT-2002]	1..1018 1..1008	984/1040 (94%) 984/1040 (94%)	0.0
AAE19158	Human kinase polypeptide (PKIN-16) - Homo sapiens, 1009 aa. [WO200208399-A2, 31-JAN-2002]	1..1018 1..1009	973/1041 (93%) 975/1041 (93%)	0.0
AAU03553	Human protein kinase #53 - Homo sapiens, 1009 aa. [WO200138503-A2, 31-MAY-2001]	1..1018 1..1009	973/1041 (93%) 975/1041 (93%)	0.0
AAU76874	Human EphA full length kinase - Homo sapiens, 974 aa. [WO200208253-A2, 31-JAN-2002]	37..1018 1..974	948/1004 (94%) 948/1004 (94%)	0.0
AAM47209	Human NOV3 protein - Homo sapiens, 1000 aa. [WO200174851-A2, 11-OCT-2001]	36..1018 33..1000	946/1003 (94%) 946/1003 (94%)	0.0

In a BLAST search of public sequence databases, the NOV20a protein was found to have homology to the proteins shown in the BLASTP data in Table 20E.

Table 20E. Public BLASTP Results for NOV20a				
Protein Accession Number	**Protein/Organism/Length**	**NOV20a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
CAD23752	Sequence 1 from Patent WO0208253 - Homo sapiens (Human), 974 aa (fragment).	37..1018 1..974	948/1004 (94%) 948/1004 (94%)	0.0
CAD13085	Sequence 5 from Patent WO0174851 - Homo sapiens (Human), 1000 aa.	36..1018 33..1000	946/1003 (94%) 946/1003 (94%)	0.0
CAD23753	Sequence 3 from Patent WO0208253 - Homo sapiens (Human), 1022 aa (fragment).	37..1018 1..1022	947/1052 (90%) 947/1052 (90%)	0.0
CAD23754	Sequence 5 from Patent WO0208253 - Homo sapiens (Human), 647 aa (fragment).	37..663 1..647	625/647 (96%) 625/647 (96%)	0.0
CAD23755	Sequence 7 from Patent WO0208253 - Homo sapiens (Human), 695 aa (fragment).	37..663 1..695	624/695 (89%) 624/695 (89%)	0.0

PFam analysis indicates that the NOV20a protein contains the domains shown in the Table 20F.

Table 20F. Domain Analysis of NOV20a			
Pfam Domain	**NOV20a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
EPH_lbd	35..211	116/178 (65%) 159/178 (89%)	8.8e-122
fn3	337..432	25/100 (25%) 65/100 (65%)	1.9e-05
fn3	437..517	23/88 (26%) 58/88 (66%)	7.3e-09
pkinase	656..910	81/298 (27%) 176/298 (59%)	8.3e-38
SAM	941..1005	27/68 (40%) 51/68 (75%)	7.1e-16

Example 21.

The NOV21 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 21A.

Table 21A. NOV21 Sequence Analysis		
NOV21a, CG51965-01 DNA Sequence	SEQ ID NO: 275	9087 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 9085

ATGGCGCCGCCGCCGCCGCCCGTGCTGCCCGTGCTGCTGCTCCTGGCCGCCGCCGCCGCCCTGCCGGC
GATGGGGCTGCGAGCGGCCGCCTGGGAGCCGCGCGTACCCGGCGGGACCCGCGCCTTCGCCCTCCGGC
CCGGCTGTACCTACGCGGTGGGCGCCGCTTGCACGCCCCGGGCGCCGCGGGAGCTGCTGGACGTGGGC
CGCGATGGGCGGCTGGCAGGACGTCGGCGCGTCTCGGGCGCGGGGCGCCCGCTGCCGCTGCAAGTCCG
CTTGGTGGCCCGCAGTGCCCCGACGGCGCTGAGCCGCCGCCTGCGGGCGCGCACGCACCTTCCCGGCT
GCGGAGCCCGTGCCCGGCTCTGCGGAACCGGTGCCCGGCTCTGCGGGGCGCTCTGCTTCCCCGTCCCC
GGCGGCTGCGCGGCCGCGCAGCATTCGGCGCTCGCAGCTCCGACCACCTTACCCGCCTGCCGCTGCCC
GCCGCGCCCCAGGCCCCGCTGTCCCGGCCGTCCCATCTGCCTGCCGCCGGGCGGCTCGGTCCGCCTGC
GTCTGCTGTGCGCCCTGCGGCGCGCGGCTGGCGCCGTCCGGGTGGGACTGGCGCTGGAGGCCGCCACC
GCGGGGACGCCCTCCGCGTCGCCATCCCCATCGCCGCCCCTGCCGCCGAACTTGCCCGAAGCCCGGGC
GGGGCCGGCGCGACGGGCCCGGCGGGGCACGAGCGGCAGAGGGAGCCTGAAGTTTCCGATGCCCAACT
ACCAGGTGGCGTTGTTTGAGAACGAACCGGCGGGCACCCTCATCCTCCAGCTGCACGCGCACTACACC
ATCGAGGGCGAGGAGGAGCGCGTGAGCTATTACATGGAGGGGCTGTTCGACGAGCGCTCCCGGGGCTA
CTTCCGAATCGACTCTGCCACGGGCGCCGTGAGCACGGACAGCGTACTGGACCGCGAGACCAAGGAGA
CGCACGTCCTCAGGGTGAAAGCCGTGGACTACAGTACGCCGCCGCGCTCGGCCACCACCTACATCACT
GTCTTGGTCAAAGACACCAACGACCACAGCCCGGTCTTCGAGCAGTCGGAGTACCGCGAGCGCGTGCG
GGAGAACCTGGAGGTGGGCTACGAGGTGCTGACCATCCGCGCCAGCGACCGCGACTCGCCCATCAACG
CCAACTTGCGTTACCGCGTGTTGGGGGGCGCGTGGGACGTCTTCCAGCTCAACGAGAGCTCTGGCGTG
GTGAGCACACGGGCGGTGCTGGACCGGGAGGAGGCGGCCGAGTACCAGCTCCTGGTGGAGGCCAACGA
CCAGGGGCGCAATCCGGGCCCGCTCAGTGCCACGGCCACCGTGTACATCGAGGTGGAGGACGAGAACG
ACAACTACCCCCAGTTCAGCGAGCAGAACTACGTGGTCCAGGTGCCCGAGGACGTGGGGCTCAACACG
GCTGTGCTGCGAGTGCAGGCCACGGACCGGGACCAGGGCCAGAACGCGGCCATTCACTACAGCATCCT
CAGCGGGAACGTGGCCGGCCAGTTCTACCTGCACTCGCTGAGCGGGATCCTGGATGTGATCAACCCCT
TGGATTTCGAGGATGTCCAGAAATACTCGCTGAGCATTAAGGCCCAGGATGGGGGCCGGCCCCCGCTC
ATCAATTCTTCAGGGGTGGTGTCTGTGCAGGTGCTGGATGTCAACGACAACGAGCCTATCTTTGTGAG
CAGCCCCTTCCAGGCCACGGTGCTGGAGAATGTGCCCCTGGGCTACCCCGTGGTGCACATTCAGGCGG
TGGACGCGGACTCTGGAGAGAACGCCCGGCTGCACTATCGCCTGGTGGACACGGCCTCCACCTTTCTG
GGGGGCGGCAGCGCTGGGCCTAAGAATCCTGCCCCCACCCCTGACTTCCCCTTCCAGATCCACAACAG
CTCCGGTTGGATCACAGTGTGTGCCGAGCTGGACCGCGAGGAGGTGGAGCACTACAGCTTCGGGGTGG
AGGCGGTGGACCACGGCTCGCCCCCCATGAGCTCCTCCACCAGCGTGTCCATCACGGTGCTGGACGTG
AATGACAACGACCCGGTGTTCACGCAGCCCACCTACGAGCTTCGTCTGAATGAGGATGCGGCCGTGGG
GAGCAGCGTGCTGACCCTGCAGGCCCGCGACCGTGACGCCAACAGTGTGATTACCTACCAGCTCACAG
GCGGCAACACCCGGAACCGCTTTGCACTCAGCAGCCAGAGAGGGGGCGGCCTCATCACCCTGGCGCTA
CCTCTGGACTACAAGCAGGAGCAGCAGTACGTGCTGGCGGTGACAGCATCCGACGGCACACGGTCGCA
CACTGCGCATGTCCTAATCAACGTCACTGATGCCAACACCCACAGGCCTGTCTTTCAGAGCTCCCATT
ACACAGTGAGTGTCAGTGAGGACAGGCCTGTGGGCACCTCCATTGCTACCCTCAGTGCCAACGATGAG
GACACAGGAGAGAATGCCCGCATCACCTACGTGATTCAGGACCCCGTGCCGCAGTTCCGCATTGACCC
CGACAGTGGCACCATGTACACCATGATGGAGCTGGACTATGAGAACCAGGTCGCCTACACGCTGACCA
TCATGGCCCAGGACAACGGCATCCCGCAGAAATCAGACACCACCACCCTAGAGATCCTCATCCTCGAT
GCCAATGACAATGCACCCCAGTTCCTGTGGGATTTCTACCAGGGTTCCATCTTTGAGGATGCTCCACC
CTCGACCAGCATCCTCCAGGTCTCTGCCACGGACCGGGACTCAGGTCCCAATGGGCGTCTGCTGTACA
CCTTCCAGGGTGGGGACGACGGCGATGGGGACTTCTACATCGAGCCCACGTCCGGTGTGATTCGCACC
CAGCGCCGGCTGGACCGGGAGAATGTGGCCGTGTACAACCTTTGGGCTCTGGCTGTGGATCGGGGCAG
TCCCACTCCCCTTAGCGCCTCGGTAGAAATCCAGGTGACCATCTTGGACATTAATGACAATGCCCCCA
TGTTTGAGAAGGACGAACTGGAGCTGTTTGTTGAGGAGAACAACCCAGTGGGGTCGGTGGTGGCAAAG
ATTCGTGCTAACGACCCTGATGAAGGCCCTAATGCCCAGATCATGTATCAGATTGTGGAAGGGGACAT
GCGGCATTTCTTCCAGCTGGACCTGCTCAACGGGGACCTGCGTGCCATGGTGGAGCTGGACTTTGAGG
TCCGGCGGGAGTATGTGCTGGTGGTGCAGGCCACGTCGGCTCCGCTGGTGAGCCGAGCCACGGTGCAC
ATCCTTCTCGTGGACCAGAATGACAACCCGCCTGTGCTGCCCGACTTCCAGATCCTCTTCAACAACTA

TGTCACCAACAAGTCCAACAGTTTCCCCACCGGCGTGATCGGCTGCATCCCGGCCCATGACCCCGACG
TGTCAGACAGCCTCAACTACACCTTCGTGCAGGGCAACGAGCTGCGCCTGTTGCTGCTGGACCCCGCC
ACGGGCGAACTGCAGCTCAGCCGCGACCTGGACAACAACCGGCCGCTGGAGGCGCTCATGGAGGTGTC
TGTGTCTGCAGATGGCATCCACAGCGTCACGGCCTTCTGCACCCTGCGTGTCACCATCATCACGGACG
ACATGCTGACCAACAGCATCACTGTCCGCCTGGAGAACATGTCCCAGGAGAAGTTCCTGTCCCCGCTG
CTGGCCCTCTTCGTGGAGGGGGTGGCCGCCGTGCTGTCCACCACCAAGGACGACGTCTTCGTCTTCAA
CGTCCAGAACGACACCGACGTCAGCTCCAACATCCTGAACGTGACCTTCTCGGCGCTGCTGCCTGGCG
GCGTCCGCGGCCAGTTCTTCCCGTCGGAGGACCTGCAGGAGCAGATCTACCTGAATCGGACGCTGCTG
ACCACCATCTCCACGCAGCGCGTGCTGCCCTTCGACGACAACATCTGCCTGCGCGAGCCCTGCGAGAA
CTACATGAAGTGCGTGTCCGTTCTGCGATTCGACAGCTCCGCGCCCTTCCTCAGCTCCACCACCGTGC
TCTTCCGGCCCATCCACCCCATCAACGGCCTGCGCTGCCGCTGCCCGCCCGGCTTCACCGGCGACTAC
TGCGAGACGGAGATCGACCTCTGCTACTCCGACCCGTGCGGCGCCAACGGCCGCTGCCGCAGCCGCGA
GGGCGGCTACACCTGCGAGTGCTTCGAGGACTTCACTGGAGAGCACTGTGAGGTGGATGCCCGCTCAG
GCCGCTGTGCCAACGGGGTGTGCAAGAACGGGGGCACCTGCGTGAACCTGCTCATCGGCGGCTTCCAC
TGCGTGTGTCCTCCTGGCGAGTATGAGAGGCCCTACTGTGAGGTGACCACCAGGAGCTTCCCGCCCCA
GTCCTTCGTCACCTTCCGGGGCCTGAGACAGCGCTTCCACTTCACCATCTCCCTCACGTTTGCCACTC
AGGAAAGGAACGGCTTGCTTCTCTACAACGGCCGCTTCAATGAGAAGCACGACTTCATCGCCCTGGAG
ATCGTGGACGAGCAGGTGCAGCTCACCTTCTCTGCAGGTGCAGGCGAGACAACAACGACCGTGGCACC
GAAGGTTCCCAGTGGTGTGAGTGACGGGCGGTGGCACTCTGTGCAGGTGCAGTACTACAACAAGGTAA
GATGGGCCCCACCACTTCCCCCTGGCCCCCAGCCCAATATTGGCCACCTGGGCCTGCCCCATGGGCCG
TCCGGGGAAAAGATGGCCGTGGTGACAGTGGATGATTGTGACACAACCATGGCTGTGCGCTTTGGAAA
GGACATCGGGAACTACAGCTGCGCTGCCCAGGGCACTCAGACCGGCTCCAAGAAGTCCCTGGATCTGA
CCGGCCCTCTACTCCTGGGGGGTGTCCCCAACCTGCCAGAAGACTTCCCAGTGCACAACCGGCAGTTC
GTGGGCTGCATGCGGAACCTGTCAGTCGACGGCAAAAATGTGGACATGGCCGGATTCATCGCCAACAA
TGGCACCCGGGAAGGCTGCGCTGCTCGGAGGAACTTCTGCGATGGGAGGCGGTGTCAGAATGGAGGCA
CCTGTGTCAACAGGTGGAATATGTATCTGTGTGAGTGTCCACTCCGATTCGGCGGGAAGAACTGTGAG
CAAGCCATGCCTCACCCCCAGCTCTTCAGCGGTGAGAGCGTCGTGTCCTGGAGTGACCTGAACATCAT
CATCTCTGTGCCCTGGTACCTGGGGCTCATGTTCCGGACCCGGAAGGAGGACAGCGTTCTGATGGAGG
CCACCAGTGGTGGGCCCACCAGCTTTCGCCTCCAGATCCTGAACAACTACCTCCAGTTTGAGGTGTCC
CACGGCCCCTCCGATGTGGAGTCCGTGATGCTGTCCGGGTTGCGGGTGACCGACGGGGAGTGGCACCA
CCTGCTGATCGAGCTGAAGAATGTTAAGGAGGACAGTGAGATGAAGCACCTGGTCACCATGACCTTGG
ACTATGGGATGGACCAGAACAAGGCAGATATCGGGGGCATGCTTCCCGGGCTGACGGTAAGGAGCGTG
GTGGTCGGAGGCGCCTCTGAAGACAAGGTCTCCGTGCGCCGTGGATTCCGAGGCTGCATGCAGGGAGT
GAGGATGGGGGGGACGCCCACCAACGTCGCCACCCTGAACATGAACAACGCACTCAAGGTCAGGGTGA
AGGACGGCTGTGATGTGGACGACCCCTGTACCTCGAGCCCCTGTCCCCCCAATAGCCGCTGCCACGAC
GCCTGGGAGGACTACAGCTGCGTCTGTGACAAAGGGTACCTTGGAATAAACTGTGTGGATGCCTGTCA
CCTGAACCCCTGCGAGAACATGGGGGCCTGCGTGCGCTCCCCCGGCTCCCCGCAGGGCTACGTGTGCG
AGTGTGGGCCCAGTCACTACGGGCCGTACTGTGAGAACAAACTCGACCTTCCGTGCCCCAGAGGCTGG
TGGGGGAACCCCGTCTGTGGACCCTGCCACTGTGCCGTCAGCAAAGGCTTTGATCCCGACTGTAATAA
GACCAACGGCCAGTGCCAATGCAAGGAGAATTACTACAAGCTCCTAGCCCAGGACACCTGTCTGCCCT
GCGACTGCTTCCCCCATGGCTCCCACAGCCGCACTTGCGACATGGCCACCGGGCAGTGTGCCTGCAAG
CCCGGCGTCATCGGCCGCCAGTGCAACCGCTGCGACAACCCGTTTGCCGAGGTCACCACGCTCGGCTG
TGAAGTGATCTACAATGGCTGTCCCAAAGCATTTGAGGCCGGCATCTGGTGGCCACAGACCAAGTTCG
GGCAGCCGGCTGCGGTGCCATGCCCTAAGGGATCCGTTGGAAATGCGGTCCGACACTGCAGCGGGGAG
AAGGGCTGGCTGCCCCCAGAGCTCTTTAACTGTACCACCATCTCCTTCGTGGACCTCAGGGCCATGAA
TGAGAAGCTGAGCCGCAATGAGACGCAGGTGGACGGCGCCAGGGCCCTGCAGCTGGTGAGGGCGCTGC
GCAGTGCTACACAGCACACGGGCACGCTCTTTGGCAATGACGTGCGCACGGCCTACCAGCTGCTGGGC
CACGTCCTTCAGCACGAGAGCTGGCAGCAGGGCTTCGACCTGGCAGCCACGCAGGACGCCGACTTTCA
CGAGGACGTCATCCACTCGGGCAGCGCCCTCCTGGCCCCAGCCACCAGGGCGGCGTGGGAGCAGATCC

AGCGGAGCGAGGGCGGCACGGCACAGCTGCTCCGGCGCCTCGAGGGCTACTTCAGCAACGTGGCACGC
AACGTGCGGCGGACGTACCTGCGGCCCTTCGTCATCGTCACCGCCAACATGGTTCTTGCTGTCGACAT
CTTTGACAAGTTCAACTTTACGGGAGCCAGGGTCCCGCGATTCGACACCATCCATGAAGAGTTCCCCA
GGGAGCTGGAGTCCTCCGTCTCCTTCCCAGCCGACTTCTTCAGACCACCTGAAGAAAAAGAAGGCCCC
CTGCTGAGGCCGGCTGGCCGGAGGACCACCCCGCAGACCACGCGCCCGGGGCCTGGCACCGAGAGGGA
GGCCCCGATCAGCAGGCGGAGGCGACACCCTGATGACGCTGGCCAGTTCGCCGTCGCTCTGGTCATCA
TTTACCGCACCCTGGGGCAGCTCCTGCCCGAGCGCTACGACCCCGACCGTCGCAGCCTCCGGTTGCCT
CACCGGCCCATCATTAATACCCCGATGGTGAGCACGCTGGTGTACAGCGAGGGGGCTCCGCTCCCGAG
ACCCCTGGAGAGGCCCGTCCTGGTGGAGTTCGCCCTGCTGGAGGTGGAGGAGCGAACCAAGCCTGTCT
GCGTGTTCTGGAACCACTCCCTGGCCGTTGGTGGGACGGGAGGGTGGTCTGCCCGGGGCTGCGAGCTC

CTGTCCAGGAACCGGACACATGTCGCCTGCCAGTGCAGCCACACAGCCAGCTTTGCGGTGCTCATGGA
TATCTCCAGGCGTGAGAACGGGGAGGTCCTGCCTCTGAAGATTGTCACCTATGCCGCTGTGTCCTTGT
CACTGGCAGCCCTGCTGGTGGCCTTCGTCCTCCTGAGCCTGGTCCGCATGCTGCGCTCCAACCTGCAC
AGCATTCACAAGCACCTCGCCGTGGCGCTCTTCCTCTCTCAGCTGGTGTTCGTGATTGGGATCAACCA
GACGGAAAACCCGTTTCTGTGCACAGTGGTTGCCATCCTCCTCCACTACATCTACATGAGCACCTTTG
CCTGGACCCTCGTGGAGAGCCTGCATGTCTACCGCATGCTGACCGAGGTGCGCAACATCGACACGGGG
CCCATGCGGTTCTACTACGTCGTGGGCTGGGGCATCCCGGCCATTGTCACAGGACTGGCGGTCGGCCT
GGACCCCCAGGGCTACGGGAACCCCGACTTCTGCTGGCTGTCGCTTCAAGACACCCTGATTTGGAGCT
TTGCGGGGCCCATCGGAGCTGTTATAATCATCAACACAGTCACTTCTGTCCTATCTGCAAAGGTTTCC
TGCCAAAGAAAGCACCATTATTATGGGAAAAAAGGGATCGTCTCCCTGCTGAGGACCGCATTCCTCCT
GCTGCTGCTCATCAGCGCCACCTGGCTGCTGGGGCTGCTGGCTGTGAACCGCGATGCACTGAGCTTTC
ACTACCTCTTCGCCATCTTCAGCGGCTTACAGGGCCCCTTCGTCCTCCTTTTCCACTGCGTGCTCAAC
CAGGAGGTCCGGAAGCACCTGAAGGGCGTGCTCGGCGGGAGGAAGCTGCACCTGGAGGACTCCGCCAC
CACCAGGGCCACCCTGCTGACGCGCTCCCTCAACTGCAACACCACCTTCGGTGACGGGCCTGACATGC
TGCGCACAGACTTGGGCGAGTCCACCGCCTCGCTGGACAGCATCGTCAGGGATGAAGGGATCCAGAAG
CTCGGCGTGTCCTCTGGGCTGGTGAGGGGCAGCCACGGAGAGCCAGACGCGTCCCTCATGCCCAGGAG
CTGCAAGGATCCCCCTGGCCACGATTCCGACTCAGATAGCGAGCTGTCCCTGGATGAGCAGAGCAGCT
CTTACGCCTCCTCACACTCGTCAGACAGCGAGGACGATGGGGTGGGAGCTGAGGAAAAATGGGACCCG
GCCAGGGGCGCCGTCCACAGCACCCCCAAAGGGGACGCTGTGGCCAACCACGTTCCGGCCGGCTGGCC
CGACCAGAGCCTGGCTGAGAGTGACAGTGAGGACCCCAGCGGCAAGCCCCGCCTGAAGGTGGAGACCA
AGGTCAGCGTGGAGCTGCACCGCGAGGAGCAGGGCAGTCACCGTGGAGAGTACCCCCCGGACCAGGAG
AGCGGGGGCGCAGCCAGGCTTGCTAGCAGCCAGCCCCCAGAGCAGAGGAGCATCTTGAAAAATAAAGT
CACCTACCCGCCGCCGCTGACGCTGACGGAGCAGACGCTGAAGGGCCGGCTCCGGGAGAAGCTGGCCG
ACTGTGAGCAGAGCCCCACATCCTCGCGCACGTCTTCCCTGGGCTCTGGCGGCCCCGACTGCGCCATC
ACAGTCAAGAGCCCTGGGAGGGAGCCGGGGCGTGACCACCTCAACGGGGTGGCCATGAATGTGCGCAC
TGGGAGCGCCCAGGCCGATGGCTCCGACTCTGAGAAACCGTGA

NOV21a, CG51965-01 Protein Sequence	SEQ ID NO: 276	3028 aa	MW at 330865.9kD

MAPPPPPVLPVLLLLAAAAALPAMGLRAAAWEPRVPGGTRAFALRPGCTYAVGAACTPRAPRELLDVG
RDGRLAGRRRVSGAGRPLPLQVRLVARSAPTALSRRLRARTHLPGCGARARLCGTGARLCGALCFPVP
GGCAAAQHSALAAPTTLPACRCPPRPRPRCPGRPICLPPGGSVRLRLLCALRRAAGAVRVGLALEAAT
AGTPSASPSPSPPLPPNLPEARAGPARRARRGTSGRGSLKFPMPNYQVALFENEPAGTLILQLHAHYT
IEGEEERVSYYMEGLFDERSRGYFRIDSATGAVSTDSVLDRETKETHVLRVKAVDYSTPPRSATTYIT
VLVKDTNDHSPVFEQSEYRERVRENLEVGYEVLTIRASDRDSPINANLRYRVLGGAWDVFQLNESSGV
VSTRAVLDREEAAEYQLLVEANDQGRNPGPLSATATVYIEVEDENDNYPQFSEQNYVVQVPEDVGLNT
AVLRVQATDRDQGQNAAIHYSILSGNVAGQFYLHSLSGILDVINPLDFEDVQKYSLSIKAQDGGRPPL
INSSGVVSVQVLDVNDNEPIFVSSPFQATVLENVPLGYPVVHIQAVDADSGENARLHYRLVDTASTFL
GGGSAGPKNPAPTPDFPFQIHNSSGWITVCAELDREEVEHYSFGVEAVDHGSPPMSSSTSVSITVLDV
NDNDPVFTQPTYELRLNEDAAVGSSVLTLQARDRDANSVITYQLTGGNTRNRFALSSQRGGGLITLAL
PLDYKQEQQYVLAVTASDGTRSHTAHVLINVTDANTHRPVFQSSHYTVSVSEDRPVGTSIATLSANDE
DTGENARITYVIQDPVPQFRIDPDSGTMYTMMELDYENQVAYTLTIMAQDNGIPQKSDTTTLEILILD
ANDNAPQFLWDFYQGSIFEDAPPSTSILQVSATDRDSGPNGRLLYTFQGGDDGDGDFYIEPTSGVIRT
QRRLDRENVAVYNLWALAVDRGSPTPLSASVEIQVTILDINDNAPMFEKDELELFVEENNPVGSVVAK
IRANDPDEGPNAQIMYQIVEGDMRHFFQLDLLNGDLRAMVELDFEVRREYVLVVQATSAPLVSRATVH
ILLVDQNDNPPVLPDFQILFNNYVTNKSNSFPTGVIGCIPAHDPDVSDSLNYTFVQGNELRLLLLDPA
TGELQLSRDLDNNRPLEALMEVSVSADGIHSVTAFCTLRVTIITDDMLTNSITVRLENMSQEKFLSPL
LALFVEGVAAVLSTTKDDVFVFNVQNDTDVSSNILNVTFSALLPGGVRGQFFPSEDLQEQIYLNRTLL
TTISTQRVLPFDDNICLREPCENYMKCVSVLRFDSSAPFLSSTTVLFRPIHPINGLRCRCPPGFTGDY
CETEIDLCYSDPCGANGRCRSREGGYTCECFEDFTGEHCEVDARSGRCANGVCKNGGTCVNLLIGGFH
CVCPPGEYERPYCEVTTRSFPPQSFVTFRGLRQRFHFTISLTFATQERNGLLLYNGRFNEKHDFIALE
IVDEQVQLTFSAGAGETTTTVAPKVPSGVSDGRWHSVQVQYYNKVRWAPPLPPGPQPNIGHLGLPHGP
SGEKMAVVTVDDCDTTMAVRFGKDIGNYSCAAQGTQTGSKKSLDLTGPLLLGGVPNLPEDFPVHNRQF
VGCMRNLSVDGKNVDMAGFIANNGTREGCAARRNFCDGRRCQNGGTCVNRWNMYLCECPLRFGGKNCE
QAMPHPQLFSGESVVSWSDLNIIISVPWYLGLMFRTRKEDSVLMEATSGGPTSFRLQILNNYLQFEVS
HGPSDVESVMLSGLRVTDGEWHHLLIELKNVKEDSEMKHLVTMTLDYGMDQNKADIGGMLPGLTVRSV
VVGGASEDKVSVRRGFRGCMQGVRMGGTPTNVATLNMNNALKVRVKDGCDVDDPCTSSPCPPNSRCHD
AWEDYSCVCDKGYLGINCVDACHLNPCENMGACVRSPGSPQGYVCECGPSHYGPYCENKLDLPCPRGW
WGNPVCGPCHCAVSKGFDPDCNKTNGQCQCKENYYKLLAQDTCLPCDCFPHGSHSRTCDMATGQCACK

PGVIGRQCNRCDNPFAEVTTLGCEVIYNGCPKAFEAGIWWPQTKFGQPAAVPCPKGSVGNAVRHCSGE
KGWLPPELFNCTTISFVDLRAMNEKLSRNETQVDGARALQLVRALRSATQHTGTLFGNDVRTAYQLLG
HVLQHESWQQGFDLAATQDADFHEDVIHSGSALLAPATRAAWEQIQRSEGGTAQLLRRLEGYFSNVAR
NVRRTYLRPFVIVTANMVLAVDIFDKFNFTGARVPRFDTIHEEFPRELESSVSFPADFFRPPEEKEGP
LLRPAGRRTTPQTTRPGPGTEREAPISRRRRHPDDAGQFAVALVIIYRTLGQLLPERYDPDRRSLRLP
HRPIINTPMVSTLVYSEGAPLPRPLERPVLVEFALLEVEERTKPVCVFWNHSLAVGGTGGWSARGCEL
LSRNRTHVACQCSHTASFAVLMDISRRENGEVLPLKIVTYAAVSLSLAALLVAFVLLSLVRMLRSNLH
SIHKHLAVALFLSQLVFVIGINQTENPFLCTVVAILLHYIYMSTFAWTLVESLHVYRMLTEVRNIDTG
PMRFYYVVGWGIPAIVTGLAVGLDPQGYGNPDFCWLSLQDTLIWSFAGPIGAVIIINTVTSVLSAKVS
CQRKHHYYGKKGIVSLLRTAFLLLLLISATWLLGLLAVNRDALSFHYLFAIFSGLQGPFVLLFHCVLN
QEVRKHLKGVLGGRKLHLEDSATTRATLLTRSLNCNTTFGDGPDMLRTDLGESTASLDSIVRDEGIQK
LGVSSGLVRGSHGEPDASLMPRSCKDPPGHDSDSDSELSLDEQSSSYASSHSSDSEDDGVGAEEKWDP
ARGAVHSTPKGDAVANHVPAGWPDQSLAESDSEDPSGKPRLKVETKVSVELHREEQGSHRGEYPPDQE
SGGAARLASSQPPEQRSILKNKVTYPPPLTLTEQTLKGRLREKLADCEQSPTSSRTSSLGSGGPDCAI
TVKSPGREPGRDHLNGVAMNVRTGSAQADGSDSEKP

NOV21b, 258076370 DNA Sequence	SEQ ID NO: 277	1263 bp
	ORF Start: at 1	ORF Stop: end of sequence

CTCAAGCTTGAAGTGATCTACAATGGCTGTCCCAAAGCATTTGAGGCCGGCATCTGGTGGCCACAGAC
CAAGTTCGGGCAGCCGGCTGCGGTGCCATGCCCTAAGGGATCCGTTGGAAATGCGGTCCGACACTGCA
GCGGGGAGAAGGGCTGGCTGCCCCCAGAGCTCTTTAACTGTACCACCATCTCCTTCGTGGACCTCAGG
GCCATGAATGAGAAGCTGAGCCGCAATGAGACGCAGGTGGACGGCGCCAGGGCCCTGCAGCTGGTGAG
GGCGCTGCGCAGTGCTACACAGCACACGGGCACGCTCTTTGGCAATGACGTGCGCACGGCCTACCAGC
TGCTGGGCCACGTCCTTCAGCACGAGAGCTGGCAGCAGGGCTTCGACCTGGCAGCCACGCAGGACGCC
GACTTTCACGAGGACGTCATCCACTCGGGCAGCGCCCTCCTGGCCCCAGCCACCAGGGCGGCGTGGGA
GCAGATCCAGCGGAGCGAGGGCGGCACGGCACAGCTGCTCCGGCGCCTCGAGGGCTACTTCAGCAACG
TGGCACGCAACGTGCGGCGGACGTACCTGCGGCCCTTCGTCATCGTCACCGCCAACATGATTCTTGCT
GTCGACATCTTTGACAAGTTCAACTTTACGGGAGCCAGGGTCCCGCGATTCGACACCATCCATGAAGA
GTTCCCCAGGGAGCTGGAGTCCTCCGTCTCCTTCCCAGCCGACTTCTTCAGACCACCTGAAGAAAAAG
AAGGCCCCCTGCTGAGGCCGGCTGGCCGGAGGACCACCCCGCAGACCACGCGCCCGGGGCCTGGCACC
GAGAGGGAGGCCCCGATCAGCAGGCGGAGGCGACACCCTGATGACGCTGGCCAGTTCGCCGTCGCTCT
GGTCATCATTTACCGCACCCTGGGGCAGCTCCTGCCCGAGCGCTACGACCCCGACCGTCGCAGCCTCC
GGTTGCCTCACCGGCCCATCATTAATACCCCGATGGTGAGCACGCTGGTGTACAGCGAGGGGGCTCCG
CTCCCGAGACCCCTGGAGAGGCCCGTCCTGGTGGAGTTCGCCCTGCTGGAGGTGGAGGAGCGAACCAA
GCCTGTCTGCGTGTTCTGGAACCACTCCCTGGCCGTTGGTGGGACGGGAGGGTGGTCTGCCCGGGGCT
GCGAGCTCCTGTCCAGGAACCGGACACATGTCGCCTGCCAGTGCAGCCACACAGCCAGCTTTGCGGTG
CTCATGGATATCTCCAGGCGTGAGAACGGGGAGAAGCTT

NOV21b, 258076370 Protein Sequence	SEQ ID NO: 278	421 aa	MW at 47178.1kD

LKLEVIYNGCPKAFEAGIWWPQTKFGQPAAVPCPKGSVGNAVRHCSGEKGWLPPELFNCTTISFVDLR
AMNEKLSRNETQVDGARALQLVRALRSATQHTGTLFGNDVRTAYQLLGHVLQHESWQQGFDLAATQDA
DFHEDVIHSGSALLAPATRAAWEQIQRSEGGTAQLLRRLEGYFSNVARNVRRTYLRPFVIVTANMILA
VDIFDKFNFTGARVPRFDTIHEEFPRELESSVSFPADFFRPPEEKEGPLLRPAGRRTTPQTTRPGPGT
EREAPISRRRRHPDDAGQFAVALVIIYRTLGQLLPERYDPDRRSLRLPHRPIINTPMVSTLVYSEGAP
LPRPLERPVLVEFALLEVEERTKPVCVFWNHSLAVGGTGGWSARGCELLSRNRTHVACQCSHTASFAV
LMDISRRENGEKL

NOV21c, 317619862 DNA Sequence	SEQ ID NO: 279	750 bp
	ORF Start: at 1	ORF Stop: end of sequence

CGCAAGCTTGTCCTGCCTCTGAAGATTGTCACCTATGCCGCTGTGTCCTTGTCACTGGCAGCCCTGCT
GGTGGCCTTCGTCCTCCTGAGCCTGGTCCGCATGCTGCGCTCCAACCTGCACAGCATTCACAAGCACC
TCGCCGTGGCGCTCTTCCTCTCTCAGCTGGTGTTCGTGATTGGGATCAACCAGACGGAAAACCCGTTT
CTGTGCACAGTGGTTGCCATCCTCCTCCACTACATCTACATGAGCACCTTTGCCTGGACCCTCGTGGA
GAGCCTGCATGTCTACCGCATGCTGACCGAGGTGCGCAACATCGACACGGGGCCCATGCGGTTCTACT
ACGTCGTGGGCTGGGGCATCCCGGCCATTGTCACAGGACTGGCGGTCGGCCTGGACCCCCAGGGCTAC
GGGAACCCCGACTTCTGCTGGCTGTCGCTTCAAGACACCCTGATTTGGAGCTTTGCGGGGCCCATCGG
AGCTGTTATAATCATCAACACAGTCACTTCTGTCCTATCTGCAAAGGTTTCCTGCCAAAGAAAGCACC

ATTATTATGGGAAAAAAGGGATCGTCTCCCTGCTGAGGACCGCATTCCTCCTGCTGCTGCTCATCAGC
GCCACCTGGCTGCTGGGGCTGCTGGCTGTGAACCGCGACGCACTGAGCTTTCACTACCTCTTCGCCAT
CTTCAGCGGCTTACAGGGCCCCTTCGTCCTCCTTTTCCACTGCGTGCTCAACCAGGAGGTCAAGCTTG
CG

NOV21c, 317619862 Protein Sequence	SEQ ID NO: 280	250 aa	MW at 27799.9kD

RKLVLPLKIVTYAAVSLSLAALLVAFVLLSLVRMLRSNLHSIHKHLAVALFLSQLVFVIGINQTENPF
LCTVVAILLHYIYMSTFAWTLVESLHVYRMLTEVRNIDTGPMRFYYVVGWGIPAIVTGLAVGLDPQGY
GNPDFCWLSLQDTLIWSFAGPIGAVIIINTVTSVLSAKVSCQRKHHYYGKKGIVSLLRTAFLLLLLIS
ATWLLGLLAVNRDALSFHYLFAIFSGLQGPFVLLFHCVLNQEVKLA

NOV21d, 317460050 DNA Sequence	SEQ ID NO: 281	2541 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTTACCAGGTGGCGTTGTTTGAGAACGAACCGGCGGGCACCCTCATCCTCCAGCTGCACGCGCA
CCACACCATCGAGGGCGAGGAGGAGCGCGTGAGCTATTACATGGAGGGGCTGTTCGACGAGCGCTCCC
GGGGCTACTTCCGAATCGACTCTGCCGCGGGCGCCGTGAGCACGGACAGCGTACTGGACCGCGAGACC
AAGGAGACGCACGTCCTCAGGGTGAAAGCCGTGGACTACAGTACGCCGCCGCGCTCGGCCACCACCTA
CATCACTGTCTTGGTCAAAGACACCAACGACCACAGCCCGGTCTTCGAGCAGTCGGAGTACCGCGAGC
GCGTGCGGGAGAACCTGGAGGTGGGCTACGAGGTGCTGACCATCCGCGCCAGCGACCGCGACTCGCCC
ATCAACGCCAACTTGCGTTACCGCGTGTTGGGGGGCGCGTGGGACGTCTTCCAGCTCAACGAGAGCTC
TGGCGTGGTGAGTACACGGGCGGTGCTGGACCGGGAGGAGGCGGCCGAGTACCAGCTGCTGGTGGAGG
CCAACGACCAGGGGCGCAATCCGGGCCCGCTCAGTGCCACGGCCACCGTGTACATCGAGGTGGAGGAC
GAGAACGACAACTACCCCCAGTTCAGCGAGCAGAACTACGTGGTCCAGGTGCCCGAGGACGTGGGGCT
CAACACGGCTGTGCTGCGAGTGCAGGCCACGGACCGGGACCAGGGCCAGAACGCGGCCATTCACTACA
GCATCCTCAGCGGGAACGTGGCCGGCCAGTTCTACCTGCACTCGCTGAGCGGGATCCTGGATGTGATC
AACCCCTTGGATTTCGAGGATGTCCAGAAATACTCGCTGAGCATTAAGGCCCAGGATGGGGGCCGGCC
CCCGCTCATCAATTCTTCAGGGGTGGTGTCTGTGCAGGTGCTGGATGTCAACGACAACGAGCCTATCT
TTGTGAGCAGCCCCTTCCAGGCCACGGTGCTGGAGAATGTGCCCCTGGGCTACCCCGTGGTGCACATT
CAGGCGGTGGACGCGGACTCTGGAGAGAACGCCCGGCTGCACTATCGCCTGGTGGACACGGCCTCCAC
CTTTCTGGGGGGCGGCAGCGCTGGGCCTAAGAATCCTGCCCCCACCCCTGACTTCCCCTTCCAGATCC
GCAACAGCTCCGGTTGGATCACAGTGTGTGCCGAGCTGGACCGCGAGGAGGTGGAGCACTACAGCTTC
GGGGTGGAGGCGGTGGACCACGGCTGGCCCCCCATGAGCTCCTCCACCAGCGTGTCCATCACGGTGCT
GGACGTGAATGACAACGACCCGGTGTTCACGCAGCCCACCTACGAGCTTCGTCTGAATGAGGATGCGG
CCGTGGGGAGCAGCGTGCTGACCCTGCAGGCCCGCGACCGTGACGCCAACAGTGTGATTACCTACCAG
CTCACAGGCGGCAACACCCGGAACCGCTTTGCACTCAGCAGCCAGAGAGGGGGCGGCCTCATCACCCT
GGCGCTACCTCTGGACTACAAGCAGGAGCAGCAGTACGTGCTGGCGGTGACAGCATCCGACGGCACAC
GGTCGCACACTGCGCATGTCCTAATCAACGTCACTGATGCCAACACCCACAGGCCTGTCTTTCAGAGC
TCCCATTACACAGTGAGTGTCAGTGAGGACAGGCCTGTGGGCACCTCCATTGCTACCCTCAGTGCCAA
CGATGAGGACACAGGAGAGAATGCCCGCATCACCTACGTGATTCAGGACCCCGTGCCGCAGTTCCGCA
TTGACCCCGACAGTGGCACCATGTACACCATGATGGAGCTGGACTATGAGAACCAGGTCGCCTACACG
CTGACCATCATGGCCCAGGACAACGGCATCCCGCAGAAATCAGACACCACCACCCTAGAGATCCTCAT
CCTCGATGCCAATGACAATGCACCCCAGTTCCTGTGGGATTTCTACCAGGGTTCCATCTTTGAGGATG
CTCCACCCTCGACCAGCATCCTCCAGGTCTCTGCCACGGACCGGGACTCAGGTCCCAATGGGCGTCTG
CTGTACACCTTCCAGGGTGGGGACGACGGCGATGGGGACTTCTACATCGAGCCCACGTCCGGTGTGAT
TCGCACCCAGCGCCGGCTGGACCGGGAGAATGTGGCCGTGTACAACCTTTGGGCTCTGGCTGTGGATC
GGGGCAGTCCCACTCCCCTTAGCGCCTCGGTAGAAATCCAGGTGACCATCTTGGACATTAATGACAAT
GCCCCCATGTTTGAGAAGGACGAACTGGAGCTGTTTGTTGAGGAGAACAACCCAGTGGGGTCGGTGGT
GGCAAAGATTCGTGCTAACGACCCTGATGAAGGCCCTAATGCCCAGATCATGTATCAGATTGTGGAAG
GGGACATGCGGCATTTCTTCCAGCTGGACCTGCTCAACGGGGACCTGCGTGCCATGGTGGAGCTGGAC
TTTGAGGTCCGGCGGGAGTATGTGCTGGTGGTGCAGGCCACGTCGGCTCCGCTGGTGAGCCGAGCCAC
GGTGCACATCCTTCTCGTGCTCGAG

NOV21d, 317460050 Protein Sequence	SEQ ID NO: 282	847 aa	MW at 93611.5kD

KLYQVALFENEPAGTLILQLHAHHTIEGEEERVSYYMEGLFDERSRGYFRIDSAAGAVSTDSVLDRET
KETHVLRVKAVDYSTPPRSATTYITVLVKDTNDHSPVFEQSEYRERVRENLEVGYEVLTIRASDRDSP
INANLRYRVLGGAWDVFQLNESSGVVSTRAVLDREEAAEYQLLVEANDQGRNPGPLSATATVYIEVED
ENDNYPQFSEQNYVVQVPEDVGLNTAVLRVQATDRDQGQNAAIHYSILSGNVAGQFYLHSLSGILDVI

NPLDFEDVQKYSLSIKAQDGGRPPLINSSGVVSVQVLDVNDNEPIFVSSPFQATVLENVPLGYPVVHI
QAVDADSGENARLHYRLVDTASTFLGGGSAGPKNPAPTPDFPFQIRNSSGWITVCAELDREEVEHYSF
GVEAVDHGWPPMSSSTSVSITVLDVNDNDPVFTQPTYELRLNEDAAVGSSVLTLQARDRDANSVITYQ
LTGGNTRNRFALSSQRGGGLITLALPLDYKQEQQYVLAVTASDGTRSHTAHVLINVTDANTHRPVFQS
SHYTVSVSEDRPVGTSIATLSANDEDTGENARITYVIQDPVPQFRIDPDSGTMYTMMELDYENQVAYT
LTIMAQDNGIPQKSDTTTLEILILDANDNAPQFLWDFYQGSIFEDAPPSTSILQVSATDRDSGPNGRL
LYTFQGGDDGDGDFYIEPTSGVIRTQRRLDRENVAVYNLWALAVDRGSPTPLSASVEIQVTILDINDN
APMFEKDELELFVEENNPVGSVVAKIRANDPDEGPNAQIMYQIVEGDMRHFFQLDLLNGDLRAMVELD
FEVRREYVLVVQATSAPLVSRATVHILLVLE

NOV21e, SNP13382483 of CG51965-01, DNA Sequence	SEQ ID NO: 283	9087 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 9085
	SNP Pos: 8435	SNP Change: G to C

ATGGCGCCGCCGCCGCCGCCCGTGCTGCCCGTGCTGCTGCTCCTGGCCGCCGCCGCCGCCCTGCCGGC
GATGGGGCTGCGAGCGGCCGCCTGGGAGCCGCGCGTACCCGGCGGGACCCGCGCCTTCGCCCTCCGGC
CCGGCTGTACCTACGCGGTGGGCGCCGCTTGCACGCCCCGGGCGCCGCGGGAGCTGCTGGACGTGGGC
CGCGATGGGCGGCTGGCAGGACGTCGGCGCGTCTCGGGCGCGGGGCGCCCGCTGCCGCTGCAAGTCCG
CTTGGTGGCCCGCAGTGCCCCGACGGCGCTGAGCCGCCGCCTGCGGGCGCGCACGCACCTTCCCGGCT
GCGGAGCCCGTGCCCGGCTCTGCGGAACCGGTGCCCGGCTCTGCGGGGCGCTCTGCTTCCCCGTCCCC
GGCGGCTGCGCGGCCGCGCAGCATTCGGCGCTCGCAGCTCCGACCACCTTACCCGCCTGCCGCTGCCC
GCCGCGCCCCAGGCCCCGCTGTCCCGGCCGTCCCATCTGCCTGCCGCCGGGCGGCTCGGTCCGCCTGC
GTCTGCTGTGCGCCCTGCGGCGCGCGGCTGGCGCCGTCCGGGTGGGACTGGCGCTGGAGGCCGCCACC
GCGGGGACGCCCTCCGCGTCGCCATCCCCATCGCCGCCCCTGCCGCCGAACTTGCCCGAAGCCCGGGC
GGGGCCGGCGCGACGGGCCCGGCGGGGCACGAGCGGCAGAGGGAGCCTGAAGTTTCCGATGCCCAACT
ACCAGGTGGCGTTGTTTGAGAACGAACCGGCGGGCACCCTCATCCTCCAGCTGCACGCGCACTACACC
ATCGAGGGCGAGGAGGAGCGCGTGAGCTATTACATGGAGGGGCTGTTCGACGAGCGCTCCCGGGGCTA
CTTCCGAATCGACTCTGCCACGGGCGCCGTGAGCACGGACAGCGTACTGGACCGCGAGACCAAGGAGA
CGCACGTCCTCAGGGTGAAAGCCGTGGACTACAGTACGCCGCCGCGCTCGGCCACCACCTACATCACT
GTCTTGGTCAAAGACACCAACGACCACAGCCCGGTCTTCGAGCAGTCGGAGTACCGCGAGCGCGTGCG
GGAGAACCTGGAGGTGGGCTACGAGGTGCTGACCATCCGCGCCAGCGACCGCGACTCGCCCATCAACG
CCAACTTGCGTTACCGCGTGTTGGGGGGCGCGTGGGACGTCTTCCAGCTCAACGAGAGCTCTGGCGTG
GTGAGCACACGGGCGGTGCTGGACCGGGAGGAGGCGGCCGAGTACCAGCTCCTGGTGGAGGCCAACGA
CCAGGGGCGCAATCCGGGCCCGCTCAGTGCCACGGCCACCGTGTACATCGAGGTGGAGGACGAGAACG
ACAACTACCCCCAGTTCAGCGAGCAGAACTACGTGGTCCAGGTGCCCGAGGACGTGGGGCTCAACACG
GCTGTGCTGCGAGTGCAGGCCACGGACCGGGACCAGGGCCAGAACGCGGCCATTCACTACAGCATCCT
CAGCGGGAACGTGGCCGGCCAGTTCTACCTGCACTCGCTGAGCGGGATCCTGGATGTGATCAACCCCT
TGGATTTCGAGGATGTCCAGAAATACTCGCTGAGCATTAAGGCCCAGGATGGGGGCCGGCCCCCGCTC
ATCAATTCTTCAGGGGTGGTGTCTGTGCAGGTGCTGGATGTCAACGACAACGAGCCTATCTTTGTGAG
CAGCCCCTTCCAGGCCACGGTGCTGGAGAATGTGCCCCTGGGCTACCCCGTGGTGCACATTCAGGCGG
TGGACGCGGACTCTGGAGAGAACGCCCGGCTGCACTATCGCCTGGTGGACACGGCCTCCACCTTTCTG
GGGGGCGGCAGCGCTGGGCCTAAGAATCCTGCCCCCACCCCTGACTTCCCCTTCCAGATCCACAACAG
CTCCGGTTGGATCACAGTGTGTGCCGAGCTGGACCGCGAGGAGGTGGAGCACTACAGCTTCGGGGTGG
AGGCGGTGGACCACGGCTCGCCCCCCATGAGCTCCTCCACCAGCGTGTCCATCACGGTGCTGGACGTG
AATGACAACGACCCGGTGTTCACGCAGCCCACCTACGAGCTTCGTCTGAATGAGGATGCGGCCGTGGG
GAGCAGCGTGCTGACCCTGCAGGCCCGCGACCGTGACGCCAACAGTGTGATTACCTACCAGCTCACAG
GCGGCAACACCCGGAACCGCTTTGCACTCAGCAGCCAGAGAGGGGGCGGCCTCATCACCCTGGCGCTA
CCTCTGGACTACAAGCAGGAGCAGCAGTACGTGCTGGCGGTGACAGCATCCGACGGCACACGGTCGCA
CACTGCGCATGTCCTAATCAACGTCACTGATGCCAACACCCACAGGCCTGTCTTTCAGAGCTCCCATT
ACACAGTGAGTGTCAGTGAGGACAGGCCTGTGGGCACCTCCATTGCTACCCTCAGTGCCAACGATGAG
GACACAGGAGAGAATGCCCGCATCACCTACGTGATTCAGGACCCCGTGCCGCAGTTCCGCATTGACCC
CGACAGTGGCACCATGTACACCATGATGGAGCTGGACTATGAGAACCAGGTCGCCTACACGCTGACCA
TCATGGCCCAGGACAACGGCATCCCGCAGAAATCAGACACCACCACCCTAGAGATCCTCATCCTCGAT
GCCAATGACAATGCACCCCAGTTCCTGTGGGATTTCTACCAGGGTTCCATCTTTGAGGATGCTCCACC
CTCGACCAGCATCCTCCAGGTCTCTGCCACGGACCGGGACTCAGGTCCCAATGGGCGTCTGCTGTACA
CCTTCCAGGGTGGGGACGACGGCGATGGGGACTTCTACATCGAGCCCACGTCCGGTGTGATTCGCACC
CAGCGCCGGCTGGACCGGGAGAATGTGGCCGTGTACAACCTTTGGGCTCTGGCTGTGGATCGGGGCAG
TCCCACTCCCCTTAGCGCCTCGGTAGAAATCCAGGTGACCATCTTGGACATTAATGACAATGCCCCCA
TGTTTGAGAAGGACGAACTGGAGCTGTTTGTTGAGGAGAACAACCCAGTGGGGTCGGTGGTGGCAAAG

ATTCGTGCTAACGACCCTGATGAAGGCCCTAATGCCCAGATCATGTATCAGATTGTGGAAGGGGACAT
GCGGCATTTCTTCCAGCTGGACCTGCTCAACGGGGACCTGCGTGCCATGGTGGAGCTGGACTTTGAGG
TCCGGCGGGAGTATGTGCTGGTGGTGCAGGCCACGTCGGCTCCGCTGGTGAGCCGAGCCACGGTGCAC
ATCCTTCTCGTGGACCAGAATGACAACCCGCCTGTGCTGCCCGACTTCCAGATCCTCTTCAACAACTA

TGTCACCAACAAGTCCAACAGTTTCCCCACCGGCGTGATCGGCTGCATCCCGGCCCATGACCCCGACG
TGTCAGACAGCCTCAACTACACCTTCGTGCAGGGCAACGAGCTGCGCCTGTTGCTGCTGGACCCCGCC
ACGGGCGAACTGCAGCTCAGCCGCGACCTGGACAACAACCGGCCGCTGGAGGCGCTCATGGAGGTGTC
TGTGTCTGCAGATGGCATCCACAGCGTCACGGCCTTCTGCACCCTGCGTGTCACCATCATCACGGACG
ACATGCTGACCAACAGCATCACTGTCCGCCTGGAGAACATGTCCCAGGAGAAGTTCCTGTCCCCGCTG
CTGGCCCTCTTCGTGGAGGGGGTGGCCGCCGTGCTGTCCACCACCAAGGACGACGTCTTCGTCTTCAA
CGTCCAGAACGACACCGACGTCAGCTCCAACATCCTGAACGTGACCTTCTCGGCGCTGCTGCCTGGCG
GCGTCCGCGGCCAGTTCTTCCCGTCGGAGGACCTGCAGGAGCAGATCTACCTGAATCGGACGCTGCTG
ACCACCATCTCCACGCAGCGCGTGCTGCCCTTCGACGACAACATCTGCCTGCGCGAGCCCTGCGAGAA
CTACATGAAGTGCGTGTCCGTTCTGCGATTCGACAGCTCCGCGCCCTTCCTCAGCTCCACCACCGTGC
TCTTCCGGCCCATCCACCCCATCAACGGCCTGCGCTGCCGCTGCCCGCCCGGCTTCACCGGCGACTAC
TGCGAGACGGAGATCGACCTCTGCTACTCCGACCCGTGCGGCGCCAACGGCCGCTGCCGCAGCCGCGA
GGGCGGCTACACCTGCGAGTGCTTCGAGGACTTCACTGGAGAGCACTGTGAGGTGGATGCCCGCTCAG
GCCGCTGTGCCAACGGGGTGTGCAAGAACGGGGGCACCTGCGTGAACCTGCTCATCGGCGGCTTCCAC
TGCGTGTGTCCTCCTGGCGAGTATGAGAGGCCCTACTGTGAGGTGACCACCAGGAGCTTCCCGCCCCA
GTCCTTCGTCACCTTCCGGGGCCTGAGACAGCGCTTCCACTTCACCATCTCCCTCACGTTTGCCACTC
AGGAAAGGAACGGCTTGCTTCTCTACAACGGCCGCTTCAATGAGAAGCACGACTTCATCGCCCTGGAG
ATCGTGGACGAGCAGGTGCAGCTCACCTTCTCTGCAGGTGCAGGCGAGACAACAACGACCGTGGCACC
GAAGGTTCCCAGTGGTGTGAGTGACGGGCGGTGGCACTCTGTGCAGGTGCAGTACTACAACAAGGTAA
GATGGGCCCCACCACTTCCCCCTGGCCCCCAGCCCAATATTGGCCACCTGGGCCTGCCCCATGGGCCG
TCCGGGGAAAAGATGGCCGTGGTGACAGTGGATGATTGTGACACAACCATGGCTGTGCGCTTTGGAAA
GGACATCGGGAACTACAGCTGCGCTGCCCAGGGCACTCAGACCGGCTCCAAGAAGTCCCTGGATCTGA
CCGGCCCTCTACTCCTGGGGGGTGTCCCCAACCTGCCAGAAGACTTCCCAGTGCACAACCGGCAGTTC
GTGGGCTGCATGCGGAACCTGTCAGTCGACGGCAAAAATGTGGACATGGCCGGATTCATCGCCAACAA
TGGCACCCGGGAAGGCTGCGCTGCTCGGAGGAACTTCTGCGATGGGAGGCGGTGTCAGAATGGAGGCA
CCTGTGTCAACAGGTGGAATATGTATCTGTGTGAGTGTCCACTCCGATTCGGCGGGAAGAACTGTGAG
CAAGCCATGCCTCACCCCCAGCTCTTCAGCGGTGAGAGCGTCGTGTCCTGGAGTGACCTGAACATCAT
CATCTCTGTGCCCTGGTACCTGGGGCTCATGTTCCGGACCCGGAAGGAGGACAGCGTTCTGATGGAGG
CCACCAGTGGTGGGCCCACCAGCTTTCGCCTCCAGATCCTGAACAACTACCTCCAGTTTGAGGTGTCC
CACGGCCCCTCCGATGTGGAGTCCGTGATGCTGTCCGGGTTGCGGGTGACCGACGGGGAGTGGCACCA
CCTGCTGATCGAGCTGAAGAATGTTAAGGAGGACAGTGAGATGAAGCACCTGGTCACCATGACCTTGG
ACTATGGGATGGACCAGAACAAGGCAGATATCGGGGGCATGCTTCCCGGGCTGACGGTAAGGAGCGTG
GTGGTCGGAGGCGCCTCTGAAGACAAGGTCTCCGTGCGCCGTGGATTCCGAGGCTGCATGCAGGGAGT
GAGGATGGGGGGGACGCCCACCAACGTCGCCACCCTGAACATGAACAACGCACTCAAGGTCAGGGTGA
AGGACGGCTGTGATGTGGACGACCCCTGTACCTCGAGCCCCTGTCCCCCCAATAGCCGCTGCCACGAC
GCCTGGGAGGACTACAGCTGCGTCTGTGACAAAGGGTACCTTGGAATAAACTGTGTGGATGCCTGTCA
CCTGAACCCCTGCGAGAACATGGGGGCCTGCGTGCGCTCCCCCGGCTCCCCGCAGGGCTACGTGTGCG
AGTGTGGGCCCAGTCACTACGGGCCGTACTGTGAGAACAAACTCGACCTTCCGTGCCCCAGAGGCTGG
TGGGGGAACCCCGTCTGTGGACCCTGCCACTGTGCCGTCAGCAAAGGCTTTGATCCCGACTGTAATAA
GACCAACGGCCAGTGCCAATGCAAGGAGAATTACTACAAGCTCCTAGCCCAGGACACCTGTCTGCCCT
GCGACTGCTTCCCCCATGGCTCCCACAGCCGCACTTGCGACATGGCCACCGGGCAGTGTGCCTGCAAG
CCCGGCGTCATCGGCCGCCAGTGCAACCGCTGCGACAACCCGTTTGCCGAGGTCACCACGCTCGGCTG
TGAAGTGATCTACAATGGCTGTCCCAAAGCATTTGAGGCCGGCATCTGGTGGCCACAGACCAAGTTCG
GGCAGCCGGCTGCGGTGCCATGCCCTAAGGGATCCGTTGGAAATGCGGTCCGACACTGCAGCGGGGAG
AAGGGCTGGCTGCCCCCAGAGCTCTTTAACTGTACCACCATCTCCTTCGTGGACCTCAGGGCCATGAA
TGAGAAGCTGAGCCGCAATGAGACGCAGGTGGACGGCGCCAGGGCCCTGCAGCTGGTGAGGGCGCTGC
GCAGTGCTACACAGCACACGGGCACGCTCTTTGGCAATGACGTGCGCACGGCCTACCAGCTGCTGGGC
CACGTCCTTCAGCACGAGAGCTGGCAGCAGGGCTTCGACCTGGCAGCCACGCAGGACGCCGACTTTCA
CGAGGACGTCATCCACTCGGGCAGCGCCCTCCTGGCCCCAGCCACCAGGGCGGCGTGGGAGCAGATCC

AGCGGAGCGAGGGCGGCACGGCACAGCTGCTCCGGCGCCTCGAGGGCTACTTCAGCAACGTGGCACGC
AACGTGCGGCGGACGTACCTGCGGCCCTTCGTCATCGTCACCGCCAACATGGTTCTTGCTGTCGACAT
CTTTGACAAGTTCAACTTTACGGGAGCCAGGGTCCCGCGATTCGACACCATCCATGAAGAGTTCCCCA
GGGAGCTGGAGTCCTCCGTCTCCTTCCCAGCCGACTTCTTCAGACCACCTGAAGAAAAAGAAGGCCCC
CTGCTGAGGCCGGCTGGCCGGAGGACCACCCCGCAGACCACGCGCCCGGGGCCTGGCACCGAGAGGGA

GGCCCCGATCAGCAGGCGGAGGCGACACCCTGATGACGCTGGCCAGTTCGCCGTCGCTCTGGTCATCA
TTTACCGCACCCTGGGGCAGCTCCTGCCCGAGCGCTACGACCCCGACCGTCGCAGCCTCCGGTTGCCT
CACCGGCCCATCATTAATACCCCGATGGTGAGCACGCTGGTGTACAGCGAGGGGGCTCCGCTCCCGAG
ACCCCTGGAGAGGCCCGTCCTGGTGGAGTTCGCCCTGCTGGAGGTGGAGGAGCGAACCAAGCCTGTCT
GCGTGTTCTGGAACCACTCCCTGGCCGTTGGTGGGACGGGAGGGTGGTCTGCCCGGGGCTGCGAGCTC
CTGTCCAGGAACCGGACACATGTCGCCTGCCAGTGCAGCCACACAGCCAGCTTTGCGGTGCTCATGGA
TATCTCCAGGCGTGAGAACGGGGAGGTCCTGCCTCTGAAGATTGTCACCTATGCCGCTGTGTCCTTGT
CACTGGCAGCCCTGCTGGTGGCCTTCGTCCTCCTGAGCCTGGTCCGCATGCTGCGCTCCAACCTGCAC
AGCATTCACAAGCACCTCGCCGTGGCGCTCTTCCTCTCTCAGCTGGTGTTCGTGATTGGGATCAACCA
GACGGAAAACCCGTTTCTGTGCACAGTGGTTGCCATCCTCCTCCACTACATCTACATGAGCACCTTTG
CCTGGACCCTCGTGGAGAGCCTGCATGTCTACCGCATGCTGACCGAGGTGCGCAACATCGACACGGGG
CCCATGCGGTTCTACTACGTCGTGGGCTGGGGCATCCCGGCCATTGTCACAGGACTGGCGGTCGGCCT
GGACCCCCAGGGCTACGGGAACCCCGACTTCTGCTGGCTGTCGCTTCAAGACACCCTGATTTGGAGCT
TTGCGGGGCCCATCGGAGCTGTTATAATCATCAACACAGTCACTTCTGTCCTATCTGCAAAGGTTTCC
TGCCAAAGAAAGCACCATTATTATGGGAAAAAAGGGATCGTCTCCCTGCTGAGGACCGCATTCCTCCT
GCTGCTGCTCATCAGCGCCACCTGGCTGCTGGGGCTGCTGGCTGTGAACCGCGATGCACTGAGCTTTC
ACTACCTCTTCGCCATCTTCAGCGGCTTACAGGGCCCCTTCGTCCTCCTTTTCCACTGCGTGCTCAAC
CAGGAGGTCCGGAAGCACCTGAAGGGCGTGCTCGGCGGGAGGAAGCTGCACCTGGAGGACTCCGCCAC
CACCAGGGCCACCCTGCTGACGCGCTCCCTCAACTGCAACACCACCTTCGGTGACGGGCCTGACATGC
TGCGCACAGACTTGGGCGAGTCCACCGCCTCGCTGGACAGCATCGTCAGGGATGAAGGGATCCAGAAG
CTCGGCGTGTCCTCTGGGCTGGTGAGGGGCAGCCACGGAGAGCCAGACGCGTCCCTCATGCCCAGGAG
CTCCAAGGATCCCCCTGGCCACGATTCCGACTCAGATAGCGAGCTGTCCCTGGATGAGCAGAGCAGCT
CTTACGCCTCCTCACACTCGTCAGACAGCGAGGACGATGGGGTGGGAGCTGAGGAAAAATGGGACCCG
GCCAGGGGCGCCGTCCACAGCACCCCCAAAGGGGACGCTGTGGCCAACCACGTTCCGGCCGGCTGGCC
CGACCAGAGCCTGGCTGAGAGTGACAGTGAGGACCCCAGCGGCAAGCCCCGCCTGAAGGTGGAGACCA
AGGTCAGCGTGGAGCTGCACCGCGAGGAGCAGGGCAGTCACCGTGGAGAGTACCCCCCGGACCAGGAG
AGCGGGGGCGCAGCCAGGCTTGCTAGCAGCCAGCCCCCAGAGCAGAGGAGCATCTTGAAAAATAAAGT
CACCTACCCGCCGCCGCTGACGCTGACGGAGCAGACGCTGAAGGGCCGGCTCCGGGAGAAGCTGGCCG
ACTGTGAGCAGAGCCCCACATCCTCGCGCACGTCTTCCCTGGGCTCTGGCGGCCCCGACTGCGCCATC
ACAGTCAAGAGCCCTGGGAGGGAGCCGGGGCGTGACCACCTCAACGGGGTGGCCATGAATGTGCGCAC
TGGGAGCGCCCAGGCCGATGGCTCCGACTCTGAGAAACCGTGA

NOV21e, SNP13382483 of CG51965-01, Protein Sequence	SEQ ID NO: 284	3028 aa	MW at 330849.8kD
	SNP Pos: 2812		SNP Change: Cys to Ser

MAPPPPPVLPVLLLLAAAAALPAMGLRAAAWEPRVPGGTRAFALRPGCTYAVGAACTPRAPRELLDVG
RDGRLAGRRRVSGAGRPLPLQVRLVARSAPTALSRRLRARTHLPGCGARARLCGTGARLCGALCFPVP
GGCAAAQHSALAAPTTLPACRCPPRPRPRCPGRPICLPPGGSVRLRLLCALRRAAGAVRVGLALEAAT
AGTPSASPSPSPPLPPNLPEARAGPARRARRGTSGRGSLKFPMPNYQVALFENEPAGTLILQLHAHYT
IEGEEERVSYYMEGLFDERSRGYFRIDSATGAVSTDSVLDRETKETHVLRVKAVDYSTPPRSATTYIT
VLVKDTNDHSPVFEQSEYRERVRENLEVGYEVLTIRASDRDSPINANLRYRVLGGAWDVFQLNESSGV
VSTRAVLDREEAAEYQLLVEANDQGRNPGPLSATATVYIEVEDENDNYPQFSEQNYVVQVPEDVGLNT
AVLRVQATDRDQGQNAAIHYSILSGNVAGQFYLHSLSGILDVINPLDFEDVQKYSLSIKAQDGGRPPL
INSSGVVSVQVLDVNDNEPIFVSSPFQATVLENVPLGYPVVHIQAVDADSGENARLHYRLVDTASTFL
GGGSAGPKNPAPTPDFPFQIHNSSGWITVCAELDREEVEHYSFGVEAVDHGSPPMSSSTSVSITVLDV
NDNDPVFTQPTYELRLNEDAAVGSSVLTLQARDRDANSVITYQLTGGNTRNRFALSSQRGGGLITLAL
PLDYKQEQQYVLAVTASDGTRSHTAHVLINVTDANTHRPVFQSSHYTVSVSEDRPVGTSIATLSANDE
DTGENARITYVIQDPVPQFRIDPDSGTMYTMMELDYENQVAYTLTIMAQDNGIPQKSDTTTLEILILD
ANDNAPQFLWDFYQGSIFEDAPPSTSILQVSATDRDSGPNGRLLYTFQGGDDGDGDFYIEPTSGVIRT
QRRLDRENVAVYNLWALAVDRGSPTPLSASVEIQVTILDINDNAPMFEKDELELFVEENNPVGSVVAK
IRANDPDEGPNAQIMYQIVEGDMRHFFQLDLLNGDLRAMVELDFEVRREYVLVVQATSAPLVSRATVH
ILLVDQNDNPPVLPDFQILFNNYVTNKSNSFPTGVIGCIPAHDPDVSDSLNYTFVQGNELRLLLLDPA
TGELQLSRDLDNNRPLEALMEVSVSADGIHSVTAFCTLRVTIITDDMLTNSITVRLENMSQEKFLSPL
LALFVEGVAAVLSTTKDDVFVFNVQNDTDVSSNILNVTFSALLPGGVRGQFFPSEDLQEQIYLNRTLL
TTISTQRVLPFDDNICLREPCENYMKCVSVLRFDSSAPFLSSTTVLFRPIHPINGLRCRCPPGFTGDY
CETEIDLCYSDPCGANGRCRSREGGYTCECFEDFTGEHCEVDARSGRCANGVCKNGGTCVNLLIGGFH
CVCPPGEYERPYCEVTTRSFPPQSFVTFRGLRQRFHFTISLTFATQERNGLLLYNGRFNEKHDFIALE
IVDEQVQLTFSAGAGETTTTVAPKVPSGVSDGRWHSVQVQYYNKVRWAPPLPPGPQPNIGHLGLPHGP
SGEKMAVVTVDDCDTTMAVRFGKDIGNYSCAAQGTQTGSKKSLDLTGPLLLGGVPNLPEDFPVHNRQF
VGCMRNLSVDGKNVDMAGFIANNGTREGCAARRNFCDGRRCQNGGTCVNRWNMYLCECPLRFGGKNCE

QAMPHPQLFSGESVVSWSDLNIIISVPWYLGLMFRTRKEDSVLMEATSGGPTSFRLQILNNYLQFEVS
HGPSDVESVMLSGLRVTDGEWHHLLIELKNVKEDSEMKHLVTMTLDYGMDQNKADIGGMLPGLTVRSV
VVGGASEDKVSVRRGFRGCMQGVRMGGTPTNVATLNMNNALKVRVKDGCDVDDPCTSSPCPPNSRCHD
AWEDYSCVCDKGYLGINCVDACHLNPCENMGACVRSPGSPQGYVCECGPSHYGPYCENKLDLPCPRGW
WGNPVCGPCHCAVSKGFDPDCNKTNGQCQCKENYYKLLAQDTCLPCDCFPHGSHSRTCDMATGQCACK
PGVIGRQCNRCDNPFAEVTTLGCEVIYNGCPKAFEAGIWWPQTKFGQPAAVPCPKGSVGNAVRHCSGE
KGWLPPELFNCTTISFVDLRAMNEKLSRNETQVDGARALQLVRALRSATQHTGTLFGNDVRTAYQLLG
HVLQHESWQQGFDLAATQDADFHEDVIHSGSALLAPATRAAWEQIQRSEGGTAQLLRRLEGYFSNVAR
NVRRTYLRPFVIVTANMVLAVDIFDKFNFTGARVPRFDTIHEEFPRELESSVSFPADFFRPPEEKEGP
LLRPAGRRTTPQTTRPGPGTEREAPISRRRRHPDDAGQFAVALVIIYRTLGQLLPERYDPDRRSLRLP
HRPIINTPMVSTLVYSEGAPLPRPLERPVLVEFALLEVEERTKPVCVFWNHSLAVGGTGGWSARGCEL
LSRNRTHVACQCSHTASFAVLMDISRRENGEVLPLKIVTYAAVSLSLAALLVAFVLLSLVRMLRSNLH
SIHKHLAVALFLSQLVFVIGINQTENPFLCTVVAILLHYIYMSTFAWTLVESLHVYRMLTEVRNIDTG
PMRFYYVVGWGIPAIVTGLAVGLDPQGYGNPDFCWLSLQDTLIWSFAGPIGAVIIINTVTSVLSAKVS
CQRKHHYYGKKGIVSLLRTAFLLLLLISATWLLGLLAVNRDALSFHYLFAIFSGLQGPFVLLFHCVLN
QEVRKHLKGVLGGRKLHLEDSATTRATLLTRSLNCNTTFGDGPDMLRTDLGESTASLDSIVRDEGIQK
LGVSSGLVRGSHGEPDASLMPRSSKDPPGHDSDSDSELSLDEQSSSYASSHSSDSEDDGVGAEEKWDP
ARGAVHSTPKGDAVANHVPAGWPDQSLAESDSEDPSGKPRLKVETKVSVELHREEQGSHRGEYPPDQE
SGGAARLASSQPPEQRSILKNKVTYPPPLTLTEQTLKGRLREKLADCEQSPTSSRTSSLGSGGPDCAI
TVKSPGREPGRDHLNGVAMNVRTGSAQADGSDSEKP

NOV21f, SNP13382484 of CG51965-01, DNA Sequence	SEQ ID NO: 285	9087 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 9085
	SNP Pos: 8592	SNP Change: C to A

ATGGCGCCGCCGCCGCCGCCCGTGCTGCCCGTGCTGCTGCTCCTGGCCGCCGCCGCCGCCCTGCCGGC
GATGGGGCTGCGAGCGGCCGCCTGGGAGCCGCGCGTACCCGGCGGGACCCGCGCCTTCGCCCTCCGGC
CCGGCTGTACCTACGCGGTGGGCGCCGCTTGCACGCCCCGGGCGCCGCGGGAGCTGCTGGACGTGGGC
CGCGATGGGCGGCTGGCAGGACGTCGGCGCGTCTCGGGCGCGGGGCGCCCGCTGCCGCTGCAAGTCCG
CTTGGTGGCCCGCAGTGCCCCGACGGCGCTGAGCCGCCGCCTGCGGGCGCGCACGCACCTTCCCGGCT
GCGGAGCCCGTGCCCGGCTCTGCGGAACCGGTGCCCGGCTCTGCGGGGCGCTCTGCTTCCCCGTCCCC
GGCGGCTGCGCGGCCGCGCAGCATTCGGCGCTCGCAGCTCCGACCACCTTACCCGCCTGCCGCTGCCC
GCCGCGCCCCAGGCCCCGCTGTCCCGGCCGTCCCATCTGCCTGCCGCCGGGCGGCTCGGTCCGCCTGC
GTCTGCTGTGCGCCCTGCGGCGCGCGGCTGGCGCCGTCCGGGTGGGACTGGCGCTGGAGGCCGCCACC
GCGGGGACGCCCTCCGCGTCGCCATCCCCATCGCCGCCCCTGCCGCCGAACTTGCCCGAAGCCCGGGC
GGGGCCGGCGCGACGGGCCCGGCGGGGCACGAGCGGCAGAGGGAGCCTGAAGTTTCCGATGCCCAACT
ACCAGGTGGCGTTGTTTGAGAACGAACCGGCGGGCACCCTCATCCTCCAGCTGCACGCGCACTACACC
ATCGAGGGCGAGGAGGAGCGCGTGAGCTATTACATGGAGGGGCTGTTCGACGAGCGCTCCCGGGGCTA
CTTCCGAATCGACTCTGCCACGGGCGCCGTGAGCACGGACAGCGTACTGGACCGCGAGACCAAGGAGA
CGCACGTCCTCAGGGTGAAAGCCGTGGACTACAGTACGCCGCCGCGCTCGGCCACCACCTACATCACT
GTCTTGGTCAAAGACACCAACGACCACAGCCCGGTCTTCGAGCAGTCGGAGTACCGCGAGCGCGTGCG
GGAGAACCTGGAGGTGGGCTACGAGGTGCTGACCATCCGCGCCAGCGACCGCGACTCGCCCATCAACG
CCAACTTGCGTTACCGCGTGTTGGGGGGCGCGTGGGACGTCTTCCAGCTCAACGAGAGCTCTGGCGTG
GTGAGCACACGGGCGGTGCTGGACCGGGAGGAGGCGGCCGAGTACCAGCTCCTGGTGGAGGCCAACGA
CCAGGGGCGCAATCCGGGCCCGCTCAGTGCCACGGCCACCGTGTACATCGAGGTGGAGGACGAGAACG
ACAACTACCCCCAGTTCAGCGAGCAGAACTACGTGGTCCAGGTGCCCGAGGACGTGGGGCTCAACACG
GCTGTGCTGCGAGTGCAGGCCACGGACCGGGACCAGGGCCAGAACGCGGCCATTCACTACAGCATCCT
CAGCGGGAACGTGGCCGGCCAGTTCTACCTGCACTCGCTGAGCGGGATCCTGGATGTGATCAACCCCT
TGGATTTCGAGGATGTCCAGAAATACTCGCTGAGCATTAAGGCCCAGGATGGGGGCCGGCCCCCGCTC
ATCAATTCTTCAGGGGTGGTGTCTGTGCAGGTGCTGGATGTCAACGACAACGAGCCTATCTTTGTGAG
CAGCCCCTTCCAGGCCACGGTGCTGGAGAATGTGCCCCTGGGCTACCCCGTGGTGCACATTCAGGCGG
TGGACGCGGACTCTGGAGAGAACGCCCGGCTGCACTATCGCCTGGTGGACACGGCCTCCACCTTTCTG
GGGGGCGGCAGCGCTGGGCCTAAGAATCCTGCCCCCACCCCTGACTTCCCCTTCCAGATCCACAACAG
CTCCGGTTGGATCACAGTGTGTGCCGAGCTGGACCGCGAGGAGGTGGAGCACTACAGCTTCGGGGTGG
AGGCGGTGGACCACGGCTCGCCCCCCATGAGCTCCTCCACCAGCGTGTCCATCACGGTGCTGGACGTG
AATGACAACGACCCGGTGTTCACGCAGCCCACCTACGAGCTTCGTCTGAATGAGGATGCGGCCGTGGG
GAGCAGCGTGCTGACCCTGCAGGCCCGCGACCGTGACGCCAACAGTGTGATTACCTACCAGCTCACAG
GCGGCAACACCCGGAACCGCTTTGCACTCAGCAGCCAGAGAGGGGGCGGCCTCATCACCCTGGCGCTA
CCTCTGGACTACAAGCAGGAGCAGCAGTACGTGCTGGCGGTGACAGCATCCGACGGCACACGGTCGCA

CACTGCGCATGTCCTAATCAACGTCACTGATGCCAACACCCACAGGCCTGTCTTTCAGAGCTCCCATT
ACACAGTGAGTGTCAGTGAGGACAGGCCTGTGGGCACCTCCATTGCTACCCTCAGTGCCAACGATGAG
GACACAGGAGAGAATGCCCGCATCACCTACGTGATTCAGGACCCCGTGCCGCAGTTCCGCATTGACCC
CGACAGTGGCACCATGTACACCATGATGGAGCTGGACTATGAGAACCAGGTCGCCTACACGCTGACCA
TCATGGCCCAGGACAACGGCATCCCGCAGAAATCAGACACCACCACCCTAGAGATCCTCATCCTCGAT
GCCAATGACAATGCACCCCAGTTCCTGTGGGATTTCTACCAGGGTTCCATCTTTGAGGATGCTCCACC
CTCGACCAGCATCCTCCAGGTCTCTGCCACGGACCGGGACTCAGGTCCCAATGGGCGTCTGCTGTACA
CCTTCCAGGGTGGGGACGACGGCGATGGGGACTTCTACATCGAGCCCACGTCCGGTGTGATTCGCACC
CAGCGCCGGCTGGACCGGGAGAATGTGGCCGTGTACAACCTTTGGGCTCTGGCTGTGGATCGGGGCAG
TCCCACTCCCCTTAGCGCCTCGGTAGAAATCCAGGTGACCATCTTGGACATTAATGACAATGCCCCCA
TGTTTGAGAAGGACGAACTGGAGCTGTTTGTTGAGGAGAACAACCCAGTGGGGTCGGTGGTGGCAAAG
ATTCGTGCTAACGACCCTGATGAAGGCCCTAATGCCCAGATCATGTATCAGATTGTGGAAGGGGACAT
GCGGCATTTCTTCCAGCTGGACCTGCTCAACGGGGACCTGCGTGCCATGGTGGAGCTGGACTTTGAGG
TCCGGCGGGAGTATGTGCTGGTGGTGCAGGCCACGTCGGCTCCGCTGGTGAGCCGAGCCACGGTGCAC
ATCCTTCTCGTGGACCAGAATGACAACCCGCCTGTGCTGCCCGACTTCCAGATCCTCTTCAACAACTA

TGTCACCAACAAGTCCAACAGTTTCCCCACCGGCGTGATCGGCTGCATCCCGGCCCATGACCCCGACG
TGTCAGACAGCCTCAACTACACCTTCGTGCAGGGCAACGAGCTGCGCCTGTTGCTGCTGGACCCCGCC
ACGGGCGAACTGCAGCTCAGCCGCGACCTGGACAACAACCGGCCGCTGGAGGCGCTCATGGAGGTGTC
TGTGTCTGCAGATGGCATCCACAGCGTCACGGCCTTCTGCACCCTGCGTGTCACCATCATCACGGACG
ACATGCTGACCAACAGCATCACTGTCCGCCTGGAGAACATGTCCCAGGAGAAGTTCCTGTCCCCGCTG
CTGGCCCTCTTCGTGGAGGGGGTGGCCGCCGTGCTGTCCACCACCAAGGACGACGTCTTCGTCTTCAA
CGTCCAGAACGACACCGACGTCAGCTCCAACATCCTGAACGTGACCTTCTCGGCGCTGCTGCCTGGCG
GCGTCCGCGGCCAGTTCTTCCCGTCGGAGGACCTGCAGGAGCAGATCTACCTGAATCGGACGCTGCTG
ACCACCATCTCCACGCAGCGCGTGCTGCCCTTCGACGACAACATCTGCCTGCGCGAGCCCTGCGAGAA
CTACATGAAGTGCGTGTCCGTTCTGCGATTCGACAGCTCCGCGCCCTTCCTCAGCTCCACCACCGTGC
TCTTCCGGCCCATCCACCCCATCAACGGCCTGCGCTGCCGCTGCCCGCCCGGCTTCACCGGCGACTAC
TGCGAGACGGAGATCGACCTCTGCTACTCCGACCCGTGCGGCGCCAACGGCCGCTGCCGCAGCCGCGA
GGGCGGCTACACCTGCGAGTGCTTCGAGGACTTCACTGGAGAGCACTGTGAGGTGGATGCCCGCTCAG
GCCGCTGTGCCAACGGGGTGTGCAAGAACGGGGGCACCTGCGTGAACCTGCTCATCGGCGGCTTCCAC
TGCGTGTGTCCTCCTGGCGAGTATGAGAGGCCCTACTGTGAGGTGACCACCAGGAGCTTCCCGCCCCA
GTCCTTCGTCACCTTCCGGGGCCTGAGACAGCGCTTCCACTTCACCATCTCCCTCACGTTTGCCACTC
AGGAAAGGAACGGCTTGCTTCTCTACAACGGCCGCTTCAATGAGAAGCACGACTTCATCGCCCTGGAG
ATCGTGGACGAGCAGGTGCAGCTCACCTTCTCTGCAGGTGCAGGCGAGACAACAACGACCGTGGCACC
GAAGGTTCCCAGTGGTGTGAGTGACGGGCGGTGGCACTCTGTGCAGGTGCAGTACTACAACAAGGTAA
GATGGGCCCCACCACTTCCCCCTGGCCCCCAGCCCAATATTGGCCACCTGGGCCTGCCCCATGGGCCG
TCCGGGGAAAAGATGGCCGTGGTGACAGTGGATGATTGTGACACAACCATGGCTGTGCGCTTTGGAAA
GGACATCGGGAACTACAGCTGCGCTGCCCAGGGCACTCAGACCGGCTCCAAGAAGTCCCTGGATCTGA
CCGGCCCTCTACTCCTGGGGGGGTGTCCCCAACCTGCCAGAAGACTTCCCAGTGCACAACCGGCAGTTC
GTGGGCTGCATGCGGAACCTGTCAGTCGACGGCAAAAATGTGGACATGGCCGGATTCATCGCCAACAA
TGGCACCCGGGAAGGCTGCGCTGCTCGGAGGAACTTCTGCGATGGGAGGCGGTGTCAGAATGGAGGCA
CCTGTGTCAACAGGTGGAATATGTATCTGTGTGAGTGTCCACTCCGATTCGGCGGGAAGAACTGTGAG
CAAGCCATGCCTCACCCCCAGCTCTTCAGCGGTGAGAGCGTCGTGTCCTGGAGTGACCTGAACATCAT
CATCTCTGTGCCCCTGGTACCTGGGGCTCATGTTCCGGACCCGGAAGGAGGACAGCGTTCTGATGGAGG
CCACCAGTGGTGGGCCCACCAGCTTTCGCCTCCAGATCCTGAACAACTACCTCCAGTTTGAGGTGTCC
CACGGCCCCTCCGATGTGGAGTCCGTGATGCTGTCCGGGTTGCGGGTGACCGACGGGGAGTGGCACCA
CCTGCTGATCGAGCTGAAGAATGTTAAGGAGGACAGTGAGATGAAGCACCTGGTCACCATGACCTTGG
ACTATGGGATGGACCAGAACAAGGCAGATATCGGGGGCATGCTTCCCGGGCTGACGGTAAGGAGCGTG
GTGGTCGGAGGCGCCTCTGAAGACAAGGTCTCCGTGCGCCGTGGATTCCGAGGCTGCATGCAGGGAGT
GAGGATGGGGGGGACGCCCACCAACGTCGCCACCCTGAACATGAACAACGCACTCAAGGTCAGGGTGA
AGGACGGCTGTGATGTGGACGACCCCTGTACCTCGAGCCCCTGTCCCCCCAATAGCCGCTGCCACGAC
GCCTGGGAGGACTACAGCTGCGTCTGTGACAAAGGGTACCTTGGAATAAACTGTGTGGATGCCTGTCA
CCTGAACCCCTGCGAGAACATGGGGGCCTGCGTGCGCTCCCCCGGCTCCCCGCAGGGCTACGTGTGCG
AGTGTGGGCCCAGTCACTACGGGCCGTACTGTGAGAACAAACTCGACCTTCCGTGCCCCAGAGGCTGG
TGGGGGAACCCCGTCTGTGGACCCTGCCACTGTGCCGTCAGCAAAGGCTTTGATCCCGACTGTAATAA
GACCAACGGCCAGTGCCAATGCAAGGAGAATTACTACAAGCTCCTAGCCCAGGACACCTGTCTGCCCT
GCGACTGCTTCCCCCATGGCTCCCACAGCCGCACTTGCGACATGGCCACCGGGCAGTGTGCCTGCAAG
CCCGGCGTCATCGGCCGCCAGTGCAACCGCTGCGACAACCCGTTTGCCGAGGTCACCACGCTCGGCTG
TGAAGTGATCTACAATGGCTGTCCCAAAGCATTTGAGGCCGGCATCTGGTGGCCACAGACCAAGTTCG
GGCAGCCGGCTGCGGTGCCATGCCCTAAGGGATCCGTTGGAAATGCGGTCCGACACTGCAGCGGGAG

AAGGGCTGGCTGCCCCCAGAGCTCTTTAACTGTACCACCATCTCCTTCGTGGACCTCAGGGCCATGAA
TGAGAAGCTGAGCCGCAATGAGACGCAGGTGGACGGCGCCAGGGCCCTGCAGCTGGTGAGGGCGCTGC
GCAGTGCTACACAGCACACGGGCACGCTCTTTGGCAATGACGTGCGCACGGCCTACCAGCTGCTGGGC
CACGTCCTTCAGCACGAGAGCTGGCAGCAGGGCTTCGACCTGGCAGCCACGCAGGACGCCGACTTTCA
CGAGGACGTCATCCACTCGGGCAGCGCCCTCCTGGCCCCAGCCACCAGGGCGGCGTGGGAGCAGATCC

AGCGGAGCGAGGGCGGCACGGCACAGCTGCTCCGGCGCCTCGAGGGCTACTTCAGCAACGTGGCACGC
AACGTGCGGCGGACGTACCTGCGGCCCTTCGTCATCGTCACCGCCAACATGGTTCTTGCTGTCGACAT
CTTTGACAAGTTCAACTTTACGGGAGCCAGGGTCCCGCGATTCGACACCATCCATGAAGAGTTCCCCA
GGGAGCTGGAGTCCTCCGTCTCCTTCCCAGCCGACTTCTTCAGACCACCTGAAGAAAAAGAAGGCCCC
CTGCTGAGGCCGGCTGGCCGGAGGACCACCCCGCAGACCACGCGCCCGGGGCCTGGCACCGAGAGGGA
GGCCCCGATCAGCAGGCGGAGGCGACACCCTGATGACGCTGGCCAGTTCGCCGTCGCTCTGGTCATCA
TTTACCGCACCCTGGGGCAGCTCCTGCCCGAGCGCTACGACCCCGACCGTCGCAGCCTCCGGTTGCCT
CACCGGCCCATCATTAATACCCCGATGGTGAGCACGCTGGTGTACAGCGAGGGGGCTCCGCTCCCGAG
ACCCCTGGAGAGGCCCGTCCTGGTGGAGTTCGCCCTGCTGGAGGTGGAGGAGCGAACCAAGCCTGTCT
GCGTGTTCTGGAACCACTCCCTGGCCGTTGGTGGGACGGGAGGGTGGTCTGCCCGGGGCTGCGAGCTC
CTGTCCAGGAACCGGACACATGTCGCCTGCCAGTGCAGCCACACAGCCAGCTTTGCGGTGCTCATGGA
TATCTCCAGGCGTGAGAACGGGGAGGTCCTGCCTCTGAAGATTGTCACCTATGCCGCTGTGTCCTTGT
CACTGGCAGCCCTGCTGGTGGCCTTCGTCCTCCTGAGCCTGGTCCGCATGCTGCGCTCCAACCTGCAC
AGCATTCACAAGCACCTCGCCGTGGCGCTCTTCCTCTCTCAGCTGGTGTTCGTGATTGGGATCAACCA
GACGGAAAACCCGTTTCTGTGCACAGTGGTTGCCATCCTCCTCCACTACATCTACATGAGCACCTTTG
CCTGGACCCTCGTGGAGAGCCTGCATGTCTACCGCATGCTGACCGAGGTGCGCAACATCGACACGGGG
CCCATGCGGTTCTACTACGTCGTGGGCTGGGGCATCCCGGCCATTGTCACAGGACTGGCGGTCGGCCT
GGACCCCCAGGGCTACGGGAACCCCGACTTCTGCTGGCTGTCGCTTCAAGACACCCTGATTTGGAGCT
TTGCGGGGCCCATCGGAGCTGTTATAATCATCAACACAGTCACTTCTGTCCTATCTGCAAAGGTTTCC
TGCCAAAGAAAGCACCATTATTATGGGAAAAAAGGGATCGTCTCCCTGCTGAGGACCGCATTCCTCCT
GCTGCTGCTCATCAGCGCCACCTGGCTGCTGGGGCTGCTGGCTGTGAACCGCGATGCACTGAGCTTTC
ACTACCTCTTCGCCATCTTCAGCGGCTTACAGGGCCCCTTCGTCCTCCTTTTCCACTGCGTGCTCAAC
CAGGAGGTCCGGAAGCACCTGAAGGGCGTGCTCGGCGGGAGGAAGCTGCACCTGGAGGACTCCGCCAC
CACCAGGGCCACCCTGCTGACGCGCTCCCTCAACTGCAACACCACCTTCGGTGACGGGCCTGACATGC
TGCGCACAGACTTGGGCGAGTCCACCGCCTCGCTGGACAGCATCGTCAGGGATGAAGGGATCCAGAAG
CTCGGCGTGTCCTCTGGGCTGGTGAGGGGCAGCCACGGAGAGCCAGACGCGTCCCTCATGCCCAGGAG
CTGCAAGGATCCCCCTGGCCACGATTCCGACTCAGATAGCGAGCTGTCCCTGGATGAGCAGAGCAGCT
CTTACGCCTCCTCACACTCGTCAGACAGCGAGGACGATGGGGTGGGAGCTGAGGAAAAATGGGACCCG
GCCAGGGGCGCCGTCCACAGCACACCCAAAGGGGACGCTGTGGCCAACCACGTTCCGGCCGGCTGGCC
CGACCAGAGCCTGGCTGAGAGTGACAGTGAGGACCCCAGCGGCAAGCCCCGCCTGAAGGTGGAGACCA
AGGTCAGCGTGGAGCTGCACCGCGAGGAGCAGGGCAGTCACCGTGGAGAGTACCCCCCGGACCAGGAG
AGCGGGGGCGCAGCCAGGCTTGCTAGCAGCCAGCCCCCAGAGCAGAGGAGCATCTTGAAAAATAAAGT
CACCTACCCGCCGCCGCTGACGCTGACGGAGCAGACGCTGAAGGGCCGGCTCCGGGAGAAGCTGGCCG
ACTGTGAGCAGAGCCCCACATCCTCGCGCACGTCTTCCCTGGGCTCTGGCGGCCCCGACTGCGCCATC
ACAGTCAAGAGCCCTGGGAGGGAGCCGGGGCGTGACCACCTCAACGGGGTGGCCATGAATGTGCGCAC
TGGGAGCGCCCAGGCCGATGGCTCCGACTCTGAGAAACCGTGA

NOV21f, SNP13382484 of CG51965-01, Protein Sequence	SEQ ID NO: 286	3028 aa	MW at 330865.9kD
	SNP Pos: 2864		SNP Change: Thr to Thr

MAPPPPPVLPVLLLLAAAAALPAMGLRAAAWEPRVPGGTRAFALRPGCTYAVGAACTPRAPRELLDVG
RDGRLAGRRRVSGAGRPLPLQVRLVARSAPTALSRRLRARTHLPGCGARARLCGTGARLCGALCFPVP
GGCAAAQHSALAAPTTLPACRCPPRPRPRCPGRPICLPPGGSVRLRLLCALRRAAGAVRVGLALEAAT
AGTPSASPSPSPPLPPNLPEARAGPARRARRGTSGRGSLKFPMPNYQVALFENEPAGTLILQLHAHYT
IEGEEERVSYYMEGLFDERSRGYFRIDSATGAVSTDSVLDRETKETHVLRVKAVDYSTPPRSATTYIT
VLVKDTNDHSPVFEQSEYRERVRENLEVGYEVLTIRASDRDSPINANLRYRVLGGAWDVFQLNESSGV
VSTRAVLDREEAAEYQLLVEANDQGRNPGPLSATATVYIEVEDENDNYPQFSEQNYVVQVPEDVGLNT
AVLRVQATDRDQGQNAAIHYSILSGNVAGQFYLHSLSGILDVINPLDFEDVQKYSLSIKAQDGGRPPL
INSSGVVSVQVLDVNDNEPIFVSSPFQATVLENVPLGYPVVHIQAVDADSGENARLHYRLVDTASTFL
GGGSAGPKNPAPTPDFPFQIHNSSGWITVCAELDREEVEHYSFGVEAVDHGSPPMSSSTSVSITVLDV
NDNDPVFTQPTYELRLNEDAAVGSSVLTLQARDRDANSVITYQLTGGNTRNRFALSSQRGGGLITLAL
PLDYKQEQQYVLAVTASDGTRSHTAHVLINVTDANTHRPVFQSSHYTVSVSEDRPVGTSIATLSANDE
DTGENARITYVIQDPVPQFRIDPDSGTMYTMMELDYENQVAYTLTIMAQDNGIPQKSDTTTLEILILD
ANDNAPQFLWDFYQGSIFEDAPPSTSILQVSATDRDSGPNGRLLYTFQGGDDGDGDFYIEPTSGVIRT

QRRLDRENVAVYNLWALAVDRGSPTPLSASVEIQVTILDINDNAPMFEKDELELFVEENNPVGSVVAK
IRANDPDEGPNAQIMYQIVEGDMRHFFQLDLLNGDLRAMVELDFEVRREYVLVVQATSAPLVSRATVH
ILLVDQNDNPPVLPDFQILFNNYVTNKSNSFPTGVIGCIPAHDPDVSDSLNYTFVQGNELRLLLLDPA
TGELQLSRDLDNNRPLEALMEVSVSADGIHSVTAFCTLRVTIITDDMLTNSITVRLENMSQEKFLSPL
LALFVEGVAAVLSTTKDDVFVFNVQNDTDVSSNILNVTFSALLPGGVRGQFFPSEDLQEQIYLNRTLL
TTISTQRVLPFDDNICLREPCENYMKCVSVLRFDSSAPFLSSTTVLFRPIHPINGLRCRCPPGFTGDY
CETEIDLCYSDPCGANGRCRSREGGYTCECFEDFTGEHCEVDARSGRCANGVCKNGGTCVNLLIGGFH
CVCPPGEYERPYCEVTTRSFPPQSFVTFRGLRQRFHFTISLTFATQERNGLLLYNGRFNEKHDFIALE
IVDEQVQLTFSAGAGETTTTVAPKVPSGVSDGRWHSVQVQYYNKVRWAPPLPPGPQPNIGHLGLPHGP
SGEKMAVVTVDDCDTTMAVRFGKDIGNYSCAAQGTQTGSKKSLDLTGPLLLGGVPNLPEDFPVHNRQF
VGCMRNLSVDGKNVDMAGFIANNGTREGCAARRNFCDGRRCQNGGTCVNRWNMYLCECPLRFGGKNCE
QAMPHPQLFSGESVVSWSDLNIIISVPWYLGLMFRTRKEDSVLMEATSGGPTSFRLQILNNYLQFEVS
HGPSDVESVMLSGLRVTDGEWHHLLIELKNVKEDSEMKHLVTMTLDYGMDQNKADIGGMLPGLTVRSV
VVGGASEDKVSVRRGFRGCMQGVRMGGTPTNVATLNMNNALKVRVKDGCDVDDPCTSSPCPPNSRCHD
AWEDYSCVCDKGYLGINCVDACHLNPCENMGACVRSPGSPQGYVCECGPSHYGPYCENKLDLPCPRGW
WGNPVCGPCHCAVSKGFDPDCNKTNGQCQCKENYYKLLAQDTCLPCDCFPHGSHSRTCDMATGQCACK
PGVIGRQCNRCDNPFAEVTTLGCEVIYNGCPKAFEAGIWWPQTKFGQPAAVPCPKGSVGNAVRHCSGE
KGWLPPELFNCTTISFVDLRAMNEKLSRNETQVDGARALQLVRALRSATQHTGTLFGNDVRTAYQLLG
HVLQHESWQQGFDLAATQDADFHEDVIHSGSALLAPATRAAWEQIQRSEGGTAQLLRRLEGYFSNVAR
NVRRTYLRPFVIVTANMVLAVDIFDKFNFTGARVPRFDTIHEEFPRELESSVSFPADFFRPPEEKEGP
LLRPAGRRTTPQTTRPGPGTEREAPISRRRRHPDDAGQFAVALVIIYRTLGQLLPERYDPDRRSLRLP
HRPIINTPMVSTLVYSEGAPLPRPLERPVLVEFALLEVEERTKPVCVFWNHSLAVGGTGGWSARGCEL
LSRNRTHVACQCSHTASFAVLMDISRRENGEVLPLKIVTYAAVSLSLAALLVAFVLLSLVRMLRSNLH
SIHKHLAVALFLSQLVFVIGINQTENPFLCTVVAILLHYIYMSTFAWTLVESLHVYRMLTEVRNIDTG
PMRFYYVVGWGIPAIVTGLAVGLDPQGYGNPDFCWLSLQDTLIWSFAGPIGAVIIINTVTSVLSAKVS
CQRKHHYYGKKGIVSLLRTAFLLLLLISATWLLGLLAVNRDALSFHYLFAIFSGLQGPFVLLFHCVLN
QEVRKHLKGVLGGRKLHLEDSATTRATLLTRSLNCNTTFGDGPDMLRTDLGESTASLDSIVRDEGIQK
LGVSSGLVRGSHGEPDASLMPRSCKDPPGHDSDSDSELSLDEQSSSYASSHSSDSEDDGVGAEEKWDP
ARGAVHSTPKGDAVANHVPAGWPDQSLAESDSEDPSGKPRLKVETKVSVELHREEQGSHRGEYPPDQE
SGGAARLASSQPPEQRSILKNKVTYPPPLTLTEQTLKGRLREKLADCEQSPTSSRTSSLGSGGPDCAI
TVKSPGREPGRDHLNGVAMNVRTGSAQADGSDSEKP

NOV21g, SNP13382485 of CG51965-01, DNA Sequence	SEQ ID NO: 287	9087 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 9085
	SNP Pos: 8752	SNP Change: G to C

ATGGCGCCGCCGCCGCCGCCCGTGCTGCCCGTGCTGCTGCTCCTGGCCGCCGCCGCCGCCCTGCCGGC
GATGGGGCTGCGAGCGGCCGCCTGGGAGCCGCGCGTACCCGGCGGGACCCGCGCCTTCGCCCTCCGGC
CCGGCTGTACCTACGCGGTGGGCGCCGCTTGCACGCCCCGGGCGCCGCGGGAGCTGCTGGACGTGGGC
CGCGATGGGCGGCTGGCAGGACGTCGGCGCGTCTCGGGCGCGGGGCGCCCGCTGCCGCTGCAAGTCCG
CTTGGTGGCCCGCAGTGCCCCGACGGCGCTGAGCCGCCGCCTGCGGGCGCGCACGCACCTTCCCGGCT
GCGGAGCCCGTGCCCGGCTCTGCGGAACCGGTGCCCGGCTCTGCGGGGCGCTCTGCTTCCCCGTCCCC
GGCGGCTGCGCGGCCGCGCAGCATTCGGCGCTCGCAGCTCCGACCACCTTACCCGCCTGCCGCTGCCC
GCCGCGCCCCAGGCCCCGCTGTCCCGGCCGTCCCATCTGCCTGCCGCCGGGCGGCTCGGTCCGCCTGC
GTCTGCTGTGCGCCCTGCGGCGCGCGGCTGGCGCCGTCCGGGTGGGACTGGCGCTGGAGGCCGCCACC
GCGGGGACGCCCTCCGCGTCGCCATCCCCATCGCCGCCCCTGCCGCCGAACTTGCCCGAAGCCCGGGC
GGGGCCGGCGCGACGGGCCCGGCGGGGCACGAGCGGCAGAGGGAGCCTGAAGTTTCCGATGCCCAACT
ACCAGGTGGCGTTGTTTGAGAACGAACCGGCGGGCACCCTCATCCTCCAGCTGCACGCGCACTACACC
ATCGAGGGCGAGGAGGAGCGCGTGAGCTATTACATGGAGGGGCTGTTCGACGAGCGCTCCCGGGGCTA
CTTCCGAATCGACTCTGCCACGGGCGCCGTGAGCACGGACAGCGTACTGGACCGCGAGACCAAGGAGA
CGCACGTCCTCAGGGTGAAAGCCGTGGACTACAGTACGCCGCCGCGCTCGGCCACCACCTACATCACT
GTCTTGGTCAAAGACACCAACGACCACAGCCCGGTCTTCGAGCAGTCGGAGTACCGCGAGCGCGTGCG
GGAGAACCTGGAGGTGGGCTACGAGGTGCTGACCATCCGCGCCAGCGACCGCGACTCGCCCATCAACG
CCAACTTGCGTTACCGCGTGTTGGGGGGCGCGTGGGACGTCTTCCAGCTCAACGAGAGCTCTGGCGTG
GTGAGCACACGGGCGGTGCTGGACCGGGAGGAGGCGGCCGAGTACCAGCTCCTGGTGGAGGCCAACGA
CCAGGGGCGCAATCCGGGCCCGCTCAGTGCCACGGCCACCGTGTACATCGAGGTGGAGGACGAGAACG
ACAACTACCCCCAGTTCAGCGAGCAGAACTACGTGGTCCAGGTGCCCGAGGACGTGGGGCTCAACACG
GCTGTGCTGCGAGTGCAGGCCACGGACCGGGACCAGGGCCAGAACGCGGCCATTCACTACAGCATCCT
CAGCGGGAACGTGGCCGGCCAGTTCTACCTGCACTCGCTGAGCGGGATCCTGGATGTGATCAACCCCT

TGGATTTCGAGGATGTCCAGAAATACTCGCTGAGCATTAAGGCCCAGGATGGGGGCCGGCCCCCGCTC
ATCAATTCTTCAGGGGTGGTGTCTGTGCAGGTGCTGGATGTCAACGACAACGAGCCTATCTTTGTGAG
CAGCCCCTTCCAGGCCACGGTGCTGGAGAATGTGCCCCTGGGCTACCCCGTGGTGCACATTCAGGCGG
TGGACGCGGACTCTGGAGAGAACGCCCGGCTGCACTATCGCCTGGTGGACACGGCCTCCACCTTTCTG
GGGGGCGGCAGCGCTGGGCCTAAGAATCCTGCCCCCACCCCTGACTTCCCCTTCCAGATCCACAACAG
CTCCGGTTGGATCACAGTGTGTGCCGAGCTGGACCGCGAGGAGGTGGAGCACTACAGCTTCGGGGTGG
AGGCGGTGGACCACGGCTCGCCCCCCATGAGCTCCTCCACCAGCGTGTCCATCACGGTGCTGGACGTG
AATGACAACGACCCGGTGTTCACGCAGCCCACCTACGAGCTTCGTCTGAATGAGGATGCGGCCGTGGG
GAGCAGCGTGCTGACCCTGCAGGCCCGCGACCGTGACGCCAACAGTGTGATTACCTACCAGCTCACAG
GCGGCAACACCCGGAACCGCTTTGCACTCAGCAGCCAGAGAGGGGGCGGCCTCATCACCCTGGCGCTA
CCTCTGGACTACAAGCAGGAGCAGCAGTACGTGCTGGCGGTGACAGCATCCGACGGCACACGGTCGCA
CACTGCGCATGTCCTAATCAACGTCACTGATGCCAACACCCACAGGCCTGTCTTTCAGAGCTCCCATT
ACACAGTGAGTGTCAGTGAGGACAGGCCTGTGGGCACCTCCATTGCTACCCTCAGTGCCAACGATGAG
GACACAGGAGAGAATGCCCGCATCACCTACGTGATTCAGGACCCCGTGCCGCAGTTCCGCATTGACCC
CGACAGTGGCACCATGTACACCATGATGGAGCTGGACTATGAGAACCAGGTCGCCTACACGCTGACCA
TCATGGCCCAGGACAACGGCATCCCGCAGAAATCAGACACCACCACCCTAGAGATCCTCATCCTCGAT
GCCAATGACAATGCACCCCAGTTCCTGTGGGATTTCTACCAGGGTTCCATCTTTGAGGATGCTCCACC
CTCGACCAGCATCCTCCAGGTCTCTGCCACGGACCGGGACTCAGGTCCCAATGGGCGTCTGCTGTACA
CCTTCCAGGGTGGGGACGACGGCGATGGGGACTTCTACATCGAGCCCACGTCCGGTGTGATTCGCACC
CAGCGCCGGCTGGACCGGGAGAATGTGGCCGTGTACAACCTTTGGGCTCTGGCTGTGGATCGGGGCAG
TCCCACTCCCCTTAGCGCCTCGGTAGAAATCCAGGTGACCATCTTGGACATTAATGACAATGCCCCCA
TGTTTGAGAAGGACGAACTGGAGCTGTTTGTTGAGGAGAACAACCCAGTGGGGTCGGTGGTGGCAAAG
ATTCGTGCTAACGACCCTGATGAAGGCCCTAATGCCCAGATCATGTATCAGATTGTGGAAGGGGACAT
GCGGCATTTCTTCCAGCTGGACCTGCTCAACGGGGACCTGCGTGCCATGGTGGAGCTGGACTTTGAGG
TCCGGCGGGAGTATGTGCTGGTGGTGCAGGCCACGTCGGCTCCGCTGGTGAGCCGAGCCACGGTGCAC
ATCCTTCTCGTGGACCAGAATGACAACCCGCCTGTGCTGCCCGACTTCCAGATCCTCTTCAACAACTA

TGTCACCAACAAGTCCAACAGTTTCCCCACCGGCGTGATCGGCTGCATCCCGGCCCATGACCCCGACG
TGTCAGACAGCCTCAACTACACCTTCGTGCAGGGCAACGAGCTGCGCCTGTTGCTGCTGGACCCCGCC
ACGGGCGAACTGCAGCTCAGCCGCGACCTGGACAACAACCGGCCGCTGGAGGCGCTCATGGAGGTGTC
TGTGTCTGCAGATGGCATCCACAGCGTCACGGCCTTCTGCACCCTGCGTGTCACCATCATCACGGACG
ACATGCTGACCAACAGCATCACTGTCCGCCTGGAGAACATGTCCCAGGAGAAGTTCCTGTCCCCGCTG
CTGGCCCTCTTCGTGGAGGGGGTGGCCGCCGTGCTGTCCACCACCAAGGACGACGTCTTCGTCTTCAA
CGTCCAGAACGACACCGACGTCAGCTCCAACATCCTGAACGTGACCTTCTCGGCGCTGCTGCCTGGCG
GCGTCCGCGGCCAGTTCTTCCCGTCGGAGGACCTGCAGGAGCAGATCTACCTGAATCGGACGCTGCTG
ACCACCATCTCCACGCAGCGCGTGCTGCCCTTCGACGACAACATCTGCCTGCGCGAGCCCTGCGAGAA
CTACATGAAGTGCGTGTCCGTTCTGCGATTCGACAGCTCCGCGCCCTTCCTCAGCTCCACCACCGTGC
TCTTCCGGCCCATCCACCCCATCAACGGCCTGCGCTGCCGCTGCCCGCCCGGCTTCACCGGCGACTAC
TGCGAGACGGAGATCGACCTCTGCTACTCCGACCCGTGCGGCGCCAACGGCCGCTGCCGCAGCCGCGA
GGGCGGCTACACCTGCGAGTGCTTCGAGGACTTCACTGGAGAGCACTGTGAGGTGGATGCCCGCTCAG
GCCGCTGTGCCAACGGGGTGTGCAAGAACGGGGGCACCTGCGTGAACCTGCTCATCGGCGGCTTCCAC
TGCGTGTGTCCTCCTGGCGAGTATGAGAGGCCCTACTGTGAGGTGACCACCAGGAGCTTCCCGCCCCA
GTCCTTCGTCACCTTCCGGGGCCTGAGACAGCGCTTCCACTTCACCATCTCCCTCACGTTTGCCACTC
AGGAAAGGAACGGCTTGCTTCTCTACAACGGCCGCTTCAATGAGAAGCACGACTTCATCGCCCTGGAG
ATCGTGGACGAGCAGGTGCAGCTCACCTTCTCTGCAGGTGCAGGCGAGACAACAACGACCGTGGCACC
GAAGGTTCCCAGTGGTGTGAGTGACGGGCGGTGGCACTCTGTGCAGGTGCAGTACTACAACAAGGTAA
GATGGGCCCCACCACTTCCCCCTGGCCCCCAGCCCAATATTGGCCACCTGGGCCTGCCCCATGGGCCG
TCCGGGGAAAAGATGGCCGTGGTGACAGTGGATGATTGTGACACAACCATGGCTGTGCGCTTTGGAAA
GGACATCGGGAACTACAGCTGCGCTGCCCAGGGCACTCAGACCGGCTCCAAGAAGTCCCTGGATCTGA
CCGGCCCTCTACTCCTGGGGGGTGTCCCCAACCTGCCAGAAGACTTCCCAGTGCACAACCGGCAGTTC
GTGGGCTGCATGCGGAACCTGTCAGTCGACGGCAAAAATGTGGACATGGCCGGATTCATCGCCAACAA
TGGCACCCGGGAAGGCTGCGCTGCTCGGAGGAACTTCTGCGATGGGAGGCGGTGTCAGAATGGAGGCA
CCTGTGTCAACAGGTGGAATATGTATCTGTGTGAGTGTCCACTCCGATTCGGCGGGAAGAACTGTGAG
CAAGCCATGCCTCACCCCCAGCTCTTCAGCGGTGAGAGCGTCGTGTCCTGGAGTGACCTGAACATCAT
CATCTCTGTGCCCTGGTACCTGGGGCTCATGTTCCGGACCCGGAAGGAGGACAGCGTTCTGATGGAGG
CCACCAGTGGTGGGCCCACCAGCTTTCGCCTCCAGATCCTGAACAACTACCTCCAGTTTGAGGTGTCC
CACGGCCCCTCCGATGTGGAGTCCGTGATGCTGTCCGGGTTGCGGGTGACCGACGGGGAGTGGCACCA
CCTGCTGATCGAGCTGAAGAATGTTAAGGAGGACAGTGAGATGAAGCACCTGGTCACCATGACCTTGG
ACTATGGGATGGACCAGAACAAGGCAGATATCGGGGGCATGCTTCCCGGGCTGACGGTAAGGAGCGTG
GTGGTCGGAGGCGCCTCTGAAGACAAGGTCTCCGTGCGCCGTGGATTCCGAGGCTGCATGCAGGGAGT

GAGGATGGGGGGGACGCCCACCAACGTCGCCACCCTGAACATGAACAACGCACTCAAGGTCAGGGTGA
AGGACGGCTGTGATGTGGACGACCCCTGTACCTCGAGCCCCTGTCCCCCCAATAGCCGCTGCCACGAC
GCCTGGGAGGACTACAGCTGCGTCTGTGACAAAGGGTACCTTGGAATAAACTGTGTGGATGCCTGTCA
CCTGAACCCCTGCGAGAACATGGGGGCCTGCGTGCGCTCCCCCGGCTCCCCGCAGGGCTACGTGTGCG
AGTGTGGGCCCAGTCACTACGGGCCGTACTGTGAGAACAAACTCGACCTTCCGTGCCCCAGAGGCTGG
TGGGGGAACCCCGTCTGTGGACCCTGCCACTGTGCCGTCAGCAAAGGCTTTGATCCCGACTGTAATAA
GACCAACGGCCAGTGCCAATGCAAGGAGAATTACTACAAGCTCCTAGCCCAGGACACCTGTCTGCCCT
GCGACTGCTTCCCCCATGGCTCCCACAGCCGCACTTGCGACATGGCCACCGGGCAGTGTGCCTGCAAG
CCCGGCGTCATCGGCCGCCAGTGCAACCGCTGCGACAACCCGTTTGCCGAGGTCACCACGCTCGGCTG
TGAAGTGATCTACAATGGCTGTCCCAAAGCATTTGAGGCCGGCATCTGGTGGCCACAGACCAAGTTCG
GGCAGCCGGCTGCGGTGCCATGCCCTAAGGGATCCGTTGGAAATGCGGTCCGACACTGCAGCGGGGAG
AAGGGCTGGCTGCCCCCAGAGCTCTTTAACTGTACCACCATCTCCTTCGTGGACCTCAGGGCCATGAA
TGAGAAGCTGAGCCGCAATGAGACGCAGGTGGACGGCGCCAGGGCCCTGCAGCTGGTGAGGGCGCTGC
GCAGTGCTACACAGCACACGGGCACGCTCTTTGGCAATGACGTGCGCACGGCCTACCAGCTGCTGGGC
CACGTCCTTCAGCACGAGAGCTGGCAGCAGGGCTTCGACCTGGCAGCCACGCAGGACGCCGACTTTCA
CGAGGACGTCATCCACTCGGGCAGCGCCCTCCTGGCCCCAGCCACCAGGGCGGCGTGGGAGCAGATCC

AGCGGAGCGAGGGCGGCACGGCACAGCTGCTCCGGCGCCTCGAGGGCTACTTCAGCAACGTGGCACGC
AACGTGCGGCGGACGTACCTGCGGCCCTTCGTCATCGTCACCGCCAACATGGTTCTTGCTGTCGACAT
CTTTGACAAGTTCAACTTTACGGGAGCCAGGGTCCCGCGATTCGACACCATCCATGAAGAGTTCCCCA
GGGAGCTGGAGTCCTCCGTCTCCTTCCCAGCCGACTTCTTCAGACCACCTGAAGAAAAAGAAGGCCCC
CTGCTGAGGCCGGCTGGCCGGAGGACCACCCCGCAGACCACGCGCCCGGGGCCTGGCACCGAGAGGGA
GGCCCCGATCAGCAGGCGGAGGCGACACCCTGATGACGCTGGCCAGTTCGCCGTCGCTCTGGTCATCA
TTTACCGCACCCTGGGGCAGCTCCTGCCCGAGCGCTACGACCCCGACCGTCGCAGCCTCCGGTTGCCT
CACCGGCCCATCATTAATACCCCGATGGTGAGCACGCTGGTGTACAGCGAGGGGGCTCCGCTCCCGAG
ACCCCTGGAGAGGCCCGTCCTGGTGGAGTTCGCCCTGCTGGAGGTGGAGGAGCGAACCAAGCCTGTCT
GCGTGTTCTGGAACCACTCCCTGGCCGTTGGTGGGACGGGAGGGTGGTCTGCCCGGGGCTGCGAGCTC
CTGTCCAGGAACCGGACACATGTCGCCTGCCAGTGCAGCCACACAGCCAGCTTTGCGGTGCTCATGGA
TATCTCCAGGCGTGAGAACGGGGAGGTCCTGCCTCTGAAGATTGTCACCTATGCCGCTGTGTCCTTGT
CACTGGCAGCCCTGCTGGTGGCCTTCGTCCTCCTGAGCCTGGTCCGCATGCTGCGCTCCAACCTGCAC
AGCATTCACAAGCACCTCGCCGTGGCGCTCTTCCTCTCTCAGCTGGTGTTCGTGATTGGGATCAACCA
GACGGAAAACCCGTTTCTGTGCACAGTGGTTGCCATCCTCCTCCACTACATCTACATGAGCACCTTTG
CCTGGACCCTCGTGGAGAGCCTGCATGTCTACCGCATGCTGACCGAGGTGCGCAACATCGACACGGGG
CCCATGCGGTTCTACTACGTCGTGGGCTGGGGCATCCCGGCCATTGTCACAGGACTGGCGGTCGGCCT
GGACCCCCAGGGCTACGGGAACCCCGACTTCTGCTGGCTGTCGCTTCAAGACACCCTGATTTGGAGCT
TTGCGGGGCCCATCGGAGCTGTTATAATCATCAACACAGTCACTTCTGTCCTATCTGCAAAGGTTTCC
TGCCAAAGAAAGCACCATTATTATGGGAAAAAAGGGATCGTCTCCCTGCTGAGGACCGCATTCCTCCT
GCTGCTGCTCATCAGCGCCACCTGGCTGCTGGGGCTGCTGGCTGTGAACCGCGATGCACTGAGCTTTC
ACTACCTCTTCGCCATCTTCAGCGGCTTACAGGGCCCCTTCGTCCTCCTTTTCCACTGCGTGCTCAAC
CAGGAGGTCCGGAAGCACCTGAAGGGCGTGCTCGGCGGGAGGAAGCTGCACCTGGAGGACTCCGCCAC
CACCAGGGCCACCCTGCTGACGCGCTCCCTCAACTGCAACACCACCTTCGGTGACGGGCCTGACATGC
TGCGCACAGACTTGGGCGAGTCCACCGCCTCGCTGGACAGCATCGTCAGGGATGAAGGGATCCAGAAG
CTCGGCGTGTCCTCTGGGCTGGTGAGGGGCAGCCACGGAGAGCCAGACGCGTCCCTCATGCCCAGGAG
CTGCAAGGATCCCCCTGGCCACGATTCCGACTCAGATAGCGAGCTGTCCCTGGATGAGCAGAGCAGCT
CTTACGCCTCCTCACACTCGTCAGACAGCGAGGACGATGGGGTGGGAGCTGAGGAAAAATGGGACCCG
GCCAGGGGCGCCGTCCACAGCACCCCCAAAGGGGACGCTGTGGCCAACCACGTTCCGGCCGGCTGGCC
CGACCAGAGCCTGGCTGAGAGTGACAGTGAGGACCCCAGCGGCAAGCCCCGCCTGAAGGTGGAGACCA
AGGTCAGCGTGGAGCTGCACCGCGAGGAGCAGGGCAGTCACCGTGGACAGTACCCCCCGGACCAGGAG
AGCGGGGGCGCAGCCAGGCTTGCTAGCAGCCAGCCCCCAGAGCAGAGGAGCATCTTGAAAAATAAAGT
CACCTACCCGCCGCCGCTGACGCTGACGGAGCAGACGCTGAAGGGCCGGCTCCGGGAGAAGCTGGCCG
ACTGTGAGCAGAGCCCCACATCCTCGCGCACGTCTTCCCTGGGCTCTGGCGGCCCCGACTGCGCCATC
ACAGTCAAGAGCCCTGGGAGGGAGCCGGGGCGTGACCACCTCAACGGGGTGGCCATGAATGTGCGCAC
TGGGAGCGCCCAGGCCGATGGCTCCGACTCTGAGAAACCGTGA

NOV21g, SNP13382485 of CG51965-01, Protein Sequence	SEQ ID NO: 288	3028 aa	MW at 330864.9kD
	SNP Pos: 2918		SNP Change: Glu to Gln

MAPPPPPVLPVLLLLAAAAALPAMGLRAAAWEPRVPGGTRAFALRPGCTYAVGAACTPRAPRELLDVG
RDGRLAGRRRVSGAGRPLPLQVRLVARSAPTALSRRLRARTHLPGCGARARLCGTGARLCGALCFPVP
GGCAAAQHSALAAPTTLPACRCPPRPRPRCPGRPICLPPGGSVRLRLLCALRRAAGAVRVGLALEAAT

AGTPSASPSPSPPLPPNLPEARAGPARRARRGTSGRGSLKFPMPNYQVALFENEPAGTLILQLHAHYT
IEGEEERVSYYMEGLFDERSRGYFRIDSATGAVSTDSVLDRETKETHVLRVKAVDYSTPPRSATTYIT
VLVKDTNDHSPVFEQSEYRERVRENLEVGYEVLTIRASDRDSPINANLRYRVLGGAWDVFQLNESSGV
VSTRAVLDREEAAEYQLLVEANDQGRNPGPLSATATVYIEVEDENDNYPQFSEQNYVVQVPEDVGLNT
AVLRVQATDRDQGQNAAIHYSILSGNVAGQFYLHSLSGILDVINPLDFEDVQKYSLSIKAQDGGRPPL
INSSGVVSVQVLDVNDNEPIFVSSPFQATVLENVPLGYPVVHIQAVDADSGENARLHYRLVDTASTFL
GGGSAGPKNPAPTPDFPFQIHNSSGWITVCAELDREEVEHYSFGVEAVDHGSPPMSSSTSVSITVLDV
NDNDPVFTQPTYELRLNEDAAVGSSVLTLQARDRDANSVITYQLTGGNTRNRFALSSQRGGGLITLAL
PLDYKQEQQYVLAVTASDGTRSHTAHVLINVTDANTHRPVFQSSHYTVSVSEDRPVGTSIATLSANDE
DTGENARITYVIQDPVPQFRIDPDSGTMYTMMELDYENQVAYTLTIMAQDNGIPQKSDTTTLEILILD
ANDNAPQFLWDFYQGSIFEDAPPSTSILQVSATDRDSGPNGRLLYTFQGGDDGDGDFYIEPTSGVIRT
QRRLDRENVAVYNLWALAVDRGSPTPLSASVEIQVTILDINDNAPMFEKDELELFVEENNPVGSVVAK
IRANDPDEGPNAQIMYQIVEGDMRHFFQLDLLNGDLRAMVELDFEVRREYVLVVQATSAPLVSRATVH
ILLVDQNDNPPVLPDFQILFNNYVTNKSNSFPTGVIGCIPAHDPDVSDSLNYTFVQGNELRLLLLDPA
TGELQLSRDLDNNRPLEALMEVSVSADGIHSVTAFCTLRVTIITDDMLTNSITVRLENMSQEKFLSPL
LALFVEGVAAVLSTTKDDVFVFNVQNDTDVSSNILNVTFSALLPGGVRGQFFPSEDLQEQIYLNRTLL
TTISTQRVLPFDDNICLREPCENYMKCVSVLRFDSSAPFLSSTTVLFRPIHPINGLRCRCPPGFTGDY
CETEIDLCYSDPCGANGRCRSREGGYTCECFEDFTGEHCEVDARSGRCANGVCKNGGTCVNLLIGGFH
CVCPPGEYERPYCEVTTRSFPPQSFVTFRGLRQRFHFTISLTFATQERNGLLLYNGRFNEKHDFIALE
IVDEQVQLTFSAGAGETTTTVAPKVPSGVSDGRWHSVQVQYYNKVRWAPPLPPGPQPNIGHLGLPHGP
SGEKMAVVTVDDCDTTMAVRFGKDIGNYSCAAQGTQTGSKKSLDLTGPLLLGGVPNLPEDFPVHNRQF
VGCMRNLSVDGKNVDMAGFIANNGTREGCAARRNFCDGRRCQNGGTCVNRWNMYLCECPLRFGGKNCE
QAMPHPQLFSGESVVSWSDLNIIISVPWYLGLMFRTRKEDSVLMEATSGGPTSFRLQILNNYLQFEVS
HGPSDVESVMLSGLRVTDGEWHHLLIELKNVKEDSEMKHLVTMTLDYGMDQNKADIGGMLPGLTVRSV
VVGGASEDKVSVRRGFRGCMQGVRMGGTPTNVATLNMNNALKVRVKDGCDVDDPCTSSPCPPNSRCHD
AWEDYSCVCDKGYLGINCVDACHLNPCENMGACVRSPGSPQGYVCECGPSHYGPYCENKLDLPCPRGW
WGNPVCGPCHCAVSKGFDPDCNKTNGQCQCKENYYKLLAQDTCLPCDCFPHGSHSRTCDMATGQCACK
PGVIGRQCNRCDNPFAEVTTLGCEVIYNGCPKAFEAGIWWPQTKFGQPAAVPCPKGSVGNAVRHCSGE
KGWLPPELFNCTTISFVDLRAMNEKLSRNETQVDGARALQLVRALRSATQHTGTLFGNDVRTAYQLLG
HVLQHESWQQGFDLAATQDADFHEDVIHSGSALLAPATRAAWEQIQRSEGGTAQLLRRLEGYFSNVAR
NVRRTYLRPFVIVTANMVLAVDIFDKFNFTGARVPRFDTIHEEFPRELESSVSFPADFFRPPEEKEGP
LLRPAGRRTTPQTTRPGPGTEREAPISRRRRHPDDAGQFAVALVIIYRTLGQLLPERYDPDRRSLRLP
HRPIINTPMVSTLVYSEGAPLPRPLERPVLVEFALLEVEERTKPVCVFWNHSLAVGGTGGWSARGCEL
LSRNRTHVACQCSHTASFAVLMDISRRENGEVLPLKIVTYAAVSLSLAALLVAFVLLSLVRMLRSNLH
SIHKHLAVALFLSQLVFVIGINQTENPFLCTVVAILLHYIYMSTFAWTLVESLHVYRMLTEVRNIDTG
PMRFYYVVGWGIPAIVTGLAVGLDPQGYGNPDFCWLSLQDTLIWSFAGPIGAVIIINTVTSVLSAKVS
CQRKHHYYGKKGIVSLLRTAFLLLLLISATWLLGLLAVNRDALSFHYLFAIFSGLQGPFVLLFHCVLN
QEVRKHLKGVLGGRKLHLEDSATTRATLLTRSLNCNTTFGDGPDMLRTDLGESTASLDSIVRDEGIQK
LGVSSGLVRGSHGEPDASLMPRSCKDPPGHDSDSDSELSLDEQSSSYASSHSSDSEDDGVGAEEKWDP
ARGAVHSTPKGDAVANHVPAGWPDQSLAESDSEDPSGKPRLKVETKVSVELHREEQGSHRGQYPPDQE
SGGAARLASSQPPEQRSILKNKVTYPPPLTLTEQTLKGRLREKLADCEQSPTSSRTSSLGSGGPDCAI
TVKSPGREPGRDHLNGVAMNVRTGSAQADGSDSEKP

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 21B.

Table 21B. Comparison of the NOV21 protein sequences.

```
NOV21a  MAPPPPPVLPVLLLLAAAAALPAMGLRAAAWEPRVPGGTRAFALRPGCTYAVGAACTPRA
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  PRELLDVGRDGRLAGRRRVSGAGRPLPLQVRLVARSAPTALSRRLRARTHLPGCGARARL
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------
```

```
NOV21a  CGTGARLCGALCFPVPGGCAAAQHSALAAPTTLPACRCPPRPRPRCPGRPICLPPGGSVR
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  LRLLCALRRAAGAVRVGLALEAATAGTPSASPSPSPPLPPNLPEARAGPARRARRGTSGR
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  GSLKFPMPNYQVALFENEPAGTLILQLHAHYTIEGEEERVSYYMEGLFDERSRGYFRIDS
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  -------KLYQVALFENEPAGTLILQLHAHHTIEGEEERVSYYMEGLFDERSRGYFRIDS

NOV21a  ATGAVSTDSVLDRETKETHVLRVKAVDYSTPPRSATTYITVLVKDTNDHSPVFEQSEYRE
NOV21b  ----------------------------------LKLEVIYNGCPKAFEAGIWWPQTK
NOV21c  ------------------------------------------------------------
NOV21d  AAGAVSTDSVLDRETKETHVLRVKAVDYSTPPRSATTYITVLVKDTNDHSPVFEQSEYRE

NOV21a  RVRENLEVGYEVLTIRASDRDSPINANLRYRVLGGAWDVFQLNESSGVVSTRAVLDREEA
NOV21b  FGQPAAVPCPKGSVGNAVRHCSGEKGWLPPELFNCTTISFVDLRAMNEKLSRNETQVDGA
NOV21c  --RK-LVLPLKIVTYAAVS-LS-LAALLVAFVLLSLVRMLRSNLHS-IHKHLAVALF--L
NOV21d  RVRENLEVGYEVLTIRASDRDSPINANLRYRVLGGAWDVFQLNESSGVVSTRAVLDREEA

NOV21a  AEYQLLVEANDQGRNPGPLSATATVYIEVEDENDNYPQFSEQNYVVQVPEDVGLNTAVLR
NOV21b  RALQLVRALRSATQHTGTLFGNDVRTAYQLLGHVLQHESWQQGFDLAATQDADFHEDVIH
NOV21c  SQLVFVIGIN-QTENP-FLCTVVAILLHYIYMS----TFAWT--LVES-------LHVYR
NOV21d  AEYQLLVEANDQGRNPGPLSATATVYIEVEDENDNYPQFSEQNYVVQVPEDVGLNTAVLR

NOV21a  VQATDRDQGQNAAIHYSILSGNVAGQFYLHSLSGILDVINPLDFEDVQKYSLSIKAQDGG
NOV21b  SGSALLAPATRAAWEQIQRSEG-GTAQLLRRLEGYFSNVARNVRRTYLRPFVIVTANMIL
NOV21c  MLTEVRN----------IDTG--PMRFYYVVGWGIPAIVT----G----LAVGLDPQGYG
NOV21d  VQATDRDQGQNAAIHYSILSGNVAGQFYLHSLSGILDVINPLDFEDVQKYSLSIKAQDGG

NOV21a  RPPLINSSGVVSVQVLDVNDNEPIFVSSPFQATVLENVPLGYPVVHIQAVDADSGENARL
NOV21b  AVDIFDKFNFTGARVPRFDTIHEEFPRELESSVSFPADFFRPPEEKEGPLLRPAGRRTTP
NOV21c  NPDFC----WLSLQDTLI----WSFAG-PIGAVIIINT--------VTSVLSAKVSCQRK
NOV21d  RPPLINSSGVVSVQVLDVNDNEPIFVSSPFQATVLENVPLGYPVVHIQAVDADSGENARL

NOV21a  HYRLVDTASTFLGGGSAGPKNPAPTPDFPFQIHNSSGWITVCAELDRE-EVEHYSFGVEA
NOV21b  QTTRPGPGTEREAPISRRRRHPDDAGQFAVALVIIYRTLGQLLPERYD-PDRRSLRLPHR
NOV21c  HH---------YYG--KKGIVSLLRT-AFLLLLLISATWLLGLLAVNRDALSFHYLFAIFS
NOV21d  HYRLVDTASTFLGGGSAGPKNPAPTPDFPFQIRNSSGWITVCAELDREEVE-HYSFGVEA

NOV21a  VDHGSPPMSSSTSVSITVLDVNDNDPVFTQPTYELRLNEDAAVGSSVLTLQARDRDANSV
NOV21b  PIINTPMVSTLVYSEGAPLPRPLERPVLVEFALLEVEERTKPVCVFWNHSLAVGGTGGWS
NOV21c  GLQG-P------FV-LLFHCV-LNQEVKLA----------------------------
NOV21d  VDHGWPPMSSSTSVSITVLDVNDNDPVFTQPTYELRLNEDAAVGSSVLTLQARDRDANSV

NOV21a  ITYQLTGGNTRNRFALSSQRGGGLITLALPLDYKQEQQYVLAVTASDGTRSHTAHVLINV
NOV21b  ARGCELLSRNRTHVACQCSHTASFAVLMDISRRENGEKL----------------------
NOV21c  ------------------------------------------------------------
```

```
NOV21d  ITYQLTGGNTRNRFALSSQRGGGLITLALPLDYKQEQQYVLAVTASDGTRSHTAHVLINV

NOV21a  TDANTHRPVFQSSHYTVSVSEDRPVGTSIATLSANDEDTGENARITYVIQDPVPQFRIDP
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  TDANTHRPVFQSSHYTVSVSEDRPVGTSIATLSANDEDTGENARITYVIQDPVPQFRIDP

NOV21a  DSGTMYTMMELDYENQVAYTLTIMAQDNGIPQKSDTTTLEILILDANDNAPQFLWDFYQG
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  DSGTMYTMMELDYENQVAYTLTIMAQDNGIPQKSDTTTLEILILDANDNAPQFLWDFYQG

NOV21a  SIFEDAPPSTSILQVSATDRDSGPNGRLLYTFQGGDDGDGDFYIEPTSGVIRTQRRLDRE
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  SIFEDAPPSTSILQVSATDRDSGPNGRLLYTFQGGDDGDGDFYIEPTSGVIRTQRRLDRE

NOV21a  NVAVYNLWALAVDRGSPTPLSASVEIQVTILDINDNAPMFEKDELELFVEENNPVGSVVA
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  NVAVYNLWALAVDRGSPTPLSASVEIQVTILDINDNAPMFEKDELELFVEENNPVGSVVA

NOV21a  KIRANDPDEGPNAQIMYQIVEGDMRHFFQLDLLNGDLRAMVELDFEVRREYVLVVQATSA
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  KIRANDPDEGPNAQIMYQIVEGDMRHFFQLDLLNGDLRAMVELDFEVRREYVLVVQATSA

NOV21a  PLVSRATVHILLVDQNDNPPVLPDFQILFNNYVTNKSNSFPTGVIGCIPAHDPDVSDSLN
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  PLVSRATVHILLVLE----------------------------------------------

NOV21a  YTFVQGNELRLLLLDPATGELQLSRDLDNNRPLEALMEVSVSADGIHSVTAFCTLRVTII
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  TDDMLTNSITVRLENMSQEKFLSPLLALFVEGVAAVLSTTKDDVFVFNVQNDTDVSSNIL
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  NVTFSALLPGGVRGQFFPSEDLQEQIYLNRTLLTTISTQRVLPFDDNICLREPCENYMKC
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  VSVLRFDSSAPFLSSTTVLFRPIHPINGLRCRCPPGFTGDYCETEIDLCYSDPCGANGRC
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  RSREGGYTCECFEDFTGEHCEVDARSGRCANGVCKNGGTCVNLLIGGFHCVCPPGEYERP
NOV21b  ------------------------------------------------------------
```

```
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  YCEVTTRSFPPQSFVTFRGLRQRFHFTISLTFATQERNGLLLYNGRFNEKHDFIALEIVD
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  EQVQLTFSAGAGETTTTVAPKVPSGVSDGRWHSVQVQYYNKVRWAPPLPPGPQPNIGHLG
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  LPHGPSGEKMAVVTVDDCDTTMAVRFGKDIGNYSCAAQGTQTGSKKSLDLTGPLLLGGVP
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  NLPEDFPVHNRQFVGCMRNLSVDGKNVDMAGFIANNGTREGCAARRNFCDGRRCQNGGTC
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  VNRWNMYLCECPLRFGGKNCEQAMPHPQLFSGESVVSWSDLNIIISVPWYLGLMFRTRKE
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  DSVLMEATSGGPTSFRLQILNNYLQFEVSHGPSDVESVMLSGLRVTDGEWHHLLIELKNV
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  KEDSEMKHLVTMTLDYGMDQNKADIGGMLPGLTVRSVVVGGASEDKVSVRRGFRGCMQGV
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  RMGGTPTNVATLNMNNALKVRVKDGCDVDDPCTSSPCPPNSRCHDAWEDYSCVCDKGYLG
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  INCVDACHLNPCENMGACVRSPGSPQGYVCECGPSHYGPYCENKLDLPCPRGWWGNPVCG
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  PCHCAVSKGFDPDCNKTNGQCQCKENYYKLLAQDTCLPCDCFPHGSHSRTCDMATGQCAC
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  KPGVIGRQCNRCDNPFAEVTTLGCEVIYNGCPKAFEAGIWWPQTKFGQPAAVPCPKGSVG
```

```
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  NAVRHCSGEKGWLPPELFNCTTISFVDLRAMNEKLSRNETQVDGARALQLVRALRSATQH
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  TGTLFGNDVRTAYQLLGHVLQHESWQQGFDLAATQDADFHEDVIHSGSALLAPATRAAWE
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  QIQRSEGGTAQLLRRLEGYFSNVARNVRRTYLRPFVIVTANMVLAVDIFDKFNFTGARVP
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  RFDTIHEEFPRELESSVSFPADFFRPPEEKEGPLLRPAGRRTTPQTTRPGPGTEREAPIS
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  RRRRHPDDAGQFAVALVIIYRTLGQLLPERYDPDRRSLRLPHRPIINTPMVSTLVYSEGA
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  PLPRPLERPVLVEFALLEVEERTKPVCVFWNHSLAVGGTGGWSARGCELLSRNRTHVACQ
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  CSHTASFAVLMDISRRENGEVLPLKIVTYAAVSLSLAALLVAFVLLSLVRMLRSNLHSIH
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  KHLAVALFLSQLVFVIGINQTENPFLCTVVAILLHYIYMSTFAWTLVESLHVYRMLTEVR
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  NIDTGPMRFYYVVGWGIPAIVTGLAVGLDPQGYGNPDFCWLSLQDTLIWSFAGPIGAVII
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  INTVTSVLSAKVSCQRKHHYYGKKGIVSLLRTAFLLLLLISATWLLGLLAVNRDALSFHY
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------
```

```
NOV21a  LFAIFSGLQGPFVLLFHCVLNQEVRKHLKGVLGGRKLHLEDSATTRATLLTRSLNCNTTF
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  GDGPDMLRTDLGESTASLDSIVRDEGIQKLGVSSGLVRGSHGEPDASLMPRSCKDPPGHD
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  SDSDSELSLDEQSSSYASSHSSDSEDDGVGAEEKWDPARGAVHSTPKGDAVANHVPAGWP
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  DQSLAESDSEDPSGKPRLKVETKVSVELHREEQGSHRGEYPPDQESGGAARLASSQPPEQ
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  RSILKNKVTYPPPLTLTEQTLKGRLREKLADCEQSPTSSRTSSLGSGGPDCAITVKSPGR
NOV21b  ------------------------------------------------------------
NOV21c  ------------------------------------------------------------
NOV21d  ------------------------------------------------------------

NOV21a  EPGRDHLNGVAMNVRTGSAQADGSDSEKP
NOV21b  ----------------------------
NOV21c  ----------------------------
NOV21d  ----------------------------

NOV21a  (SEQ ID NO:  276)
NOV21b  (SEQ ID NO:  278)
NOV21c  (SEQ ID NO:  280)
NOV21d  (SEQ ID NO:  282)
```

Further analysis of the NOV21a protein yielded the following properties shown in Table 21C.

Table 21C. Protein Sequence Properties NOV21a

SignalP analysis:	Cleavage site between residues 21 and 22

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 26;  peak value   9.94
      PSG score:   5.54

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   3.49
      possible cleavage site: between 20 and 21

>>> Seems to have a cleavable signal peptide (1 to 20)
```

```
ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 21
      Tentative number of TMS(s) for the threshold  0.5:  10
      INTEGRAL    Likelihood = -4.99   Transmembrane 1220 -1236
      INTEGRAL    Likelihood = -0.53   Transmembrane 2251 -2267
      INTEGRAL    Likelihood = -0.06   Transmembrane 2351 -2367
      INTEGRAL    Likelihood = -9.92   Transmembrane 2489 -2505
      INTEGRAL    Likelihood = -5.95   Transmembrane 2521 -2537
      INTEGRAL    Likelihood = -1.54   Transmembrane 2544 -2560
      INTEGRAL    Likelihood = -2.81   Transmembrane 2591 -2607
      INTEGRAL    Likelihood = -1.59   Transmembrane 2631 -2647
      INTEGRAL    Likelihood = -8.92   Transmembrane 2674 -2690
      INTEGRAL    Likelihood = -1.17   Transmembrane 2703 -2719
      PERIPHERAL  Likelihood =  2.81 (at 1076)
      ALOM score:  -9.92  (number of TMSs: 10)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 10
      Charge difference:  1.0   C( 2.0) - N( 1.0)
      C > N: C-terminal side will be inside

>>>Caution: Inconsistent mtop result with signal peptide
>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  1.33
      Hyd Moment(95):  2.60    G content:       1
      D/E content:     1       S/T content:     0
      Score: -5.63

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 69  PRA|PR

NUCDISC: discrimination of nuclear localization signals
      pat4: RRRR (5) at 2340
      pat4: RRRH (3) at 2341
      pat7: PARRARR (4) at  229
      pat7: PISRRRR (5) at 2337
      bipartite: none
      content of basic residues:   9.1%
      NLS Score:  0.80

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: DSEK

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none
```

```
RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 70.6

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

-------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       11.1 %: vacuolar
       11.1 %: mitochondrial
       11.1 %: cytoplasmic

>> prediction for CG51965-01 is end (k=9)
```

A search of the NOV21a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 21D.

Table 21D. Geneseq Results for NOV21a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV21a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAE08586	Human NOV7 protein - Homo sapiens, 3028 aa. [WO200161009-A2, 23-AUG-2001]	1..3028 1..3028	3028/3028 (100%) 3028/3028 (100%)	0.0
ABP81979	Human GPCR CELSR1/Flamingo protein SEQ ID NO:444 - Homo sapiens, 3014 aa. [WO200261087-A2, 08-AUG-2002]	1..3028 1..3014	3011/3029 (99%) 3012/3029 (99%)	0.0
AAU02196	Seven-pass transmembrane receptor-like protein, MEM1 - Homo sapiens, 3014 aa. [WO200144473-A2, 21-JUN-2001]	1..3028 1..3014	3011/3029 (99%) 3012/3029 (99%)	0.0
AAU68533	Human novel cytokine encoded by cDNA 790CIP2C_4 #1 - Homo sapiens, 3014 aa. [WO200175093-A1, 11-OCT-2001]	1..3028 1..3014	3011/3029 (99%) 3012/3029 (99%)	0.0
AAW27161	Mouse receptor ME2 - Mus musculus, 2707 aa. [WO9707209-A2, 27-FEB-1997]	256..2902 35..2663	2138/2647 (80%) 2345/2647 (87%)	0.0

In a BLAST search of public sequence databases, the NOV21a protein was found to have homology to the proteins shown in the BLASTP data in Table 21E.

Table 21E. Public BLASTP Results for NOV21a				
Protein Accession Number	**Protein/Organism/Length**	**NOV21a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9NYQ6	Cadherin EGF LAG seven-pass G-type receptor 1 precursor (Flamingo homolog 2) (hFmi2) - Homo sapiens (Human), 3014 aa.	1..3028 1..3014	3011/3029 (99%) 3012/3029 (99%)	0.0
O35161	Cadherin EGF LAG seven-pass G-type receptor 1 precursor - Mus musculus (Mouse), 3034 aa.	1..3028 1..3034	2453/3057 (80%) 2680/3057 (87%)	0.0
E1260972	SEQUENCE 3 FROM PATENT WO9707209 - unidentified, 2707 aa.	256..2902 35..2663	2138/2647 (80%) 2345/2647 (87%)	0.0
Q9HCU4	Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Epidermal growth factor-like 2) (Multiple epidermal growth factor-like domains 3) (Flamingo 1) - Homo sapiens (Human), 2923 aa.	223..3026 163..2915	1658/2839 (58%) 2067/2839 (72%)	0.0
Q9R0M0	Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Flamingo 1) (mFmi1) - Mus musculus (Mouse), 2920 aa.	2..3026 8..2912	1665/3072 (54%) 2095/3072 (67%)	0.0

PFam analysis indicates that the NOV21a protein contains the domains shown in the Table 21F.

Table 21F. Domain Analysis of NOV21a			
Pfam Domain	**NOV21a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
cadherin	250..344	35/110 (32%) 70/110 (64%)	1e-14
cadherin	358..450	44/107 (41%) 72/107 (67%)	1.8e-21
cadherin	464..556	38/107 (36%) 74/107 (69%)	2e-24
cadherin	570..678	39/124 (31%) 87/124 (70%)	3.8e-23
cadherin	692..780	35/107 (33%) 71/107 (66%)	7.4e-19
cadherin	794..883	42/107 (39%) 73/107 (68%)	1.6e-28
cadherin	897..990	41/108 (38%) 76/108 (70%)	2.2e-27
cadherin	1004..1092	38/107 (36%) 67/107 (63%)	3.8e-22
EGF	1308..1361	16/61 (26%) 38/61 (62%)	0.71
EGF	1368..1399	15/47 (32%) 26/47 (55%)	1.8e-05
EGF	1408..1441	19/47 (40%) 29/47 (62%)	5.3e-07
laminin_G	1471..1647	43/206 (21%) 138/206 (67%)	5.5e-17
EGF	1668..1699	17/47 (36%) 24/47 (51%)	4.7e-07
laminin_G	1734..1867	26/163 (16%) 90/163 (55%)	0.0034
EGF	1891..1922	11/47 (23%) 24/47 (51%)	7.7e-05
EGF	1926..1960	18/47 (38%) 24/47 (51%)	0.0057

laminin_EGF	1981..2015	13/60 (22%) 24/60 (40%)	0.87
laminin_EGF	2018..2063	26/59 (44%) 31/59 (53%)	8.8e-09
HRM	2067..2124	21/77 (27%) 47/77 (61%)	5.1e-17
GPS	2422..2475	26/56 (46%) 46/56 (82%)	1.6e-21
7tm_2	2480..2723	88/274 (32%) 210/274 (77%)	6.4e-90

Example 22.

The NOV22 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 22A.

Table 22A. NOV22 Sequence Analysis

NOV22a, CG51983-05 DNA Sequence	SEQ ID NO: 289	2268 bp
	ORF Start: ATG at 1	ORF Stop: TAG at 2266

ATGCTTCCCGGGTGTATATTCTTGATGATTTTACTCATTCCTCAGGTTAAAGAAAAGTTCATCCTTGG
AGTAGAGGGTCAACAACTGGTTCGTCCTAAAAAGCTTCCTCTGATACAGAAGCGAGATACTGGACACA
CCCATGATGATGACATACTGAAAACGTATGAAGAAGAATTGTTGTATGAAATAAAACTAAATAGAAAA
ACCTTAGTCCTTCATCTTCTAAGATCCAGGGAGTTCCTAGGCTCAAATTACAGTGAAACATTCTACTC
CATGAAAGGAGAAGCGTTCACCAGGCATCCTCAGATCATGGAACACTGTTACTATAAAGGAAACATCC
TAAATGAAAAGAATTCTGTTGCCAGCATCAGTACTTGTGACGGGTTGAGGGGATTCTTCAGAATAAAC
GACCAAAGATACCTCATTGAACCAGTGAAATACTCAGATGAGGGAGAACATTTGGTGTTCAAATATAA
CCTGAGGGTGCCGTATGGTGCCAATTATTCCTGTACAGAGCTTAATTTTACCAGAAAAACTGTTCCAG
GGGATAATGAATCTGAAGAAGACTCCAAAATAAAAGGCATCCATGATGAAAAGTATGTTGAATTGTTC
ATTGTTGCTGATGATACTGTGTATCGCAGAAATGGTCATCCTCACAATAAACTAAGGAACCGAATTTG
GGGAATGGTCAATTTTGTCAACATGATTTATAAAACCTTAAACATCCATGTGACGTTGGTTGGCATTG
AAATATGGACACATGAAGATAAAATAGAACTATATTCAAATATAGAAACTACCTTATTGCGTTTTTCA
TTTTGGCAAGAAAAGATCCTTAAAACACGGAAGGATTTTGATCATGTTGTATTACTCAGTGGGAAGTG
GCTCTACTCACATGTGCAAGGAATTTCTTATCCAGGGGGTATGTGCCTGCCCTATTATTCCACCAGTA
TCATTAAGGATCTTTTACCTGACACAAACATAATTGCAAACAGAATGGCACATCAACTGGGGCATAAC
CTTGGGATGCAGCATGACGAGTTCCCATGCACCTGTCCTTCAGGAAAATGCGTGATGGACAGTGATGG
AAGCATTCCTGCACTGAAATTCAGTAAATGCAGCCAAAACCAATACCACCAGTACTTGAAGGATTATA
AGCCAACATGCATGCTCAACATTCCATTTCCTTACAATTTTCATGATTTCCAATTTTGTGGAAACAAG
AAGTTGGATGAGGGTGAAGAGTGTGACTGTGGCCCTGCTCAGGAGTGTACTAATCCTTGCTGTGATGC
ACACACATGTGTACTGAAGCCAGGATTTACTTGTGCAGAAGGAGAATGCTGTGAATCTTGTCAGATAA
AAAAAGCAGGGTCCATATGCAGACCGGCGAAAGATGAATGTGATTTTCCTGAGATGTGCACTGGCCAC
TCGCCTGCCTGTCCTAAGGACCAGTTCAGGGTCAATGGATTTCCTTGCAAGAACTCAGAAGGCTACTG
TTTCATGGGGAAATGTCCAACTCGTGAGGATCAGTGCTCTGAACTATTTGATGATGAGGCAATAGAGA
GTCATGATATCTGCTACAAGATGAATACAAAAGGAAATAAATTTGGATACTGCAAAAACAAGGAAAAC
AGATTTCTTCCCTGTGAGGAGAAGGATGTCAGATGTGGAAAGATCTACTGCACTGGAGGGGAGCTTTC
CTCTCTCCTTGGAGAAGACAAGACTTATCACCTTAAGGATCCCCAGAAGAATGCTACTGTCAAATGCA
AAACTATTTTTTTATACCATGATTCTACAGACATTGGCCTGGTGGCGTCAGGAACAAAATGTGGAGAG
GGAATGGTATGCAACAATGGTGAATGTCTAAACATGGAAAAGGTCTATATCTCAACCAATTGCCCCTC
TCAGTGCAATGAAAATCCTGTAGATGGCCACGGACTCCAGTGCCACTGTGAGGAAGGACAGGCACCTG

TAGCCTGTGAAGAAACCTTACATGTTACCAGTATCACCATCTTGGTTGTTGTGCTTGTCCTGGTTATT
GTCGGTATCGGAGTTCTTATACTATTAGTTCGTTACCGAAAATGTATCAAGTTGAAGCAAGTTCAGAG
CCCACCTACAGAAACCCTGGGAGTGGAGAACAAAGGATACTTTGGTGATGAGCAGCAGATAAGGACTG
AGCCAATCCTGCCAGAAATTCATTTCCTAAATAAACCTGCAAGTAAAGATTCAAGAGGAATCGCAGAT
CCCAATCAAAGTGCCAAGTGG**TAG**

NOV22a, CG51983-05 Protein Sequence	SEQ ID NO: 290	755 aa	MW at 85868.5kD

MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDILKTYEEELLYEIKLNRK
TLVLHLLRSREFLGSNYSETFYSMKGEAFTRHPQIMEHCYYKGNILNEKNSVASISTCDGLRGFFRIN
DQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKTVPGDNESEEDSKIKGIHDEKYVELF
IVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYKTLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFS
FWQEKILKTRKDFDHVVLLSGKWLYSHVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHN
LGMQHDEFPCTCPSGKCVMDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNK
KLDEGEECDCGPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGH
SPACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGYCKNKEN
RFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDSTDIGLVASGTKCGE
GMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPVACEETLHVTSITILVVVLVLVI
VGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFGDEQQIRTEPILPEIHFLNKPASKDSRGIAD
PNQSAKW

NOV22b, CG51983-01 DNA Sequence	SEQ ID NO: 291	2431 bp
	ORF Start: ATG at 51	ORF Stop: TAG at 2385

GAACTCCTTTTCTCAAGCACTTCTGCTCTCCTCTACCAGAATCACTCAGA**ATG**CTTCCCGGGTGTATA
TTCTTGATGATTTTACTCATTCCTCAGGTTAAAGAAAAGTTCATCCTTGGAGTAGAGGGTCAACAACT
GGTTCGTCCTAAAAAGCTTCCTCTGATACAGAAGCGAGATACTGGACACACCCATGATGATGACATAA
AAACGTATGAAGAAGAATTGTTGTATGAAATAAAACTAAATAGAAAAACCTTAGTCCTTCATCTTCTA
AGATCCAGGAGGGAGTTCCTAGGCTCAAATTACAGTGAAACATTCTACTCCATGAAAGGAGAAGCGTT
CACCAGGCATCCTCAGATCATGGATCATTGTTTTTACCAAGGATCCATAGTACACGAATATGATTCAG
CTGCCAGTATCAGTACGTGTAATGGTCTAAGGGGATTCTTCAGAATAAACGACCAAAGATACCTCATT
GAACCAGTGAAATACTCAGATGAGGGAGAACATTTGGTGTTCAAATATAACCTGAGGGTGCCGTATGG
TGCCAATTATTCCTGTACAGAGCTTAATTTTACCAGAAAAACTGTTCCAGGGGATAATGAATCTGAAG
AAGACTCCAAAATAAAACAGGGCATCCATGATGAAAAGTATGTTGAATTGTTCATTGTTGCTGATGAT
ACTGTGTATCGCAGAAATGGTCATCCTCACAATAAACTAAGGAACCGAATTTGGGGAATGGTCAATTT
TGTCAACATGATTTATAAAACCTTAAACATCCATGTGACGTTGGTTGGCATTGAAATATGGACACATG
AAGATAAAATAGAACTATATTCAAATATAGAAACTACCTTATTGCGTTTTTCATTTTGGCAAGAAAAG
ATCCTTAAAACACGGAAGGATTTTGATCATGTTGTATTACTCAGTGGGAAGTGGCTCTACTCACATGT
GCAAGGAATTTCTTATCCAGGGGGTATGTGCCTGCCCTATTATTCCACCAGTATCATTAAGGATCTTT
TACCTGACACAAACATAATTGCAAACAGAATGGCACATCAACTGGGGCATAACCTTGGGATGCAGCAT
GACGAGTTCCCATGCACCTGTCCTTCAGGAAAATGCGTGATGGACAGTGATGGAAGCATTCCTGCACT
GAAATTCAGTAAATGCAGCCAAAACCAATACCACCAGTACTTGAAGGATTATAAGCCAACATGCATGC
TCAACATTCCATTTCCTTACAATTTTCATGATTTCCAATTTTGTGGAAACAAGAAGTTGGATGAGGGT
GAAGAGTGTGACTGTGGCCCTGCTCAGGAGTGTACTAATCCTTGCTGTGATGCACACACATGTGTACT
GAAGCCAGGATTTACTTGTGCAGAAGGAGAATGCTGTGAATCTTGTCAGATAAAAAAAGCAGGGTCCA
TATGCAGACCGGCGAAAGATGAATGTGATTTTCCTGAGATGTGCACTGGCCACTCGCCTGCCTGTCCT
AAGGACCAGTTCAGGGTCAATGGATTTCCTTGCAAGAACTCAGAAGGCTACTGTTTCATGGGGAAATG
TCCAACTCGTGAGGATCAGTGCTCTGAACTATTTGATGATGAGGCAATAGAGAGTCATGATATCTGCT
ACAAGATGAATACAAAAGGAAATAAATTTGGATACTGCAAAAACAAGGAAAACAGATTTCTTCCCTGT
GAGGAGAAGGATGTCAGATGTGGAAAGATCTACTGCACTGGAGGGGAGCTTTCCTCTCTCCTTGGAGA
AGACAAGACTTATCACCTTAAGGATCCCCAGAAGAATGCTACTGTCAAATGCAAAACTATTTTTTTAT
ACCATGATTCTACAGACATTGGCCTGGTGGCGTCAGGAACAAAATGTGGAGAGGGAATGGTATGCAAC
AATGGTGAATGTCTAAACATGGAAAAGGTCTATATCTCAACCAATTGCCCCTCTCAGTGCAATGAAAA
TCCTGTAGATGGCCACGGACTCCAGTGCCACTGTGAGGAAGGACAGGCACCTGTAGCCTGTGAAGAAA
CCTTACATGTTACCAGTATCACCATCTTGGTTGTTGTGCTTGTCCTGGTTATTGTCGGTATCGGAGTT
CTTATACTATTAGTTCGTTACCGAAAATGTATCAAGTTGAAGCAAGTTCAGAGCCCACCTACAGAAAC
CCTGGGAGTGGAGAACAAAGGATACTTTGGTGATGAGCAGCAGATAAGGACTGAGCCAATCCTGCCAG
AAATTCATTTCCTAAATCAGAGAACTCCAGAATCCTTGGAAAGCCTGCCCACTAGTTTTTCAAGTCCC
CACTACATCACACTGAAACCTGCAAGTAAAGATTCAAGAGGAATCGCAGATCCCAATCAAAGTGCCAA
GTGG**TAG**GTTACCCTGACAGATAGTACCTCCCTTTTTTATTTTTCAAATGC

NOV22b, CG51983-01 Protein Sequence	SEQ ID NO: 292	778 aa	MW at 88471.2kD

MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDIKTYEEELLYEIKLNRKT
LVLHLLRSRREFLGSNYSETFYSMKGEAFTRHPQIMDHCFYQGSIVHEYDSAASISTCNGLRGFFRIN
DQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKTVPGDNESEEDSKIKQGIHDEKYVEL
FIVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYKTLNIHVTLVGIEIWTHEDKIELYSNIETTLLRF
SFWQEKILKTRKDFDHVVLLSGKWLYSHVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGH
NLGMQHDEFPCTCPSGKCVMDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGN
KKLDEGEECDCGPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTG
HSPACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGYCKNKE
NRFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDSTDIGLVASGTKCG
EGMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPVACEETLHVTSITILVVVLVLV
IVGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFGDEQQIRTEPILPEIHFLNQRTPESLESLP
TSFSSPHYITLKPASKDSRGIADPNQSAKW

NOV22c, CG51983-02 DNA Sequence	SEQ ID NO: 293	2583 bp
	ORF Start: ATG at 85	ORF Stop: TGA at 2347

GATCCCTGCAGTGGAAGTGAGGAGGAAGAAAGGTGAACTCCTTTTCTCAAGCACTTCTGCTCTCCTCT
ACCAGAATCACTCAGA**ATG**CTTCCCGGGTGTATATTCTTGATGATTTTACTCATTCCTCAGGTTAAAG
AAAAGTTCATCCTTGGAGTAGAGGGTCAACAACTGGTTCGTCCTAAAAAGCTTCCTCTGATACAGAAG
CGAGATACTGGACACACCCATGATGATGACATACTGAAAACGTATGAAGAAGAATTGTTGTATGAAAT
AAAACTAAATAGAAAAACCTTAGTCCTTCATCTTCTAAGATCCAGGGAGTTCCTAGGCTCAAATTACA
GTGAAACATTCTACTCCATGAAAGGAGGAGCGTTCACCAGGCATCCTCAGATCATGGATCATTGTTTT
TACCAAGGATCCATAGTACACGAATATGATTCAGCTGCCAGTATCAGTACGTGTAATGGTCTAAGGGG
ATTCTTCAGAATAAACGACCAAAGATACCTCATTGAACCAGTGAAATACTCAGATGAGGGAGAACATT
TGGTGTTCAAATATAACCTGAGGGTGCCGTATGGTGCCAATTATTCCTGTACAGAGCTTAATTTTACC
AGAAAAACTGTTCCAGGGGATAATGAATCTGAAGAAGACTCCAAAATAAAAGGCATCCATGATGAAAA
GTATGTTGAATTGTTCATTGTTGCTGATGATACTGTGTATCGCAGAAATGGTCATCCTCACAATAAAC
TAAGGAACCGAATTTGGGGAATGGTCAATTTTGTCAACATGATTTATAAAACCTTAAACATCCATGTG
ACGTTGGTTGGCATTGAAATATGGACACATGAAGATAAAATAGAACTATATTCAAATATAGAAACTAC
CTTATTGCGTTTTTCATTTTGGCAAGAAAAGATCCTTAAAACACGGAAGGATTTTGATCATGTTGTAT
TACTCAGTGGGAAGTGGCTCTACTCACATGTGCAAGGAATTTCTTATCCAGGGGGTATGTGCCTGCCC
TATTATTCCACCAGTATCATTAAGGATCTTTTACCTGACACAAACATAATTGCAAACAGAATGGCACA
TCAACTGGGGCATAACCTTGGGATGCAGCATGACGAGTTCCCATGCACCTGTCCTTCAGGAAAATGCG
TGATGGACAGTGATGGAAGCATTCCTGCACTGAAATTCAGTAAATGCAGCCAAAACCAATACCACCAG
TACTTGAAGGATTATAAGCCAACATGCATGCTCAACATTCCATTTCCTTACAATTTTCATGATTTCCA
ATTTTGTGGAAACAAGAAGTTGGATGAGGGTGAAGAGTGTGACTGTGGCCCTGCTCAGGAGTGTACTA
ATCCTTGCTGTGATGCACACACATGTGTACTGAAGCCAGGATTTACTTGTGCAGAAGGAGAATGCTGT
GAATCTTGTCAGATAAAAAAAGCAGGGTCCATATGCAGACCGGCGAAAGATGAATGTGATTTTCCTGA
GATGTGCACTGGCCACTCGCCTGCCTGTCCTAAGGACCAGTTCAGGGTCAATGGATTTCCTTGCAAGA
ACTCAGAAGGCTACTGTTTCATGGGGAAATGTCCAACTCGTGAGGATCAGTGCTCTGAACTATTTGAT
GATGATGCAATAGAGAGTCATGATATCTGCTACAAGATGAATACAAAAGGAAATAAATTTGGATACTG
CAAAAACAAGGAAAACAGATTTCTTCCCTGTGAGGAGAAAGATGTCAGATGTGGAAAGATCTACTGCA
CTGGAGGGGAGCTTTCCTCTCTCCTTGGAGAAGACAAGACTTATCACCTTAAGGATCCCCAGAAGAAT
GCTACTGTCAAATGCAAAACTATTTTTTTATACCATGATTCTACAGACATTGGCCTGGTGGCGTCAGG
AACAAAATGTGGAGAGGGAATGGTGTGCAACAATGGTGAATGTCTAAACATGGAAAAGGTCTATATCT
CAACCAATTGCCCCTCTCAGTGCAATGAAAATCCTGTGGATGGCCACGGACTCCAGTGCCACTGTGAG
GAAGGACAGGCACCTGTAGCCTGTGAAGAAACCTTACATGTTACCAATATCACCATCTTGGTTGTTGT
GCTTGTCCTGGTTATTGTCGGTATCGGAGTTCTTATACTATTAGTTCGTTACCGAAAATGTATCAAGT
TGAAGCAAGTTCAGAGCCCACCTACAGAAACCCTGGGAGTGGAGAACAAAGGATACTTTGGTGATGAG
CAGCAGATAAGGACTGAGCCAATCCTGCCAGAAATTCATTTCCTAAATAAACCTGCAAGTAAAGATTC
AAGAGGAATCGCAGATCCCAATCAAAGTGCCAAG**TGA**GCTTGAAGTTGGATATCCAAAATGGCCGTGC
AAGCTTAGGCTGGGGATTCTGGATGCAACGTCTTTACAACCTTACCTAGATATCTGCTACTCACATTT
TTGGTAGTGTTTCAAACGTTCTTTATCCAGACAGACAATGTTTAAGAGAAACAACTTATTTCTGTTAA
TATTTACCGGTAGAATTCACACCCTCTATCATAAACATATGCTGCAGAAAAAAAAAAAAAAAAAAAA

NOV22c, CG51983-02 Protein Sequence	SEQ ID NO: 294	754 aa	MW at 85582.1kD

MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDILKTYEEELLYEIKLNRK
TLVLHLLRSREFLGSNYSETFYSMKGGAFTRHPQIMDHCFYQGSIVHEYDSAASISTCNGLRGFFRIN
DQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKTVPGDNESEEDSKIKGIHDEKYVELF
IVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYKTLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFS
FWQEKILKTRKDFDHVVLLSGKWLYSHVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHN
LGMQHDEFPCTCPSGKCVMDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNK
KLDEGEECDCGPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGH
SPACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDDAIESHDICYKMNTKGNKFGYCKNKEN
RFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDSTDIGLVASGTKCGE
GMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPVACEETLHVTNITILVVVLVLVI
VGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFGDEQQIRTEPILPEIHFLNKPASKDSRGIAD
PNQSAK

NOV22d, CG51983-03 DNA Sequence	SEQ ID NO: 295	2274 bp
	ORF Start: ATG at 4	ORF Stop: TAG at 2272

AGAATGCTTCCCGGGTGTATATTCTTGATGATTTTACTCATTCCTCAGGTTAAAGAAAAGTTCATCCT
TGGAGTAGAGGGTCAACAACTGGTTCGTCCTAAAAAGCTTCCTCTGATACAGAAGCGAGATACTGGAC
ACACCCATGATGATGACATAAAAACGTATGAAGAAGAATTGTTGTATGAAATAAAACTAAATAGAAAA
ACCTTAGTCCTTCATCTTCTAAGATCCAGGAGGGAGTTCCTAGGCTCAAATTACAGTGAAACATTCTA
CTCCATGAAAGGAGAAGCGTTCACCAGGCATCCTCAGATCATGGATCATTGTTTTTACCAAGGATCCA
TAGTACACGAATATGATTCAGCTGCCAGTATCAGTACGTGTAATGGTCTAAGGGGATTCTTCAGAATA
AACGACCAAAGATACCTCATTGAACCAGTGAAATACTCAGATGAGGGAGAACATTTGGTGTTCAAATA
TAACCTGAGGGTGCCGTATGGTGCCAATTATTCCTGTACAGAGCTTAATTTTACCAGAAAAACTGTTC
CAGGGGATAATGAATCTGAAGAAGACTCCAAAATAAAACAGGGCATCCATGATGAAAAGTATGTTGAA
TTGTTCATTGTTGCTGATGATACTGTGTATCGCAGAAATGGTCATCCTCACAATAAACTAAGGAACCG
AATTTGGGGAATGGTCAATTTTGTCAACATGATTTATAAAACCTTAAACATCCATGTGACGTTGGTTG
GCATTGAAATATGGACACATGAAGATAAAATAGAACTATATTCAAATATAGAAACTACCTTATTGCGT
TTTTCATTTTGGCAAGAAAAGATCCTTAAAACACGGAAGGATTTTGATCATGTTGTATTACTCAGTGG
GAAGTGGCTCTACTCACATGTGCAAGGAATTTCTTATCCAGGGGGTATGTGCCTGCCCTATTATTCCA
CCAGTATCATTAAGGATCTTTTACCTGACACAAACATAATTGCAAACAGAATGGCACATCAACTGGGG
CATAACCTTGGGATGCAGCATGACGAGTTCCCATGCACCTGTCCTTCAGGAAAATGCGTGATGGACAG
TGATGGAAGCATTCCTGCACTGAAATTCAGTAAATGCAGCCAAAACCAATACCACCAGTACTTGAAGG
ATTATAAGCCAACATGCATGCTCAACATTCCATTTCCTTACAATTTTCATGATTTCCAATTTTGTGGA
AACAAGAAGTTGGATGAGGGTGAAGAGTGTGACTGTGGCCCTGCTCAGGAGTGTACTAATCCTTGCTG
TGATGCACACACATGTGTACTGAAGCCAGGATTTACTTGTGCAGAAGGAGAATGCTGTGAATCTTGTC
AGATAAAAAAAGCAGGGTCCATATGCAGACCGGCGAAAGATGAATGTGATTTTCCTGAGATGTGCACT
GGCCACTCGCCTGCCTGTCCTAAGGACCAGTTCAGGGTCAATGGATTTCCTTGCAAGAACTCAGAAGG
CTACTGTTTCATGGGGAAATGTCCAACTCGTGAGGATCAGTGCTCTGAACTATTTGATGATGAGGCAA
TAGAGAGTCATGATATCTGCTACAAGATGAATACAAAAGGAAATAAATTTGGATACTGCAAAAACAAG
GAAAACAGATTTCTTCCCTGTGAGGAGAAGGATGTCAGATGTGGAAAGATCTACTGCACTGGAGGGGA
GCTTTCCTCTCTCCTTGGAGAAGACAAGACTTATCACCTTAAGGATCCCCAGAAGAATGCTACTGTCA
AATGCAAAACTATTTTTTTATACCATGATTCTACAGACATTGGCCTGGTGGCGTCAGGAACAAAATGT
GGAGAGGGAATGGTATGCAACAATGGTGAATGTCTAAACATGGAAAAGGTCTATATCTCAACCAATTG
CCCCTCTCAGTGCAATGAAAATCCTGTAGATGGCCACGGACTCCAGTGCCACTGTGAGGAAGGACAGG
CACCTGTAGCCTGTGAAGAAACCTTACATGTTACCAGTATCACCATCTTGGTTGTTGTGCTTGTCCTG
GTTATTGTCGGTATCGGAGTTCTTATACTATTAGTTCGTTACCGAAAATGTATCAAGTTGAAGCAAGT
TCAGAGCCCACCTACAGAAACCCTGGGAGTGGAGAACAAAGGATACTTTGGTGATGAGCAGCAGATAA
GGACTGAGCCAATCCTGCCAGAAATTCATTTCCTAAATAAACCTGCAAGTAAAGATTCAAGAGGAATC
GCAGATCCCAATCAAAGTGCCAAGTGG**TAG**

NOV22d, CG51983-03 Protein Sequence	SEQ ID NO: 296	756 aa	MW at 85998.5kD

MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDIKTYEEELLYEIKLNRKT
LVLHLLRSRREFLGSNYSETFYSMKGEAFTRHPQIMDHCFYQGSIVHEYDSAASISTCNGLRGFFRIN

DQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKTVPGDNESEEDSKIKQGIHDEKYVEL
FIVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYKTLNIHVTLVGIEIWTHEDKIELYSNIETTLLRF
SFWQEKILKTRKDFDHVVLLSGKWLYSHVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGH
NLGMQHDEFPCTCPSGKCVMDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGN
KKLDEGEECDCGPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTG
HSPACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGYCKNKE
NRFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDSTDIGLVASGTKCG
EGMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPVACEETLHVTSITILVVVLVLV
IVGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFGDEQQIRTEPILPEIHFLNKPASKDSRGIA
DPNQSAKW

NOV22e, CG51983-04 DNA Sequence	SEQ ID NO: 297	2283 bp
	ORF Start: ATG at 12	ORF Stop: TGA at 2274

AATCACTCAGAATGCTTCCCGGGTGTATATTCTTGATGATTTTACTCATTCCTCAGGTTAAAGAAAAG
TTCATCCTTGGAGTAGAGGGTCAACAACTGGTTCGTCCTAAAAAGCTTCCTCTGATACAGAAGCGAGA
TACTGGACACACCCATGATGATGACATACTGAAAACGTATGAAGAAGAATTGTTGTATGAAATAAAAC
TAAATAGAAAAACCTTAGTCCTTCATCTTCTAAGATCCAGGGAGTTCCTAGGCTCAAATTACAGTGAA
ACATTCTACTCCATGAAAGGAGGAGCGTTCACCAGGCATCCTCAGATCATGGATCATTGTTTTTACCA
AGGATCCATAGTACACGAATATGATTCAGCTGCCAGTATCAGTACGTGTAATGGTCTAAGGGGATTCT
TCAGAATAAACGACCAAAGATACCTCATTGAACCAGTGAAATACTCAGATGAGGGAGAACATTTGGTG
TTCAAATATAACCTGAGGGTGCCGTATGGTGCCAATTATTCCTGTACAGAGCTTAATTTTACCAGAAA
AACTGTTCCAGGGGATAATGAATCTGAAGAAGACTCCAAAATAAAAGGCATCCATGATGAAAAGTATG
TTGAATTGTTCATTGTTGCTGATGATACTGTGTATCGCAGAAATGGTCATCCTCACAATAAACTAAGG
AACCGAATTTGGGGAATGGTCAATTTTGTCAACATGATTTATAAAACCTTAAACATCCATGTGACGTT
GGTTGGCATTGAAATATGGACACATGAAGATAAAATAGAACTATATTCAAATATAGAAACTACCTTAT
TGCGTTTTTCATTTTGGCAAGAAAAGATCCTTAAAACACGGAAGGATTTTGATCATGTTGTATTACTC
AGTGGGAAGTGGCTCTACTCACATGTGCAAGGAATTTCTTATCCAGGGGGTATGTGCCTGCCCTATTA
TTCCACCAGTATCATTAAGGATCTTTTACCTGACACAAACATAATTGCAAACAGAATGGCACATCAAC
TGGGGCATAACCTTGGGATGCAGCATGACGAGTTCCCATGCACCTGTCCTTCAGGAAAATGCGTGATG
GACAGTGATGGAAGCATTCCTGCACTGAAATTCAGTAAATGCAGCCAAAACCAATACCACCAGTACTT
GAAGGATTATAAGCCAACATGCATGCTCAACATTCCATTTCCTTACAATTTTCATGATTTCCAATTTT
GTGGAAACAAGAAGTTGGATGAGGGTGAAGAGTGTGACTGTGGCCCTGCTCAGGAGTGTACTAATCCT
TGCTGTGATGCACACACATGTGTACTGAAGCCAGGATTTACTTGTGCAGAAGGAGAATGCTGTGAATC
TTGTCAGATAAAAAAAGCAGGGTCCATATGCAGACCGGCGAAAGATGAATGTGATTTTCCTGAGATGT
GCACTGGCCACTCGCCTGCCTGTCCTAAGGACCAGTTCAGGGTCAATGGATTTCCTTGCAAGAACTCA
GAAGGCTACTGTTTCATGGGGAAATGTCCAACTCGTGAGGATCAGTGCTCTGAACTATTTGATGATGA
TGCAATAGAGAGTCATGATATCTGCTACAAGATGAATACAAAAGGAAATAAATTTGGATACTGCAAAA
ACAAGGAAAACAGATTTCTTCCCTGTGAGGAGAAAGATGTCAGATGTGGAAAGATCTACTGCACTGGA
GGGGAGCTTTCCTCTCTCCTTGGAGAAGACAAGACTTATCACCTTAAGGATCCCCAGAAGAATGCTAC
TGTCAAATGCAAAACTATTTTTTTATACCATGATTCTACAGACATTGGCCTGGTGGCGTCAGGAACAA
AATGTGGAGAGGGAATGGTGTGCAACAATGGTGAATGTCTAAACATGGAAAAGGTCTATATCTCAACC
AATTGCCCCTCTCAGTGCAATGAAAATCCTGTGGATGGCCACGGACTCCAGTGCCACTGTGAGGAAGG
ACAGGCACCTGTAGCCTGTGAAGAAACCTTACATGTTACCAATATCACCATCTTGGTTGTTGTGCTTG
TCCTGGTTATTGTCGGTATCGGAGTTCTTATACTATTAGTTCGTTACCGAAAATGTATCAAGTTGAAG
CAAGTTCAGAGCCCACCTACAGAAACCCTGGGAGTGGAGAACAAAGGATACTTTGGTGATGAGCAGCA
GATAAGGACTGAGCCAATCCTGCCGGAAATTCATTTCCTAAATAAACCTGCAAGTAAAGATTCAAGAG
GAATCGCAGATCCCAATCAAAGTGCCAAGTGAGCTTGAA

NOV22e, CG51983-04 Protein Sequence	SEQ ID NO: 298	754 aa	MW at 85582.1kD

MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDILKTYEEELLYEIKLNRK
TLVLHLLRSREFLGSNYSETFYSMKGGAFTRHPQIMDHCFYQGSIVHEYDSAASISTCNGLRGFFRIN
DQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKTVPGDNESEEDSKIKGIHDEKYVELF
IVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYKTLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFS
FWQEKILKTRKDFDHVVLLSGKWLYSHVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHN
LGMQHDEFPCTCPSGKCVMDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNK
KLDEGEECDCGPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGH
SPACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDDAIESHDICYKMNTKGNKFGYCKNKEN
RFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDSTDIGLVASGTKCGE

GMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPVACEETLHVTNITILVVVLVLVI
VGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFGDEQQIRTEPILPEIHFLNKPASKDSRGIAD
PNQSAK

NOV22f, CG51983-06 DNA Sequence	SEQ ID NO: 299	2464 bp
	ORF Start: ATG at 4	ORF Stop: TGA at 2248

AGAATGCTTCCCGGGTGTATATTCTTGATGATTTTACTCATTCCTCAGGTTAAAGAAAAGTTCATCCT
TGGAGTAGAGGGTCAACAACTGGTTCGTCCTAAAAAGCTTCCTCTGATACAGAAGCGAGATACTGGAC
ACACCCATGATGATGACATACTGAAAACGTATGAAGAAGAATTGTTGTATGAAATAAAACTAAATAGA
AAAACCTTAGTCCTTCATCTTCTAAGATCCAGGGAGTTCCTAGGCTCAAATTACAGTGAAACATTCTA
CTCCATGAAAGGAGAAGCGTTCACCAGGCATCCTCAGATCATGGATCATTGTTTTTACCAAGGATCCA
TAGTACACGAATATGATTCAGCTGCCAGTATCAGTACGTGTAATGGTCTAAGGGGATTCTTCAGAATA
AACGACCAAAGATACCTCATTGAACCAGTGAAATACTCAGATGAGGGAGAACATTTGGTGTTCAAATA
TAACCTGAGGGTGCCGTATGGTGCCAATTATTCCTGTACAGAGCTTAATTTTACCAGAAAAACTGTTC
CAGGGGATAATGAATCTGAAGAAGACTCCAAAATAAAAGGCATCCATGATGAAAAGTATGTTGAATTG
TTCATTGTTGCTGATGATACTGTGTATCGCAGAAATGGTCATCCTCACAATAAACTAAGGAACCGAAT
TTGGGGAATGGTCAATTTTGTCAACATGATTTATAAAACCTTAAACATCCATGTGACGTTGGTTGGCA
TTGAAATATGGACACATGAAGATAAAATAGAACTATATTCAAATATAGAAACTACCTTATTGCGTTTT
TCATTTTGGCAAGAAAAGATCCTTAAAACACGGAAGGATTTTGATCATGTTGTATTACTCAGTGGGAA
GTGGCTCTACTCACATGTGCAAGGAATTTCTTATCCAGGGGGTATGTGCCTGCCCTATTATTCCACCA
GTATCATTAAGGATCTTTTACCTGACACAAACATAATTGCAAACAGAATGGCACATCAACTGGGGCAT
AACCTTGGGATGCAGCATGACGAGTTCCCATGCACCTGTCCTTCAGGAAAATGCGTGATGGACAGTGA
TGGAAGCATTCCTGCACTGAAATTCAGTAAATGCAGCCAAAACCAATACCACCAGTACTTGAAGGATT
ATAAGCCAACATGCATGCTCAACATTCCATTTCCTTACAATTTTCATGATTTCCAATTTTGTGGAAAC
AAGAAGTTGGATGAGGGTGAAGAGTGTGACTGTGGCCCTGCTCAGGAGTGTACTAATCCTTGCTGTGA
TGCACACACATGTGTACTGAAGCCAGGATTTACTTGTGCAGAAGGAGAATGCTGTGAATCTTGTCAGA
TAAAAAAAGCAGGGTCCATATGCAGACCGGCGAAAGATGAATGTGATTTTCCTGAGATGTGCACTGGC
CACTCGCCTGCCTGTCCTAAGGACCAGTTCAGGGTCAATGGATTTCCTTGCAAGAACTCAGAAGGCTA
CTGTTTCATGGGGAAATGTCCAACTCGTGAGGATCAGTGCTCTGAACTATTTGATGATGAGGCAATAG
AGAGTCATGATATCTGCTACAAGATGAATACAAAAGGAAATAAATTTGGATACTGCAAAAACAAGGAA
AACAGATTTCTTCCCTGTGAGGAGAAAGATGTCAGATGTGGAAAGACCTACTGCACTGGAGGGGAGCT
TTCCTCTCTCCTTGGAGAAGACAAGACTTATCACCTTAAGGATCCCCAGAAGAATGCTACTGTCAAAT
GCAAAACTATTTTTTTATACCATGATTCTACAGACATTGGCCTGGTGGCGTCAGGAACAAAATGTGGA
GAGGGAATGGTGGATGGCCACGGACTCCAGTGCCACTGTGAGGAAGGACAGGCACCTGTAGCCTGTGA
AGAAACCTTACATGTTACCAATATCACCATCTTGGTTGTTGTGCTTGTCCTGGTTATTGTCGGTATCG
GAGTTCTTATACTATTAGTTCGTTACCGAAAATGTATCAAGTTGAAGCAAGTTCAGAGCCCACCTACA
GAAACCCTGGGAGTGGAGAACAAAGGATACTTTGGTGATGAGCAGCAGATAAGGACTGAGCCAATCCT
GCCAGAAATTCATTTCCTAAATAGAACTCCAGAATCCTTGGAAAGCCTGCCCACTAGTTTTTCAAGTC
CCCACTACATCACACTGAAACCTGCAAGTAAAGATTCAAGAGGAATCGCAGATCCCAATCAAAGTGCC
AAGTGAGCTTGAAGTTGGATATCCAAAATGGCCGTGCAAGCTTAGGCTGGGGATTCTGGATGCAACGT
CTTTACAACCTTACCTAGATATCTGCTACTCACATTTTTGGTAGTGTTTCAAACGTTCTTTATCCAGA
CAGACAATGTTTAAGAGAAACAACTTATTTCTGTTAATATTTACCGGTAGAATTCACACCCTCTATCA
TAAACATATGCTGCAG

NOV22f, CG51983-06 Protein Sequence	SEQ ID NO: 300	748 aa	MW at 85014.4kD

MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDILKTYEEELLYEIKLNRK
TLVLHLLRSREFLGSNYSETFYSMKGEAFTRHPQIMDHCFYQGSIVHEYDSAASISTCNGLRGFFRIN
DQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKTVPGDNESEEDSKIKGIHDEKYVELF
IVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYKTLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFS
FWQEKILKTRKDFDHVVLLSGKWLYSHVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHN
LGMQHDEFPCTCPSGKCVMDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNK
KLDEGEECDCGPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGH
SPACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGYCKNKEN
RFLPCEEKDVRCGKTYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDSTDIGLVASGTKCGE
GMVDGHGLQCHCEEGQAPVACEETLHVTNITILVVVLVLVIVGIGVLILLVRYRKCIKLKQVQSPPTE
TLGVENKGYFGDEQQIRTEPILPEIHFLNRTPESLESLPTSFSSPHYITLKPASKDSRGIADPNQSAK

NOV22g, SNP13376585 of CG51983-05, DNA Sequence	SEQ ID NO: 301	2268 bp
	ORF Start: ATG at 1	ORF Stop: TAG at 2266
	SNP Pos: 119	SNP Change: A to G

ATGCTTCCCGGGTGTATATTCTTGATGATTTTACTCATTCCTCAGGTTAAAGAAAAGTTCATCCTTGG
AGTAGAGGGTCAACAACTGGTTCGTCCTAAAAAGCTTCCTCTGATACAGAGGCGAGATACTGGACACA
CCCATGATGATGACATACTGAAAACGTATGAAGAAGAATTGTTGTATGAAATAAAACTAAATAGAAAA
ACCTTAGTCCTTCATCTTCTAAGATCCAGGGAGTTCCTAGGCTCAAATTACAGTGAAACATTCTACTC
CATGAAAGGAGAAGCGTTCACCAGGCATCCTCAGATCATGGAACACTGTTACTATAAAGGAAACATCC
TAAATGAAAAGAATTCTGTTGCCAGCATCAGTACTTGTGACGGGTTGAGGGGATTCTTCAGAATAAAC
GACCAAAGATACCTCATTGAACCAGTGAAATACTCAGATGAGGGAGAACATTTGGTGTTCAAATATAA
CCTGAGGGTGCCGTATGGTGCCAATTATTCCTGTACAGAGCTTAATTTTACCAGAAAAACTGTTCCAG
GGGATAATGAATCTGAAGAAGACTCCAAAATAAAAGGCATCCATGATGAAAAGTATGTTGAATTGTTC
ATTGTTGCTGATGATACTGTGTATCGCAGAAATGGTCATCCTCACAATAAACTAAGGAACCGAATTTG
GGGAATGGTCAATTTTGTCAACATGATTTATAAAACCTTAAACATCCATGTGACGTTGGTTGGCATTG
AAATATGGACACATGAAGATAAAATAGAACTATATTCAAATATAGAAACTACCTTATTGCGTTTTTCA
TTTTGGCAAGAAAAGATCCTTAAAACACGGAAGGATTTTGATCATGTTGTATTACTCAGTGGGAAGTG
GCTCTACTCACATGTGCAAGGAATTTCTTATCCAGGGGGTATGTGCCTGCCCTATTATTCCACCAGTA
TCATTAAGGATCTTTTACCTGACACAAACATAATTGCAAACAGAATGGCACATCAACTGGGGCATAAC
CTTGGGATGCAGCATGACGAGTTCCCATGCACCTGTCCTTCAGGAAAATGCGTGATGGACAGTGATGG
AAGCATTCCTGCACTGAAATTCAGTAAATGCAGCCAAAACCAATACCACCAGTACTTGAAGGATTATA
AGCCAACATGCATGCTCAACATTCCATTTCCTTACAATTTTCATGATTTCCAATTTTGTGGAAACAAG
AAGTTGGATGAGGGTGAAGAGTGTGACTGTGGCCCTGCTCAGGAGTGTACTAATCCTTGCTGTGATGC
ACACACATGTGTACTGAAGCCAGGATTTACTTGTGCAGAAGGAGAATGCTGTGAATCTTGTCAGATAA
AAAAAGCAGGGTCCATATGCAGACCGGCGAAAGATGAATGTGATTTTCCTGAGATGTGCACTGGCCAC
TCGCCTGCCTGTCCTAAGGACCAGTTCAGGGTCAATGGATTTCCTTGCAAGAACTCAGAAGGCTACTG
TTTCATGGGGAAATGTCCAACTCGTGAGGATCAGTGCTCTGAACTATTTGATGATGAGGCAATAGAGA
GTCATGATATCTGCTACAAGATGAATACAAAAGGAAATAAATTTGGATACTGCAAAAACAAGGAAAAC
AGATTTCTTCCCTGTGAGGAGAAGGATGTCAGATGTGGAAAGATCTACTGCACTGGAGGGGAGCTTTC
CTCTCTCCTTGGAGAAGACAAGACTTATCACCTTAAGGATCCCCAGAAGAATGCTACTGTCAAATGCA
AAACTATTTTTTTATACCATGATTCTACAGACATTGGCCTGGTGGCGTCAGGAACAAAATGTGGAGAG
GGAATGGTATGCAACAATGGTGAATGTCTAAACATGGAAAAGGTCTATATCTCAACCAATTGCCCCTC
TCAGTGCAATGAAAATCCTGTAGATGGCCACGGACTCCAGTGCCACTGTGAGGAAGGACAGGCACCTG
TAGCCTGTGAAGAAACCTTACATGTTACCAGTATCACCATCTTGGTTGTTGTGCTTGTCCTGGTTATT
GTCGGTATCGGAGTTCTTATACTATTAGTTCGTTACCGAAAATGTATCAAGTTGAAGCAAGTTCAGAG
CCCACCTACAGAAACCCTGGGAGTGGAGAACAAAGGATACTTTGGTGATGAGCAGCAGATAAGGACTG
AGCCAATCCTGCCAGAAATTCATTTCCTAAATAAACCTGCAAGTAAAGATTCAAGAGGAATCGCAGAT
CCCAATCAAAGTGCCAAGTGGTAG

NOV22g, SNP13376585 of CG51983-05, Protein Sequence	SEQ ID NO: 302	755 aa	MW at 85896.5kD
	SNP Pos: 40		SNP Change: Lys to Arg

MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQRRDTGHTHDDDILKTYEEELLYEIKLNRK
TLVLHLLRSREFLGSNYSETFYSMKGEAFTRHPQIMEHCYYKGNILNEKNSVASISTCDGLRGFFRIN
DQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKTVPGDNESEEDSKIKGIHDEKYVELF
IVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYKTLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFS
FWQEKILKTRKDFDHVVLLSGKWLYSHVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHN
LGMQHDEFPCTCPSGKCVMDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNK
KLDEGEECDCGPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGH
SPACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGYCKNKEN
RFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDSTDIGLVASGTKCGE
GMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPVACEETLHVTSITILVVVLVLVI
VGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFGDEQQIRTEPILPEIHFLNKPASKDSRGIAD
PNQSAKW

NOV22h, SNP13376586 of CG51983-05, DNA Sequence	SEQ ID NO: 303	2268 bp
	ORF Start: ATG at 1	ORF Stop: TAG at 2266
	SNP Pos: 125	SNP Change: A to G

ATGCTTCCCGGGTGTATATTCTTGATGATTTTACTCATTCCTCAGGTTAAAGAAAAGTTCATCCTTGG
AGTAGAGGGTCAACAACTGGTTCGTCCTAAAAAGCTTCCTCTGATACAGAAGCGAGGTACTGGACACA
CCCATGATGATGACATACTGAAAACGTATGAAGAAGAATTGTTGTATGAAATAAAACTAAATAGAAAA
ACCTTAGTCCTTCATCTTCTAAGATCCAGGGAGTTCCTAGGCTCAAATTACAGTGAAACATTCTACTC
CATGAAAGGAGAAGCGTTCACCAGGCATCCTCAGATCATGGAACACTGTTACTATAAAGGAAACATCC
TAAATGAAAAGAATTCTGTTGCCAGCATCAGTACTTGTGACGGGTTGAGGGGATTCTTCAGAATAAAC
GACCAAAGATACCTCATTGAACCAGTGAAATACTCAGATGAGGGAGAACATTTGGTGTTCAAATATAA
CCTGAGGGTGCCGTATGGTGCCAATTATTCCTGTACAGAGCTTAATTTTACCAGAAAAACTGTTCCAG
GGGATAATGAATCTGAAGAAGACTCCAAAATAAAAGGCATCCATGATGAAAAGTATGTTGAATTGTTC
ATTGTTGCTGATGATACTGTGTATCGCAGAAATGGTCATCCTCACAATAAACTAAGGAACCGAATTTG
GGGAATGGTCAATTTTGTCAACATGATTTATAAAACCTTAAACATCCATGTGACGTTGGTTGGCATTG
AAATATGGACACATGAAGATAAAATAGAACTATATTCAAATATAGAAACTACCTTATTGCGTTTTTCA
TTTTGGCAAGAAAAGATCCTTAAAACACGGAAGGATTTTGATCATGTTGTATTACTCAGTGGGAAGTG
GCTCTACTCACATGTGCAAGGAATTTCTTATCCAGGGGGTATGTGCCTGCCCTATTATTCCACCAGTA
TCATTAAGGATCTTTTACCTGACACAAACATAATTGCAAACAGAATGGCACATCAACTGGGGCATAAC
CTTGGGATGCAGCATGACGAGTTCCCATGCACCTGTCCTTCAGGAAAATGCGTGATGGACAGTGATGG
AAGCATTCCTGCACTGAAATTCAGTAAATGCAGCCAAAACCAATACCACCAGTACTTGAAGGATTATA
AGCCAACATGCATGCTCAACATTCCATTTCCTTACAATTTTCATGATTTCCAATTTTGTGGAAACAAG
AAGTTGGATGAGGGTGAAGAGTGTGACTGTGGCCCTGCTCAGGAGTGTACTAATCCTTGCTGTGATGC
ACACACATGTGTACTGAAGCCAGGATTTACTTGTGCAGAAGGAGAATGCTGTGAATCTTGTCAGATAA
AAAAAGCAGGGTCCATATGCAGACCGGCGAAAGATGAATGTGATTTTCCTGAGATGTGCACTGGCCAC
TCGCCTGCCTGTCCTAAGGACCAGTTCAGGGTCAATGGATTTCCTTGCAAGAACTCAGAAGGCTACTG
TTTCATGGGGAAATGTCCAACTCGTGAGGATCAGTGCTCTGAACTATTTGATGATGAGGCAATAGAGA
GTCATGATATCTGCTACAAGATGAATACAAAAGGAAATAAATTTGGATACTGCAAAAACAAGGAAAAC
AGATTTCTTCCCTGTGAGGAGAAGGATGTCAGATGTGGAAAGATCTACTGCACTGGAGGGGAGCTTTC
CTCTCTCCTTGGAGAAGACAAGACTTATCACCTTAAGGATCCCCAGAAGAATGCTACTGTCAAATGCA
AAACTATTTTTTTATACCATGATTCTACAGACATTGGCCTGGTGGCGTCAGGAACAAAATGTGGAGAG
GGAATGGTATGCAACAATGGTGAATGTCTAAACATGGAAAAGGTCTATATCTCAACCAATTGCCCCTC
TCAGTGCAATGAAAATCCTGTAGATGGCCACGGACTCCAGTGCCACTGTGAGGAAGGACAGGCACCTG
TAGCCTGTGAAGAAACCTTACATGTTACCAGTATCACCATCTTGGTTGTTGTGCTTGTCCTGGTTATT
GTCGGTATCGGAGTTCTTATACTATTAGTTCGTTACCGAAAATGTATCAAGTTGAAGCAAGTTCAGAG
CCCACCTACAGAAACCCTGGGAGTGGAGAACAAAGGATACTTTGGTGATGAGCAGCAGATAAGGACTG
AGCCAATCCTGCCAGAAATTCATTTCCTAAATAAACCTGCAAGTAAAGATTCAAGAGGAATCGCAGAT
CCCAATCAAAGTGCCAAGTGG**TAG**

NOV22h, SNP13376586 of CG51983-05, Protein Sequence	SEQ ID NO: 304	755 aa	MW at 85810.5kD
	SNP Pos: 42		SNP Change: Asp to Gly

MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRGTGHTHDDDILKTYEEELLYEIKLNRK
TLVLHLLRSREFLGSNYSETFYSMKGEAFTRHPQIMEHCYYKGNILNEKNSVASISTCDGLRGFFRIN
DQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKTVPGDNESEEDSKIKGIHDEKYVELF
IVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYKTLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFS
FWQEKILKTRKDFDHVVLLSGKWLYSHVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHN
LGMQHDEFPCTCPSGKCVMDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNK
KLDEGEECDCGPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGH
SPACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGYCKNKEN
RFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDSTDIGLVASGTKCGE
GMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPVACEETLHVTSITILVVVLVLVI
VGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFGDEQQIRTEPILPEIHFLNKPASKDSRGIAD
PNQSAKW

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 22B.

Table 22B. Comparison of the NOV22 protein sequences.

```
NOV22a  MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDILKTYEEELL
NOV22b  MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDIK-TYEEELL
NOV22c  MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDILKTYEEELL
NOV22d  MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDIK-TYEEELL
NOV22e  MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDILKTYEEELL
NOV22f  MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDILKTYEEELL

NOV22a  YEIKLNRKTLVLHLLRSR-EFLGSNYSETFYSMKGEAFTRHPQIMEHCYYKGNILNEKNS
NOV22b  YEIKLNRKTLVLHLLRSRREFLGSNYSETFYSMKGEAFTRHPQIMDHCFYQGSIVHEYDS
NOV22c  YEIKLNRKTLVLHLLRSR-EFLGSNYSETFYSMKGGAFTRHPQIMDHCFYQGSIVHEYDS
NOV22d  YEIKLNRKTLVLHLLRSRREFLGSNYSETFYSMKGEAFTRHPQIMDHCFYQGSIVHEYDS
NOV22e  YEIKLNRKTLVLHLLRSR-EFLGSNYSETFYSMKGGAFTRHPQIMDHCFYQGSIVHEYDS
NOV22f  YEIKLNRKTLVLHLLRSR-EFLGSNYSETFYSMKGEAFTRHPQIMDHCFYQGSIVHEYDS

NOV22a  VASISTCDGLRGFFRINDQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKT
NOV22b  AASISTCNGLRGFFRINDQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKT
NOV22c  AASISTCNGLRGFFRINDQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKT
NOV22d  AASISTCNGLRGFFRINDQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKT
NOV22e  AASISTCNGLRGFFRINDQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKT
NOV22f  AASISTCNGLRGFFRINDQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKT

NOV22a  VPGDNESEEDSKIKG-IHDEKYVELFIVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYK
NOV22b  VPGDNESEEDSKIKQGIHDEKYVELFIVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYK
NOV22c  VPGDNESEEDSKIKG-IHDEKYVELFIVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYK
NOV22d  VPGDNESEEDSKIKQGIHDEKYVELFIVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYK
NOV22e  VPGDNESEEDSKIKG-IHDEKYVELFIVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYK
NOV22f  VPGDNESEEDSKIKG-IHDEKYVELFIVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYK

NOV22a  TLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFSFWQEKILKTRKDFDHVVLLSGKWLYS
NOV22b  TLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFSFWQEKILKTRKDFDHVVLLSGKWLYS
NOV22c  TLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFSFWQEKILKTRKDFDHVVLLSGKWLYS
NOV22d  TLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFSFWQEKILKTRKDFDHVVLLSGKWLYS
NOV22e  TLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFSFWQEKILKTRKDFDHVVLLSGKWLYS
NOV22f  TLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFSFWQEKILKTRKDFDHVVLLSGKWLYS

NOV22a  HVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHNLGMQHDEFPCTCPSGKCV
NOV22b  HVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHNLGMQHDEFPCTCPSGKCV
NOV22c  HVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHNLGMQHDEFPCTCPSGKCV
NOV22d  HVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHNLGMQHDEFPCTCPSGKCV
NOV22e  HVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHNLGMQHDEFPCTCPSGKCV
NOV22f  HVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHNLGMQHDEFPCTCPSGKCV

NOV22a  MDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNKKLDEGEECDC
NOV22b  MDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNKKLDEGEECDC
NOV22c  MDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNKKLDEGEECDC
NOV22d  MDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNKKLDEGEECDC
NOV22e  MDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNKKLDEGEECDC
NOV22f  MDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNKKLDEGEECDC

NOV22a  GPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGHSP
NOV22b  GPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGHSP
```

```
NOV22c  GPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGHSP
NOV22d  GPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGHSP
NOV22e  GPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGHSP
NOV22f  GPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGHSP

NOV22a  ACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGY
NOV22b  ACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGY
NOV22c  ACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDDAIESHDICYKMNTKGNKFGY
NOV22d  ACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGY
NOV22e  ACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDDAIESHDICYKMNTKGNKFGY
NOV22f  ACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGY

NOV22a  CKNKENRFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDS
NOV22b  CKNKENRFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDS
NOV22c  CKNKENRFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDS
NOV22d  CKNKENRFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDS
NOV22e  CKNKENRFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDS
NOV22f  CKNKENRFLPCEEKDVRCGKTYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDS

NOV22a  TDIGLVASGTKCGEGMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPV
NOV22b  TDIGLVASGTKCGEGMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPV
NOV22c  TDIGLVASGTKCGEGMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPV
NOV22d  TDIGLVASGTKCGEGMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPV
NOV22e  TDIGLVASGTKCGEGMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPV
NOV22f  TDIGLVASGTKCGEGMVD-------------------------GHGLQCHCEEGQAPV

NOV22a  ACEETLHVTSITILVVVLVLVIVGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFG
NOV22b  ACEETLHVTSITILVVVLVLVIVGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFG
NOV22c  ACEETLHVTNITILVVVLVLVIVGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFG
NOV22d  ACEETLHVTSITILVVVLVLVIVGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFG
NOV22e  ACEETLHVTNITILVVVLVLVIVGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFG
NOV22f  ACEETLHVTNITILVVVLVLVIVGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFG

NOV22a  DEQQIRTEPILPEIHFLN----------------------KPASKDSRGIADPNQSAKW
NOV22b  DEQQIRTEPILPEIHFLNQRTPESLESLPTSFSSPHYITLKPASKDSRGIADPNQSAKW
NOV22c  DEQQIRTEPILPEIHFLN----------------------KPASKDSRGIADPNQSAK-
NOV22d  DEQQIRTEPILPEIHFLN----------------------KPASKDSRGIADPNQSAKW
NOV22e  DEQQIRTEPILPEIHFLN----------------------KPASKDSRGIADPNQSAK-
NOV22f  DEQQIRTEPILPEIHFLN-RTPESLESLPTSFSSPHYITLKPASKDSRGIADPNQSAK-

NOV22a  (SEQ ID NO:  290)
NOV22b  (SEQ ID NO:  292)
NOV22c  (SEQ ID NO:  294)
NOV22d  (SEQ ID NO:  296)
NOV22e  (SEQ ID NO:  298)
NOV22f  (SEQ ID NO:  300)
```

Further analysis of the NOV22a protein yielded the following properties shown in Table 22C.

Table 22C. Protein Sequence Properties NOV22a

SignalP analysis:	Cleavage site between residues 19 and 20

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 16;  peak value  12.97
      PSG score:   8.57

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -5.58
      possible cleavage site: between 18 and 19

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   2
      INTEGRAL    Likelihood = -2.18   Transmembrane    1 -  17
      INTEGRAL    Likelihood =-18.52   Transmembrane  671 - 687
      PERIPHERAL  Likelihood =  4.35 (at 235)
      ALOM score: -18.52  (number of TMSs: 2)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 8
      Charge difference:  0.0   C( 1.0) - N( 1.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  3.80
      Hyd Moment(95):  1.48    G content:       1
      D/E content:     1       S/T content:     0
      Score: -6.33

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: RPKK (4) at   31
      pat7: none
      bipartite: none
      content of basic residues:  11.5%
      NLS Score: -0.22

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: QSAK

SKL: peroxisomal targeting signal in the C-terminus: none
```

```
PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:
      Bacterial regulatory proteins, gntR family signature (PS00043):
*** found ***
         EEELLYEIKLNRKTLVLHLLRS at 56

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 55.5

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       34.8 %: nuclear
       30.4 %: endoplasmic reticulum
       21.7 %: mitochondrial
        4.3 %: vesicles of secretory system
        4.3 %: cytoplasmic
        4.3 %: peroxisomal

>> prediction for CG51983-05 is nuc (k=23)
```

A search of the NOV22a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 22D.

Table 22D. Geneseq Results for NOV22a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV22a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB47567	Protease PRTS-9 - Homo sapiens, 776 aa. [WO200171004-A2, 27-SEP-2001]	1..754 1..776	749/776 (96%) 751/776 (96%)	0.0
AAU77409	Human NOV4b protein, homologue of ADAM proteins - Homo sapiens, 779 aa. [WO200206329-A2, 24-JAN-2002]	1..755 1..779	754/780 (96%) 754/780 (96%)	0.0
AAU77408	Human NOV4a protein, homologue of ADAM proteins - Homo sapiens, 778 aa. [WO200206329-A2, 24-JAN-2002]	1..755 1..778	744/779 (95%) 752/779 (96%)	0.0
AAU16950	Human novel secreted protein, SEQ ID 191 - Homo sapiens, 695 aa. [WO200155441-A2, 02-AUG-2001]	1..677 17..693	664/677 (98%) 675/677 (99%)	0.0
AAB64744	Gene 16 human secreted protein homologous amino acid sequence #138 - Macaca fascicularis, 375 aa. [WO200077237-A1, 21-DEC-2000]	336..710 1..375	353/375 (94%) 364/375 (96%)	0.0

In a BLAST search of public sequence databases, the NOV22a protein was found to have homology to the proteins shown in the BLASTP data in Table 22E.

Table 22E. Public BLASTP Results for NOV22a				
Protein Accession Number	**Protein/Organism/Length**	**NOV22a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9H2U9	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens (Human), 754 aa.	1..754 1..754	741/754 (98%) 751/754 (99%)	0.0
Q28475	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Epididymal apical protein I) (EAP I) - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 776 aa.	1..754 1..776	695/776 (89%) 726/776 (92%)	0.0
AAH43207	Hypothetical protein - Homo sapiens (Human), 561 aa (fragment).	212..736 29..553	524/525 (99%) 525/525 (99%)	0.0
O35227	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) - Mus musculus (Mouse), 788 aa.	1..752 1..771	507/773 (65%) 620/773 (79%)	0.0
Q63180	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Epididymal apical protein I) (EAP I) - Rattus norvegicus (Rat), 789 aa.	1..752 1..772	505/773 (65%) 614/773 (79%)	0.0

PFam analysis indicates that the NOV22a protein contains the domains shown in the Table 22F.

Table 22F. Domain Analysis of NOV22a			
Pfam Domain	**NOV22a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Pep_M12B_propep	73..188	38/119 (32%) 100/119 (84%)	6.7e-45
Reprolysin	199..394	88/203 (43%) 173/203 (85%)	5.5e-104
disintegrin	411..486	41/76 (54%) 53/76 (70%)	1.6e-25

Example 23.

The NOV23 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 23A.

Table 23A. NOV23 Sequence Analysis			
NOV23a, CG53390-02 DNA Sequence	SEQ ID NO: 305		994 bp
	ORF Start: ATG at 27		ORF Stop: TGA at 969
TGCAGCTAAAGTGCATTGTGTAAAAC**ATG**GGGGATGTGAATCAGTCGGTGGCCTCAGACTTCATTCTG GTGGGCCTCTTCAGTCACTCAGGATCACGCCAGCTCCTCTTCTCCCTGGTGGCTGTCATGTTTGTCAT AGGCCTTCTGGGCAACACCGTTCTTCTCTTCTTGATCCGTGTGGACTCCCGGCTCCATACACCCATGT ACTTCCTGCTCAGCCAGCTCTCCCTGTTTGACATTGGCTGTCCCATGGTCACCATCCCCAAGATGGCA TCAGACTTTCTGCGGGGAGAAGGTGCCACCTCCTATGGAGGTGGTGCAGCTCAAATATTCTTCCTCAC ACTGATGGGTGTGGCTGAGGGCGTCCTGTTGGTCCTCATGTCTTATGACCGTTATGTTGCTGTGTGCC AGCCCCTGCAGTATCCTGTACTTATGAGACGCCAGGTATGTCTGCTGATGATGGGCTCCTCCTGGGTG GTAGGTGTGCTCAACGCCTCCATCCAGACCTCCATCACCCTGCATTTTCCCTACTGTGCCTCCCGTAT TGTGGATCACTTCTTCTGTGAGGTGCCAGCCCTACTGAAGCTCTCCTGTGCAGATACCTGTGCCTACG AGATGGCGCTGTCCACCTCAGGGGTGCTGATCCTAATGCTCCCTCTTTCCCTCATCGCCACCTCCTAC GGCCACGTGTTGCAGGCTGTTCTAAGCATGCGCTCAGAGGAGGCCAGACACAAGGCTGTCACCACCTG CTCCTCGCACATCACGGTAGTGGGGCTCTTTTATGGTGCCGCCGTGTTCATGTACATGGTGCCTTGCG CCTACCACAGTCCACAGCAGGATAACGTGGTTTCCCTCTTCTATAGCCTTGTCACCCCTACACTCAAC CCCCTTATCTACAGTCTGAGGAATCCGGAGGTGTGGATGGCTTTGGTCAAAGTGCTTAGCAGAGCTGG ACTCAGGCAAATGTGC**TGA**CTACATAGAAACTGCTGGTGAGA			
NOV23a, CG53390-02 Protein Sequence	SEQ ID NO: 306	314 aa	MW at 34443.4kD
MGDVNQSVASDFILVGLFSHSGSRQLLFSLVAVMFVIGLLGNTVLLFLIRVDSRLHTPMYFLLSQLSL FDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMSYDRYVAVCQPLQYPVLM RRQVCLLMMGSSWVVGVLNASIQTSITLHFPYCASRIVDHFFCEVPALLKLSCADTCAYEMALSTSGV LILMLPLSLIATSYGHVLQAVLSMRSEEARHKAVTTCSSHITVVGLFYGAAVFMYMVPCAYHSPQQDN VVSLFYSLVTPTLNPLIYSLRNPEVWMALVKVLSRAGLRQMC			
NOV23b, CG53390-01 DNA Sequence	SEQ ID NO: 307		994 bp
	ORF Start: ATG at 27		ORF Stop: TGA at 969
TGCAGCTAAAGTGCATTGTGTAAAAC**ATG**GGGGATGTGAATCAGTCGGTGGCCTCAGACTTCATTCTG			

GTGGGCCTCTTCAGTCACTCAGGATCACGCCAGCTCCTCTTCTCCCTGGTGGCTGTCATGTTTGTCAT
AGGCCTTCTGGGCAACACCGTTCTTCTCTTCTTGATCCGTGTGGACTCCCGGCTCCATACACCCATGT
ACTTCCTGCTCAGCCAGCTCTCCCTGTTTGACATTGGCTGTCCCATGGTCACCATCCCCAAGATGGCA
TCAGACTTTCTGCGGGGAGAAGGTGCCACCTCCTATGGAGGTGGTGCAGCTCAAATATTCTTCCTCAC
ACTGATGGGTGTGGCTGAGGGCGTCCTGTTGGTCCTCATGTCTTATGACCGTTATGTTGCTGTGTGCC
AGCCCCTGCAGTATCCTGTACTTATGAGACGCCAGGTATGTCTGCTGATGATGGGCTCCTCCTGGGTG
GTAGGTGTGCTCAACGCCTCCATCCAGACCTCCATCACCCTGCATTTTCCCTACTGTGCCTCCCGTAT
TGTGGATCACTTCTTCTGTGAGGTGCCAGCCCTACTGAAGCTCTCCTGTGCAGATACCTGTGCCTACG
AGATGGCGCTGTCCACCTCAGGGGTGCTGATCCTAATGCTCCCTCTTTCCCTCATCGCCACCTCCTAC
GGCCACGTGTTGCAGGCTGTTCTAAGCATGCGCTCAGAGGAGGCCAGACACAAGGCTGTCACCACCTG
CTCCTCGCACATCACGGTAGTGGGGCTCTTTTATGGTGCCGCCGTGTTCATGTACATGGTGCCTTGCG
CCTACCACAGTCCACAGCAGGATAACGTGGTTTCCCTCTTCTATAGCCTTGTCACCCCTACACTCAAC
CCCCTTATCTACAGTCTGAGGAATCCGGAGGTGTGGATGGCTTTGGTCAAAGTGCTTAGCAGAGCTGG
ACTCAGGCAAATGTGCTGACTACATAGAAACTGCTGGTGAGA

NOV23b, CG53390-01 Protein Sequence	SEQ ID NO: 308	314 aa	MW at 34443.4kD

MGDVNQSVASDFILVGLFSHSGSRQLLFSLVAVMFVIGLLGNTVLLFLIRVDSRLHTPMYFLLSQLSL
FDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMSYDRYVAVCQPLQYPVLM
RRQVCLLMMGSSWVVGVLNASIQTSITLHFPYCASRIVDHFFCEVPALLKLSCADTCAYEMALSTSGV
LILMLPLSLIATSYGHVLQAVLSMRSEEARHKAVTTCSSHITVVGLFYGAAVFMYMVPCAYHSPQQDN
VVSLFYSLVTPTLNPLIYSLRNPEVWMALVKVLSRAGLRQMC

NOV23c, CG53390-03 DNA Sequence	SEQ ID NO: 309	977 bp
	ORF Start: ATG at 16	ORF Stop: TGA at 958

TGCATTGTGTAAAACATGGGGGATGTGAATCAGTCGGTGGCCTCAGACTTCATTCTGGTGGGCCTCTT
CAGTCACTCAGGATCACGCCAGCTCCTCTTCTCCCTGGTGGCTGTCATGTTTGTCATAGGCCTTCTGG
GCAACACCGTTCTTCTCTTCTTGATCCGTGTGGACTCCCGGCTCCATACACCCATGTACTTCCTGCTC
AGCCAGCTCTCCCTGTTTGACATTGGCTGTCCCATGGTCACCATCCCCAAGATGGCATCAGACTTTCT
GCGGGGAGAAGGTGCCACCTCCTATGGAGGTGGTGCAGCTCAAATATTCTTCCTCACACTGATGGGTG
TGGCTGAGGGCGTCCTGTTGGTCCTCATGTCTTATGACCGTTATGTTGCTGTGTGCCAGCCCCTGCAG
TATCCTGTACTTATGAGACGCCAGGTATGTCTGCTGATGATGGGCTCCTCCTGGGTGGTAGGTGTGCT
CAACGCCTCCATCCAGACCTCCATCACCCTGCATTTTCCCTACTGTGCCTCCCGTATTGTGGATCACT
TCTTCTGTGAGGTGCCAGCCCTACTGAAGCTCTCCTGTGCAGATACCTGTGCCTACGAGATGGCGCTG
TCCACCTCAGGGGTGCTGATCCTAATGCTCCCTCTTTCCCTCATCGCCACCTCCTACGGCCACGTGTT
GCAGGCTGTTCTAAGCATGCGCTCAGAGGAGGCCAGACACAAGGCTGTCACCACCTGCTCCTCGCACA
TCACGGTAGTGGGGCTCTTTTATGGTGCCGCCGTGTTCATGTACATGGTGCCTTGCGCCTACCACAGT
CCACAGCAGGATAACGTGGTTTCCCTCTTCTATAGCCTTGTCACCCCTACACTCAACCCCCTTATCTA
CAGTCTGAGGAATCCGGAGGTGTGGATGGCTTTGGTCAAAGTGCTTAGCAGAGCTGGACTCAGGCAAA
TGTGCTGACTACATAGAAACTGCTG

NOV23c, CG53390-03 Protein Sequence	SEQ ID NO: 310	314 aa	MW at 34443.4kD

MGDVNQSVASDFILVGLFSHSGSRQLLFSLVAVMFVIGLLGNTVLLFLIRVDSRLHTPMYFLLSQLSL
FDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMSYDRYVAVCQPLQYPVLM
RRQVCLLMMGSSWVVGVLNASIQTSITLHFPYCASRIVDHFFCEVPALLKLSCADTCAYEMALSTSGV
LILMLPLSLIATSYGHVLQAVLSMRSEEARHKAVTTCSSHITVVGLFYGAAVFMYMVPCAYHSPQQDN
VVSLFYSLVTPTLNPLIYSLRNPEVWMALVKVLSRAGLRQMC

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 23B.

Table 23B. Comparison of the NOV23 protein sequences.

NOV23a	MGDVNQSVASDFILVGLFSHSGSRQLLFSLVAVMFVIGLLGNTVLLFLIRVDSRLHTPMY
NOV23b	MGDVNQSVASDFILVGLFSHSGSRQLLFSLVAVMFVIGLLGNTVLLFLIRVDSRLHTPMY
NOV23c	MGDVNQSVASDFILVGLFSHSGSRQLLFSLVAVMFVIGLLGNTVLLFLIRVDSRLHTPMY

```
NOV23a   FLLSQLSLFDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMSY
NOV23b   FLLSQLSLFDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMSY
NOV23c   FLLSQLSLFDIGCPMVTIPKMASDFLRGEGATSYGGGAAQIFFLTLMGVAEGVLLVLMSY

NOV23a   DRYVAVCQPLQYPVLMRRQVCLLMMGSSWVVGVLNASIQTSITLHFPYCASRIVDHFFCE
NOV23b   DRYVAVCQPLQYPVLMRRQVCLLMMGSSWVVGVLNASIQTSITLHFPYCASRIVDHFFCE
NOV23c   DRYVAVCQPLQYPVLMRRQVCLLMMGSSWVVGVLNASIQTSITLHFPYCASRIVDHFFCE

NOV23a   VPALLKLSCADTCAYEMALSTSGVLILMLPLSLIATSYGHVLQAVLSMRSEEARHKAVTT
NOV23b   VPALLKLSCADTCAYEMALSTSGVLILMLPLSLIATSYGHVLQAVLSMRSEEARHKAVTT
NOV23c   VPALLKLSCADTCAYEMALSTSGVLILMLPLSLIATSYGHVLQAVLSMRSEEARHKAVTT

NOV23a   CSSHITVVGLFYGAAVFMYMVPCAYHSPQQDNVVSLFYSLVTPTLNPLIYSLRNPEVWMA
NOV23b   CSSHITVVGLFYGAAVFMYMVPCAYHSPQQDNVVSLFYSLVTPTLNPLIYSLRNPEVWMA
NOV23c   CSSHITVVGLFYGAAVFMYMVPCAYHSPQQDNVVSLFYSLVTPTLNPLIYSLRNPEVWMA

NOV23a   LVKVLSRAGLRQMC
NOV23b   LVKVLSRAGLRQMC
NOV23c   LVKVLSRAGLRQMC

NOV23a   (SEQ ID NO:  306)
NOV23b   (SEQ ID NO:  308)
NOV23c   (SEQ ID NO:  310)
```

Further analysis of the NOV23a protein yielded the following properties shown in Table 23C.

Table 23C. Protein Sequence Properties NOV23a

SignalP analysis:	Cleavage site between residues 42 and 43

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 0;  neg.chg 2
      H-region:  length 12;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -1.30
      possible cleavage site: between 41 and 42

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   5
      INTEGRAL    Likelihood = -8.60   Transmembrane   33 -  49
      INTEGRAL    Likelihood = -7.17   Transmembrane  101 - 117
      INTEGRAL    Likelihood = -1.38   Transmembrane  140 - 156
      INTEGRAL    Likelihood = -4.09   Transmembrane  198 - 214
      INTEGRAL    Likelihood = -3.77   Transmembrane  245 - 261
      PERIPHERAL  Likelihood =  1.38 (at 62)
```

```
      ALOM score:  -8.60  (number of TMSs: 5)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 40
      Charge difference:  1.0   C( 1.5) - N( 0.5)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  0.74
      Hyd Moment(95):  7.13     G content:       1
      D/E content:     2        S/T content:     2
      Score: -7.54

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   5.4%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none
```

```
checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       11.1 %: Golgi
       11.1 %: vacuolar
       11.1 %: vesicles of secretory system
       11.1 %: mitochondrial

>> prediction for CG53390-02 is end (k=9)
```

A search of the NOV23a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 23D.

Table 23D. Geneseq Results for NOV23a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV23a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU85207	G-coupled olfactory receptor #68 - Homo sapiens, 314 aa. [WO200198526-A2, 27-DEC-2001]	1..314 1..314	314/314 (100%) 314/314 (100%)	e-179
AAU95551	Human olfactory and pheromone G protein-coupled receptor #38 - Homo sapiens, 314 aa. [WO200224726-A2, 28-MAR-2002]	1..314 1..314	314/314 (100%) 314/314 (100%)	e-179
ABJ04025	Human G-protein coupled receptor SEQ ID NO: 116 - Homo sapiens, 314 aa. [WO200255558-A2, 18-JUL-2002]	1..314 1..314	314/314 (100%) 314/314 (100%)	e-179
ABP95895	Human GPCR polypeptide SEQ ID NO 600 - Homo sapiens, 314 aa. [WO200216548-A2, 28-FEB-2002]	1..314 1..314	314/314 (100%) 314/314 (100%)	e-179
AAG71902	Human olfactory receptor polypeptide, SEQ ID NO: 1583 - Homo sapiens, 314 aa. [WO200127158-A2, 19-APR-2001]	1..314 1..314	314/314 (100%) 314/314 (100%)	e-179

In a BLAST search of public sequence databases, the NOV23a protein was found to have homology to the proteins shown in the BLASTP data in Table 23E.

Table 23E. Public BLASTP Results for NOV23a				
Protein Accession Number	**Protein/Organism/Length**	**NOV23a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8NG97	Seven transmembrane helix receptor - Homo sapiens (Human), 314 aa.	1..314 1..314	314/314 (100%) 314/314 (100%)	e-179
Q8VFG8	Olfactory receptor MOR282-1 - Mus musculus (Mouse), 308 aa.	1..301 1..301	225/301 (74%) 255/301 (83%)	e-126
Q96R25	Olfactory receptor - Homo sapiens (Human), 217 aa (fragment).	68..284 1..217	217/217 (100%) 217/217 (100%)	e-122
Q8VGD8	Olfactory receptor MOR281-1 - Mus musculus (Mouse), 315 aa.	5..312 7..313	173/308 (56%) 229/308 (74%)	2e-97
Q8VGD7	Olfactory receptor MOR277-1 - Mus musculus (Mouse), 317 aa.	5..307 6..308	170/304 (55%) 218/304 (70%)	3e-91

PFam analysis indicates that the NOV23a protein contains the domains shown in the Table 23F.

Table 23F. Domain Analysis of NOV23a			
Pfam Domain	**NOV23a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	41..290	46/276 (17%) 167/276 (61%)	5.5e-25

Example 24.

The NOV24 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 24A.

Table 24A. NOV24 Sequence Analysis		
NOV24a, CG53482-01 DNA Sequence	SEQ ID NO: 311	1008 bp
	ORF Start: ATG at 27	ORF Stop: TAG at 999

AGCTGGAGATCTGGAACTTCCACAGC**ATG**GAGCTCTGGAACTACCACAGCATGGAGCTCTGGAACTTC
ACCTTGGGAAGTGGCTTCATTTTGGTGGGGATTCTGAATGACAGTGGGTCTCCTGAACTGCTCTGTGC
TACAATTACAATCCTATACTTGTTGGCCCTGATCAGCAATGGCCTACTGCTCCTGGCTATCACCATGG
AAGCCCGGCTCCACATGCCCATGTACCTCCTGCTTGGGCAGCTCTCTCTCATGGACCTCCTGTTCACA
TCTGTTGTCACTCCCAAGGCCCTTGCGGACTTTCTGCGCAGAGAAAACACCATCTCCTTTGGAGGCTG
TGCCCTTCAGATGTTCCTGGCACTGACAATGGGTGGTGCTGAGGACCTCCTACTGGCCTTCATGGCCT

ATGACAGGTATGTGGCCATTTGTCATCCTCTGACATACATGACCCTCATGAGCTCAAGAGCCTGCTGG
CTCATGGTGGCCACGTCCTGGATCCTGGCATCCCTAAGTGCCCTAATATATACCGTGTATACCATGCA
CTATCCCTTCTGCAGGGCCCAGGAGATCAGGCATCTTCTCTGTGAGATCCCACACTTGCTGAAGTTGG
CCTGTGCTGATACCTCCAGATATGAGCTCATGGTATATGTGATGGGTGTGACCTTCCTGATTCCCTCT
CTTGCTGCTATACTGGCCTCCTATACACAAATTCTACTCACTGTGCTCCATATGCCATCAAATGAGGG
GAGGAAGAAAGCCCTTGTCACCTGCTCTTCCCACCTGACTGTGGTTGGGATGTTCTATGGAGCTGCCA
CATTCATGTATGTCTTGCCCAGTTCCTTCCACAGCACCAGACAAGACAACATCATCTCTGTTTTCTAC
ACAATTGTCACTCCAGCCCTGAATCCACTCATCTACAGCCTGAGGAATAAGGAGGTCATGCGGGCCTT
GAGGAGGGTCCTGGGAAAATACATGCTGCCAGCACACTCCACGCTCTAGGGAAGGA

NOV24a, CG53482-01 Protein Sequence	SEQ ID NO: 312	324 aa	MW at 36344.8kD

MELWNYHSMELWNFTLGSGFILVGILNDSGSPELLCATITILYLLALISNGLLLLAITMEARLHMPMY
LLLGQLSLMDLLFTSVVTPKALADFLRRENTISFGGCALQMFLALTMGGAEDLLLAFMAYDRYVAICH
PLTYMTLMSSRACWLMVATSWILASLSALIYTVYTMHYPFCRAQEIRHLLCEIPHLLKLACADTSRYE
LMVYVMGVTFLIPSLAAILASYTQILLTVLHMPSNEGRKKALVTCSSHLTVVGMFYGAATFMYVLPSS
FHSTRQDNIISVFYTIVTPALNPLIYSLRNKEVMRALRRVLGKYMLPAHSTL

NOV24b, CG53482-02 DNA Sequence	SEQ ID NO: 313	1050 bp
	ORF Start: ATG at 72	ORF Stop: TAG at 1020

AAACTAGAGTTCATCTTAGCAAAAATTCATGAAGTATCCATCTTGTTCTAGGTGATGAAAGAAACCAC
AGCATGGAGCTCTGGAACTTCACCTTGGGAAGTGGCTTCATTTTGGTGGGGATTCTGAATGACAGTGG
GTCTCCTGAACTGCTCTGTGCTACAATTACAATCCTATACTTGTTGGCCCTGATCAGCAATGGCCTAC
TGCTCCTGGCTATCACCATGGAAGCCCGGCTCCACATGCCCATGTACCTCCTGCTTGGGCAGCTCTCT
CTCATGGACCTCCTGTTCACATCTGTTGTCACTCCCAAGGCCCTTGCGGACTTTCTGCGCAGAGAAAA
CACCATCTCCTTTGGAGGCTGTGCCCTTCAGATGTTCCTGGCACTGACAATGGGTGGTGCTGAGGACC
TCCTACTGGCCTTCATGGCCTATGACAGGTATGTGGCCATTTGTCATCCTCTGACATACATGACCCTC
ATGAGCTCAAGAGCCTGCTGGCTCATGGTGGCCACGTCCTGGATCCTGGCATCCCTAAGTGCCCTAAT
ATATACCGTGTATACCATGCACTATCCCTTCTGCAGGGCCCAGGAGATCAGGCATCTTCTCTGTGAGA
TCCCACACTTGCTGAAGGTGGCCTGTGCTGATACCTCCAGATATGAGCTCATGGTATATGTGATGGGT
GTGACCTTCCTGATTCCCTCTCTTGCTGCTATACTGGCCTCCTATACACAAATTCTACTCACTGTGCT
CCATATGCCATCAAATGAGGGGAGGAAGAAAGCCCTTGTCACCTGCTCTTCCCACCTGACTGTGGTTG
GGATGTTCTATGGAGCTGCCACATTCATGTATGTCTTGCCCAGTTCCTTCCACAGCACCAGACAAGAC
AACATCATCTCTGTTTTCTACACAATTGTCACTCCAGCCCTGAATCCACTCATCTACAGCCTGAGGAA
TAAGGAGGTCATGCGGGCCTTGAGGAGGGTCCTGGGAAAATACATGCTGCCAGCACACTCCACGCTCT
AGGGAAGGATCATGGCTAGCTTCCAGAATT

NOV24b, CG53482-02 Protein Sequence	SEQ ID NO: 314	316 aa	MW at 35269.6kD

MELWNFTLGSGFILVGILNDSGSPELLCATITILYLLALISNGLLLLAITMEARLHMPMYLLLGQLSL
MDLLFTSVVTPKALADFLRRENTISFGGCALQMFLALTMGGAEDLLLAFMAYDRYVAICHPLTYMTLM
SSRACWLMVATSWILASLSALIYTVYTMHYPFCRAQEIRHLLCEIPHLLKVACADTSRYELMVYVMGV
TFLIPSLAAILASYTQILLTVLHMPSNEGRKKALVTCSSHLTVVGMFYGAATFMYVLPSSFHSTRQDN
IISVFYTIVTPALNPLIYSLRNKEVMRALRRVLGKYMLPAHSTL

NOV24c, CG53482-03 DNA Sequence	SEQ ID NO: 315	1010 bp
	ORF Start: at 4	ORF Stop: TAG at 1000

TAGCTGGAGATCTGGAACTTCCACAGCATGGAGCTCTGGAACTACCACAGCATGGAGCTCTGGAACTT
CACCTTGGGAAGTGGCTTCATTTTGGTGGGGATTCTGAATGACAGTGGGTCTCCTGAACTGCTCTGTG
CTACAATTACAATCCTATACTTGTTGGCCCTGATCAGCAATGGCCTACTGCTCCTGGCTATCACCATG
GAAGCCCGGCTCCACATGCCCATGTACCTCCTGCTTGGGCAGCTCTCTCTCATGGACCTCCTGTTCAC
ATCTGTTGTCACTCCCAAGGCCCTTGCGGACTTTCTGCGCAGAGAAAACACCATCTCCTTTGGAGGCT
GTGCCCTTCAGATGTTCCTGGCACTGACAATGGGTGGTGCTGAGGACCTCCTACTGGCCTTCATGGCC
TATGACAGGTATGTGGCCATTTGTCATCCTCTGACATACATGACCCTCATGAGCTCAAGAGCCTGCTG
GCTCATGGTGGCCACGTCCTGGATCCTGGCATCCCTAAGTGCCCTAATATATACCGTGTATACCATGC
ACTATCCCTTCTGCAGGGCCCAGGAGATCAGGCATCTTCTCTGTGAGATCCCACACTTGCTGAAGTTG
GCCTGTGCTGATACCTCCAGATATGAGCTCATGGTATATGTGATGGGTGTGACCTTCCTGATTCCCTC
TCTTGCTGCTATACTGGCCTCCTATACACAAATTCTACTCACTGTGCTCCATATGCCATCAAATGAGG

GGAGGAAGAAAGCCCTTGTCACCTGCTCTTCCCACCTGACTGTGGTTGGGATGTTCTATGGAGCTGCC
ACATTCATGTATGTCTTGCCCAGTTCCTTCCACAGCACCAGACAAGACAACATCATCTCTGTTTTCTA
CACAATTGTCACTCCAGCCCTGAATCCACTCATCTACAGCCTGAGGAATAAGGAGGTCATGCGGGCCT
TGAGGAGGGTCCTGGGAAAATACATGCTGCCAGCACACTCCACGCTCTAGGGAAGGAA

NOV24c, CG53482-03 Protein Sequence	SEQ ID NO: 316	332 aa	MW at 37371.9kD

LEIWNFHSMELWNYHSMELWNFTLGSGFILVGILNDSGSPELLCATITILYLLALISNGLLLLAITME
ARLHMPMYLLLGQLSLMDLLFTSVVTPKALADFLRRENTISFGGCALQMFLALTMGGAEDLLLAFMAY
DRYVAICHPLTYMTLMSSRACWLMVATSWILASLSALIYTVYTMHYPFCRAQEIRHLLCEIPHLLKLA
CADTSRYELMVYVMGVTFLIPSLAAILASYTQILLTVLHMPSNEGRKKALVTCSSHLTVVGMFYGAAT
FMYVLPSSFHSTRQDNIISVFYTIVTPALNPLIYSLRNKEVMRALRRVLGKYMLPAHSTL

NOV24d, SNP13373787 of CG53482-01, DNA Sequence	SEQ ID NO: 317	1008 bp
	ORF Start: ATG at 27	ORF Stop: TAG at 999
	SNP Pos: 430	SNP Change: G to A

AGCTGGAGATCTGGAACTTCCACAGCATGGAGCTCTGGAACTACCACAGCATGGAGCTCTGGAACTTC
ACCTTGGGAAGTGGCTTCATTTTGGTGGGGATTCTGAATGACAGTGGGTCTCCTGAACTGCTCTGTGC
TACAATTACAATCCTATACTTGTTGGCCCTGATCAGCAATGGCCTACTGCTCCTGGCTATCACCATGG
AAGCCCGGCTCCACATGCCCATGTACCTCCTGCTTGGGCAGCTCTCTCTCATGGACCTCCTGTTCACA
TCTGTTGTCACTCCCAAGGCCCTTGCGGACTTTCTGCGCAGAGAAAACACCATCTCCTTTGGAGGCTG
TGCCCTTCAGATGTTCCTGGCACTGACAATGGGTGGTGCTGAGGACCTCCTACTGGCCTTCATGGCCT
ATGACAGGTATGTGGCCATTTATCATCCTCTGACATACATGACCCTCATGAGCTCAAGAGCCTGCTGG
CTCATGGTGGCCACGTCCTGGATCCTGGCATCCCTAAGTGCCCTAATATATACCGTGTATACCATGCA
CTATCCCTTCTGCAGGGCCCAGGAGATCAGGCATCTTCTCTGTGAGATCCCACACTTGCTGAAGTTGG
CCTGTGCTGATACCTCCAGATATGAGCTCATGGTATATGTGATGGGTGTGACCTTCCTGATTCCCTCT
CTTGCTGCTATACTGGCCTCCTATACACAAATTCTACTCACTGTGCTCCATATGCCATCAAATGAGGG
GAGGAAGAAAGCCCTTGTCACCTGCTCTTCCCACCTGACTGTGGTTGGGATGTTCTATGGAGCTGCCA
CATTCATGTATGTCTTGCCCAGTTCCTTCCACAGCACCAGACAAGACAACATCATCTCTGTTTTCTAC
ACAATTGTCACTCCAGCCCTGAATCCACTCATCTACAGCCTGAGGAATAAGGAGGTCATGCGGGCCTT
GAGGAGGGTCCTGGGAAAATACATGCTGCCAGCACACTCCACGCTCTAGGGAAGGA

NOV24d, SNP13373787 of CG53482-01, Protein Sequence	SEQ ID NO: 318	324 aa	MW at 36404.8kD
	SNP Pos: 135		SNP Change: Cys to Tyr

MELWNYHSMELWNFTLGSGFILVGILNDSGSPELLCATITILYLLALISNGLLLLAITMEARLHMPMY
LLLGQLSLMDLLFTSVVTPKALADFLRRENTISFGGCALQMFLALTMGGAEDLLLAFMAYDRYVAIYH
PLTYMTLMSSRACWLMVATSWILASLSALIYTVYTMHYPFCRAQEIRHLLCEIPHLLKLACADTSRYE
LMVYVMGVTFLIPSLAAILASYTQILLTVLHMPSNEGRKKALVTCSSHLTVVGMFYGAATFMYVLPSS
FHSTRQDNIISVFYTIVTPALNPLIYSLRNKEVMRALRRVLGKYMLPAHSTL

NOV24e, SNP13373786 of CG53482-01, DNA Sequence	SEQ ID NO: 319	1008 bp
	ORF Start: ATG at 27	ORF Stop: TAG at 999
	SNP Pos: 442	SNP Change: C to A

AGCTGGAGATCTGGAACTTCCACAGCATGGAGCTCTGGAACTACCACAGCATGGAGCTCTGGAACTTC
ACCTTGGGAAGTGGCTTCATTTTGGTGGGGATTCTGAATGACAGTGGGTCTCCTGAACTGCTCTGTGC
TACAATTACAATCCTATACTTGTTGGCCCTGATCAGCAATGGCCTACTGCTCCTGGCTATCACCATGG
AAGCCCGGCTCCACATGCCCATGTACCTCCTGCTTGGGCAGCTCTCTCTCATGGACCTCCTGTTCACA
TCTGTTGTCACTCCCAAGGCCCTTGCGGACTTTCTGCGCAGAGAAAACACCATCTCCTTTGGAGGCTG
TGCCCTTCAGATGTTCCTGGCACTGACAATGGGTGGTGCTGAGGACCTCCTACTGGCCTTCATGGCCT
ATGACAGGTATGTGGCCATTTGTCATCCTCTGAAATACATGACCCTCATGAGCTCAAGAGCCTGCTGG
CTCATGGTGGCCACGTCCTGGATCCTGGCATCCCTAAGTGCCCTAATATATACCGTGTATACCATGCA
CTATCCCTTCTGCAGGGCCCAGGAGATCAGGCATCTTCTCTGTGAGATCCCACACTTGCTGAAGTTGG
CCTGTGCTGATACCTCCAGATATGAGCTCATGGTATATGTGATGGGTGTGACCTTCCTGATTCCCTCT
CTTGCTGCTATACTGGCCTCCTATACACAAATTCTACTCACTGTGCTCCATATGCCATCAAATGAGGG
GAGGAAGAAAGCCCTTGTCACCTGCTCTTCCCACCTGACTGTGGTTGGGATGTTCTATGGAGCTGCCA
CATTCATGTATGTCTTGCCCAGTTCCTTCCACAGCACCAGACAAGACAACATCATCTCTGTTTTCTAC
ACAATTGTCACTCCAGCCCTGAATCCACTCATCTACAGCCTGAGGAATAAGGAGGTCATGCGGGCCTT

GAGGAGGGTCCTGGGAAAATACATGCTGCCAGCACACTCCACGCTCTAGGGAAGGA			
NOV24e, SNP13373786 of CG53482-01, Protein Sequence	SEQ ID NO: 320	324 aa	MW at 36371.9kD
	SNP Pos: 139		SNP Change: Thr to Lys
MELWNYHSMELWNFTLGSGFILVGILNDSGSPELLCATITILYLLALISNGLLLLAITMEARLHMPMY LLLGQLSLMDLLFTSVVTPKALADFLRRENTISFGGCALQMFLALTMGGAEDLLLAFMAYDRYVAICH PLKYMTLMSSRACWLMVATSWILASLSALIYTVYTMHYPFCRAQEIRHLLCEIPHLLKLACADTSRYE LMVYVMGVTFLIPSLAAILASYTQILLTVLHMPSNEGRKKALVTCSSHLTVVGMFYGAATFMYVLPSS FHSTRQDNIISVFYTIVTPALNPLIYSLRNKEVMRALRRVLGKYMLPAHSTL			
NOV24f, SNP13373785 of CG53482-01, DNA Sequence	SEQ ID NO: 321		1008 bp
	ORF Start: ATG at 27		ORF Stop: TAG at 999
	SNP Pos: 754		SNP Change: A to G
AGCTGGAGATCTGGAACTTCCACAGCATGGAGCTCTGGAACTACCACAGCATGGAGCTCTGGAACTTC ACCTTGGGAAGTGGCTTCATTTTGGTGGGGATTCTGAATGACAGTGGGTCTCCTGAACTGCTCTGTGC TACAATTACAATCCTATACTTGTTGGCCCTGATCAGCAATGGCCTACTGCTCCTGGCTATCACCATGG AAGCCCGGCTCCACATGCCCATGTACCTCCTGCTTGGGCAGCTCTCTCTCATGGACCTCCTGTTCACA TCTGTTGTCACTCCCAAGGCCCTTGCGGACTTTCTGCGCAGAGAAAACACCATCTCCTTTGGAGGCTG TGCCCTTCAGATGTTCCTGGCACTGACAATGGGTGGTGCTGAGGACCTCCTACTGGCCTTCATGGCCT ATGACAGGTATGTGGCCATTTGTCATCCTCTGACATACATGACCCTCATGAGCTCAAGAGCCTGCTGG CTCATGGTGGCCACGTCCTGGATCCTGGCATCCCTAAGTGCCCTAATATATACCGTGTATACCATGCA CTATCCCTTCTGCAGGGCCCAGGAGATCAGGCATCTTCTCTGTGAGATCCCACACTTGCTGAAGTTGG CCTGTGCTGATACCTCCAGATATGAGCTCATGGTATATGTGATGGGTGTGACCTTCCTGATTCCCTCT CTTGCTGCTATACTGGCCTCCTATACACAAATTCTACTCACTGTGCTCCATATGCCATCAAATGAGGG GAGGAGGAAAGCCCTTGTCACCTGCTCTTCCCACCTGACTGTGGTTGGGATGTTCTATGGAGCTGCCA CATTCATGTATGTCTTGCCCAGTTCCTTCCACAGCACCAGACAAGACAACATCATCTCTGTTTTCTAC ACAATTGTCACTCCAGCCCTGAATCCACTCATCTACAGCCTGAGGAATAAGGAGGTCATGCGGGCCTT GAGGAGGGTCCTGGGAAAATACATGCTGCCAGCACACTCCACGCTCTAGGGAAGGA			
NOV24f, SNP13373785 of CG53482-01, Protein Sequence	SEQ ID NO: 322	324 aa	MW at 36372.8kD
	SNP Pos: 243		SNP Change: Lys to Arg
MELWNYHSMELWNFTLGSGFILVGILNDSGSPELLCATITILYLLALISNGLLLLAITMEARLHMPMY LLLGQLSLMDLLFTSVVTPKALADFLRRENTISFGGCALQMFLALTMGGAEDLLLAFMAYDRYVAICH PLTYMTLMSSRACWLMVATSWILASLSALIYTVYTMHYPFCRAQEIRHLLCEIPHLLKLACADTSRYE LMVYVMGVTFLIPSLAAILASYTQILLTVLHMPSNEGRRKALVTCSSHLTVVGMFYGAATFMYVLPSS FHSTRQDNIISVFYTIVTPALNPLIYSLRNKEVMRALRRVLGKYMLPAHSTL			

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 24B.

Table 24B. Comparison of the NOV24 protein sequences.

```
NOV24a  --------MELWNYHSMELWNFTLGSGFILVGILNDSGSPELLCATITILYLLALISNGL
NOV24b  ----------------MELWNFTLGSGFILVGILNDSGSPELLCATITILYLLALISNGL
NOV24c  LEIWNFHSMELWNYHSMELWNFTLGSGFILVGILNDSGSPELLCATITILYLLALISNGL

NOV24a  LLLAITMEARLHMPMYLLLGQLSLMDLLFTSVVTPKALADFLRRENTISFGGCALQMFLA
NOV24b  LLLAITMEARLHMPMYLLLGQLSLMDLLFTSVVTPKALADFLRRENTISFGGCALQMFLA
NOV24c  LLLAITMEARLHMPMYLLLGQLSLMDLLFTSVVTPKALADFLRRENTISFGGCALQMFLA

NOV24a  LTMGGAEDLLLAFMAYDRYVAICHPLTYMTLMSSRACWLMVATSWILASLSALIYTVYTM
NOV24b  LTMGGAEDLLLAFMAYDRYVAICHPLTYMTLMSSRACWLMVATSWILASLSALIYTVYTM
NOV24c  LTMGGAEDLLLAFMAYDRYVAICHPLTYMTLMSSRACWLMVATSWILASLSALIYTVYTM

NOV24a  HYPFCRAQEIRHLLCEIPHLLKLACADTSRYELMVYVMGVTFLIPSLAAILASYTQILLT
```

```
NOV24b  HYPFCRAQEIRHLLCEIPHLLKVACADTSRYELMVYVMGVTFLIPSLAAILASYTQILLT
NOV24c  HYPFCRAQEIRHLLCEIPHLLKLACADTSRYELMVYVMGVTFLIPSLAAILASYTQILLT

NOV24a  VLHMPSNEGRKKALVTCSSHLTVVGMFYGAATFMYVLPSSFHSTRQDNIISVFYTIVTPA
NOV24b  VLHMPSNEGRKKALVTCSSHLTVVGMFYGAATFMYVLPSSFHSTRQDNIISVFYTIVTPA
NOV24c  VLHMPSNEGRKKALVTCSSHLTVVGMFYGAATFMYVLPSSFHSTRQDNIISVFYTIVTPA

NOV24a  LNPLIYSLRNKEVMRALRRVLGKYMLPAHSTL
NOV24b  LNPLIYSLRNKEVMRALRRVLGKYMLPAHSTL
NOV24c  LNPLIYSLRNKEVMRALRRVLGKYMLPAHSTL

NOV24a  (SEQ ID NO:  312)
NOV24b  (SEQ ID NO:  314)
NOV24c  (SEQ ID NO:  316)
```

Further analysis of the NOV24a protein yielded the following properties shown in Table 24C.

Table 24C. Protein Sequence Properties NOV24a

SignalP analysis:	Cleavage site between residues 62 and 63

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 10;  pos.chg 0;  neg.chg 2
      H-region:  length 17;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.29
      possible cleavage site: between 56 and 57

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood = -7.75   Transmembrane   41 -  57
      INTEGRAL    Likelihood = -1.65   Transmembrane   69 -  85
      INTEGRAL    Likelihood = -2.07   Transmembrane  150 - 166
      INTEGRAL    Likelihood = -5.10   Transmembrane  207 - 223
      INTEGRAL    Likelihood = -0.80   Transmembrane  253 - 269
      INTEGRAL    Likelihood = -0.80   Transmembrane  281 - 297
      PERIPHERAL  Likelihood =  0.85 (at 109)
      ALOM score:  -7.75  (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 48
      Charge difference:  2.5   C( 0.5) - N(-2.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b
```

```
MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  4.76
      Hyd Moment(95):  6.14    G content:       0
      D/E content:     2       S/T content:     1
      Score: -6.93

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   6.2%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues
```

```
--------------------------

Final Results (k = 9/23):

       44.4 %: endoplasmic reticulum
       22.2 %: vacuolar
       11.1 %: Golgi
       11.1 %: vesicles of secretory system
       11.1 %: mitochondrial

>> prediction for CG53482-01 is end (k=9)
```

A search of the NOV24a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 24D.

Table 24D. Geneseq Results for NOV24a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV24a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU97928	Novel odourant receptor NOV7 protein - Unidentified, 324 aa. [WO200236632-A2, 10-MAY-2002]	1..324 1..324	324/324 (100%) 324/324 (100%)	0.0
AAU97927	Novel odourant receptor NOV6 protein - Unidentified, 324 aa. [WO200236632-A2, 10-MAY-2002]	1..324 1..324	324/324 (100%) 324/324 (100%)	0.0
AAU07088	Human odorant receptor (OR) polypeptide #5 - Homo sapiens, 324 aa. [WO200157215-A2, 09-AUG-2001]	1..324 1..324	324/324 (100%) 324/324 (100%)	0.0
AAU85234	G-coupled olfactory receptor #95 - Homo sapiens, 316 aa. [WO200198526-A2, 27-DEC-2001]	9..324 1..316	315/316 (99%) 316/316 (99%)	e-180
AAU95574	Human olfactory and pheromone G protein-coupled receptor #61 - Homo sapiens, 316 aa. [WO200224726-A2, 28-MAR-2002]	9..324 1..316	315/316 (99%) 316/316 (99%)	e-180

In a BLAST search of public sequence databases, the NOV24a protein was found to have homology to the proteins shown in the BLASTP data in Table 24E.

Table 24E. Public BLASTP Results for NOV24a				
Protein Accession Number	**Protein/Organism/Length**	**NOV24a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
CAD57894	Sequence 11 from Patent WO0236632 - Homo sapiens (Human), 324 aa.	1..324 1..324	324/324 (100%) 324/324 (100%)	0.0
Q9H205	Olfactory receptor 2AG1 (HT3) - Homo sapiens (Human), 316 aa.	9..324 1..316	315/316 (99%) 316/316 (99%)	e-179
Q8VGU0	Olfactory receptor MOR283-2 - Mus musculus (Mouse), 316 aa.	9..324 1..316	264/316 (83%) 284/316 (89%)	e-149
Q9EPF7	T2 olfactory receptor - Mus musculus (Mouse), 316 aa.	9..324 1..316	262/316 (82%) 284/316 (88%)	e-149
Q8VFM5	Olfactory receptor MOR283-4 - Mus musculus (Mouse), 316 aa.	9..323 1..315	253/315 (80%) 280/315 (88%)	e-145

PFam analysis indicates that the NOV24a protein contains the domains shown in the Table 24F.

Table 24F. Domain Analysis of NOV24a			
Pfam Domain	**NOV24a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	49..298	62/277 (22%) 173/277 (62%)	1.3e-33

Example 25.

The NOV25 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 25A.

Table 25A. NOV25 Sequence Analysis		
NOV25a, CG53530-03 DNA Sequence	SEQ ID NO: 323	1016 bp
	ORF Start: ATG at 7	ORF Stop: TAG at 961

TACCAG**ATG**AATCCAGCAAATCATTCCCAGGTGGCAGGATTTGTTCTACTGGGGCTCTCTCAGGTTTG
GGAGCTTCGGTTTGTTTTCTTCACTGTTTTCTCTGCTGTGTATTTTATGACTGTAGTGGGAAACCTTC
TTATTGTGGTCATAGTGACCTCCGACCCACACCTGCACACAACCATGTATTTTCTCTTGGGCAATCTT

TCTTTCCTGGACTTTTGCTACTCTTCCATCACAGCACCTAGGATGCTGGTTGACTTGCTCTCAGGCAA
CCCTACCATTTCCTTTGGTGGATGCCTGACTCAACTCTTCTTCTTCCACTTCATTGGAGGCATCAAGA
TCTTCCTGCTGACTGTCATGGCGTATGACCGCTACATTGCCATTTCCCAGCCCCTGCACTACACGCTC
ATTATGAATCAGACTGTCTGTGCACTCCTTATGGCAGCCTCCTGGGTGGGGGGCTTCATCCACTCCAT
AGTACAGATTGCATTGACTATCCAGCTGCCATTCTGTGGGCCTGACAAGCTGGACAACTTTTATTGTG
ATGTGCCTCAGCTGATCAAATTGGCCTGCACAGATACCTTTGTCTTAGAGCTTTTAATGGTGTCTAAC
AATGGCCTGGTGACCCTGATGTGTTTTCTGGTGCTTCTGGGATCGTACACAGCACTGCTAGTCATGCT
CCGAAGCCACTCACGGGAGGGCCGCAGCAAGGCCCTGTCTACCTGTGCCTCTCACATTGCTGTGGTGA
CCTTAATCTTTGTGCCTTGCATCTACGTCTATACAAGGCCTTTTCGGACATTCCCCATGGACAAGGCC
GTCTCTGTGCTATACACAATTGTCACCCCCATGCTGAATCCTGCCATCTATACCCTGAGAAACAAGGA
AGTGATCATGGCCATGAAGAAGCTGTGGAGGAGGAAAAAGGACCCTATTGGTCCCCTGGAGCACAGAC
CCTTACATTAGCAGAGGCAGTGACCTGAGAATCTGAAAGATGCTACAGGGTATTAGCAGAGGCA

NOV25a, CG53530-03 Protein Sequence	SEQ ID NO: 324	318 aa	MW at 35770.3kD

MNPANHSQVAGFVLLGLSQVWELRFVFFTVFSAVYFMTVVGNLLIVVIVTSDPHLHTTMYFLLGNLSF
LDFCYSSITAPRMLVDLLSGNPTISFGGCLTQLFFFHFIGGIKIFLLTVMAYDRYIAISQPLHYTLIM
NQTVCALLMAASWVGGFIHSIVQIALTIQLPFCGPDKLDNFYCDVPQLIKLACTDTFVLELLMVSNNG
LVTLMCFLVLLGSYTALLVMLRSHSREGRSKALSTCASHIAVVTLIFVPCIYVYTRPFRTFPMDKAVS
VLYTIVTPMLNPAIYTLRNKEVIMAMKKLWRRKKDPIGPLEHRPLH

NOV25b, CG53530-01 DNA Sequence	SEQ ID NO: 325	1077 bp
	ORF Start: ATG at 31	ORF Stop: TAG at 985

CAGGTTCATTGACAAGGTCATACCAACCAGATGAATCCAGCAAATCATTCCCAGGTGGCAGGATTTGT
TCTACTGGGGCTCTCTCAGGTTTGGGAGCTTCGGTTTGTTTTCTTCACTGTTTTCTCTGCTGTGTATT
TTATGACTGTAGTGGGAAACCTTCTTATTGTGGTCATAGTGACCTCCGACCCACACCTGCACACAACC
ATGTATTTTCTCTTGGGCAATCTTTCTTTCCTGGACTTTTGCTACTCTTCCATCACAGCACCTAGGAT
GCTGGTTGACTTGCTCTCAGGCAACCCTACCATTTCCTTTGGTGGATGCCTGACTCAACTCTTCTTCT
TCCACTTCATTGGAGGCATCAAGATCTTCCTGCTGACTGTCATGGCGTATGACCGCTACATTGCCATT
TCCCAGCCCCTGCACTACACGCTCATTATGAATCAGACTGTCTGTGCACTCCTTATGGCAGCCTCCTG
GGTGGGGGGCTTCATCCACTCCATAGTACAGATTGCATTGACTATCCAGCTGCCATTCTGTGGGCCTG
ACAAGCTGGACAACTTTTATTGTGATGTGCCTCAGCTGATCAAATTGGCCTGCACAGATACCTTTGTC
TTAGAGCTTTTAATGGTGTCTAACAATGGCCTGGTGACCCTGATGTGTTTTCTGGTGCTTCTGGGATC
GTACACAGCACTGCTAGTCATGCTCCGAAGCCACTCACGGGAGGGCCGCAGCAAGGCCCTGTCTACCT
GTGCCTCTCACATTGCTGTGGTGACCTTAATCTTTGTGCCTTGCATCTACGTCTATACAAGGCCTTTT
CGGACATTCCCCATGGACAAGGCCGTCTCTGTGCTATACACAATTGTCACCCCCATGCTGAATCCTGC
CATCTATACCCTGAGAAACAAGGAAGTGATCATGGCCATGAAGAAGCTGTGGAGGAGGAAAAAGGACC
CTATTGGTCCCCTGGAGCACAGACCCTTACATTAGCAGAGGCAGTGACCTGAGAATCTGAAAGATGCT
ACAGGGTATTAGCAGAGGCAGTGACCTGAGAATCTGAAAGATGCTACAGGGTATTAG

NOV25b, CG53530-01 Protein Sequence	SEQ ID NO: 326	318 aa	MW at 35770.3kD

MNPANHSQVAGFVLLGLSQVWELRFVFFTVFSAVYFMTVVGNLLIVVIVTSDPHLHTTMYFLLGNLSF
LDFCYSSITAPRMLVDLLSGNPTISFGGCLTQLFFFHFIGGIKIFLLTVMAYDRYIAISQPLHYTLIM
NQTVCALLMAASWVGGFIHSIVQIALTIQLPFCGPDKLDNFYCDVPQLIKLACTDTFVLELLMVSNNG
LVTLMCFLVLLGSYTALLVMLRSHSREGRSKALSTCASHIAVVTLIFVPCIYVYTRPFRTFPMDKAVS
VLYTIVTPMLNPAIYTLRNKEVIMAMKKLWRRKKDPIGPLEHRPLH

NOV25c, CG53530-02 DNA Sequence	SEQ ID NO: 327	1077 bp
	ORF Start: ATG at 31	ORF Stop: TAG at 985

CAGGTTCATTGACAAGGTCATACCAACCAGATGAATCCAGCAAATCATTCCCAGGTGGCAGGATTTGT
TCTACTGGGGCTCTCTCAGGTTTGGGAGCTTCGGTTTGTTTTCTTCACTGTTTTCTCTGCTGTGTATT
TTATGACTGTAGTGGGAAACCTTCTTATTGTGGTCATAGTGACCTCCGACCCACACCTGCACACAACC
ATGTATTTTCTCTTGGGCAATCTTTCTTTCCTGGACTTTTGCTACTCTTCCATCACAGCACCTAGGAT
GCTGGTTGACTTGCTCTCAGGCAACCCTACCATTTCCTTTGGTGGATGCCTGACTCAACTCTTCTTCT
TCCACTTCATTGGAGGCATCAAGATCTTCCTGCTGACTGTCATGGCGTATGACCGCTACATTGCCATT
TCCCAGCCCCTGCACTACACGCTCATTATGAATCAGACTGTCTGTGCACTCCTTATGGCAGCCTCCTG
GGTGGGGGGCTTCATCCACTCCATAGTACAGATTGCATTGACTATCCAGCTGCCATTCTGTGGGCCTG

ACAAGCTGGACAACTTTTATTGTGATGTGCCTCAGCTGATCAAATTGGCCTGCACAGATACCTTTGTC
TTAGAGCTTTTAATGGTGTCTAACAATGGCCTGGTGACCCTGATGTGTTTTCTGGTGCTTCTGGGATC
GTACACAGCACTGCTAGTCATGCTCCGAAGCCACTCACGGGAGGGCCGCAGCAAGGCCCTGTCTACCT
GTGCCTCTCACATTGCTGTGGTGACCTTAATCTTTGTGCCTTGCATCTACGTCTATACAAGGCCTTTT
CGGACATTCCCCATGGACAAGGCCGTCTCTGTGCTATACACAATTGTCACCCCCATGCTGAATCCTGC
CATCTATACCCTGAGAAACAAGGAAGTGATCATGGCCATGAAGAAGCTGTGGAGGAGGAAAAAGGACC
CTATTGGTCCCCTGGAGCACAGACCCTTACATTAGCAGAGGCAGTGACCTGAGAATCTGAAAGATGCT
ACAGGGTATTAGCAGAGGCAGTGACCTGAGAATCTGAAAGATGCTACAGGGTATTAG

NOV25c, CG53530-02 Protein Sequence	SEQ ID NO: 328	318 aa	MW at 35770.3kD

MNPANHSQVAGFVLLGLSQVWELRFVFFTVFSAVYFMTVVGNLLIVVIVTSDPHLHTTMYFLLGNLSF
LDFCYSSITAPRMLVDLLSGNPTISFGGCLTQLFFFHFIGGIKIFLLTVMAYDRYIAISQPLHYTLIM
NQTVCALLMAASWVGGFIHSIVQIALTIQLPFCGPDKLDNFYCDVPQLIKLACTDTFVLELLMVSNNG
LVTLMCFLVLLGSYTALLVMLRSHSREGRSKALSTCASHIAVVTLIFVPCIYVYTRPFRTFPMDKAVS
VLYTIVTPMLNPAIYTLRNKEVIMAMKKLWRRKKDPIGPLEHRPLH

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 25B.

Table 25B. Comparison of the NOV25 protein sequences.

```
NOV25a  MNPANHSQVAGFVLLGLSQVWELRFVFFTVFSAVYFMTVVGNLLIVVIVTSDPHLHTTMY
NOV25b  MNPANHSQVAGFVLLGLSQVWELRFVFFTVFSAVYFMTVVGNLLIVVIVTSDPHLHTTMY
NOV25c  MNPANHSQVAGFVLLGLSQVWELRFVFFTVFSAVYFMTVVGNLLIVVIVTSDPHLHTTMY

NOV25a  FLLGNLSFLDFCYSSITAPRMLVDLLSGNPTISFGGCLTQLFFFHFIGGIKIFLLTVMAY
NOV25b  FLLGNLSFLDFCYSSITAPRMLVDLLSGNPTISFGGCLTQLFFFHFIGGIKIFLLTVMAY
NOV25c  FLLGNLSFLDFCYSSITAPRMLVDLLSGNPTISFGGCLTQLFFFHFIGGIKIFLLTVMAY

NOV25a  DRYIAISQPLHYTLIMNQTVCALLMAASWVGGFIHSIVQIALTIQLPFCGPDKLDNFYCD
NOV25b  DRYIAISQPLHYTLIMNQTVCALLMAASWVGGFIHSIVQIALTIQLPFCGPDKLDNFYCD
NOV25c  DRYIAISQPLHYTLIMNQTVCALLMAASWVGGFIHSIVQIALTIQLPFCGPDKLDNFYCD

NOV25a  VPQLIKLACTDTFVLELLMVSNNGLVTLMCFLVLLGSYTALLVMLRSHSREGRSKALSTC
NOV25b  VPQLIKLACTDTFVLELLMVSNNGLVTLMCFLVLLGSYTALLVMLRSHSREGRSKALSTC
NOV25c  VPQLIKLACTDTFVLELLMVSNNGLVTLMCFLVLLGSYTALLVMLRSHSREGRSKALSTC

NOV25a  ASHIAVVTLIFVPCIYVYTRPFRTFPMDKAVSVLYTIVTPMLNPAIYTLRNKEVIMAMKK
NOV25b  ASHIAVVTLIFVPCIYVYTRPFRTFPMDKAVSVLYTIVTPMLNPAIYTLRNKEVIMAMKK
NOV25c  ASHIAVVTLIFVPCIYVYTRPFRTFPMDKAVSVLYTIVTPMLNPAIYTLRNKEVIMAMKK

NOV25a  LWRRKKDPIGPLEHRPLH
NOV25b  LWRRKKDPIGPLEHRPLH
NOV25c  LWRRKKDPIGPLEHRPLH

NOV25a  (SEQ ID NO:  324)
NOV25b  (SEQ ID NO:  326)
NOV25c  (SEQ ID NO:  328)
```

Further analysis of the NOV25a protein yielded the following properties shown in Table 25C.

Table 25C. Protein Sequence Properties NOV25a

SignalP analysis:	Cleavage site between residues 52 and 53

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 21;  peak value   9.57
      PSG score:   5.17

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -5.51
      possible cleavage site: between 32 and 33

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood = -8.23   Transmembrane   33 -  49
      INTEGRAL    Likelihood = -4.04   Transmembrane  101 - 117
      INTEGRAL    Likelihood = -0.90   Transmembrane  142 - 158
      INTEGRAL    Likelihood = -0.90   Transmembrane  184 - 200
      INTEGRAL    Likelihood = -6.37   Transmembrane  208 - 224
      INTEGRAL    Likelihood = -4.51   Transmembrane  241 - 257
      PERIPHERAL  Likelihood =  0.79 (at 12)
      ALOM score:  -8.23  (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 40
      Charge difference:  0.0   C( 0.0) - N( 0.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  1.55
      Hyd Moment(95):  4.07    G content:       2
      D/E content:     1       S/T content:     2
      Score: -6.22

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 132  DRY|IA

NUCDISC: discrimination of nuclear localization signals
      pat4: RRKK (5) at  303
      pat7: none
      bipartite: none
      content of basic residues:   6.9%
      NLS Score: -0.16

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none
```

```
SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       22.2 %: mitochondrial
       11.1 %: nuclear

>> prediction for CG53530-03 is end (k=9)
```

A search of the NOV25a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 25D.

Table 25D. Geneseq Results for NOV25a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV25a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU85175	G-coupled olfactory receptor #36 - Homo sapiens, 318 aa. [WO200198526-A2, 27-DEC-2001]	1..318 1..318	318/318 (100%) 318/318 (100%)	0.0
AAU95525	Human olfactory and pheromone G protein-coupled receptor #12 - Homo sapiens, 318 aa. [WO200224726-A2, 28-MAR-2002]	1..318 1..318	318/318 (100%) 318/318 (100%)	0.0
AAU05140	Human odorant receptor (OR)-like protein, NOV10 - Homo sapiens, 318 aa. [WO200151632-A2, 19-JUL-2001]	1..318 1..318	318/318 (100%) 318/318 (100%)	0.0
AAU10320	G-protein coupled receptor (GCREC) #21 - Homo sapiens, 318 aa. [WO200166742-A2, 13-SEP-2001]	1..318 1..318	318/318 (100%) 318/318 (100%)	0.0
AAG71624	Human olfactory receptor polypeptide, SEQ ID NO: 1305 - Homo sapiens, 315 aa. [WO200127158-A2, 19-APR-2001]	1..315 1..315	314/315 (99%) 314/315 (99%)	0.0

In a BLAST search of public sequence databases, the NOV25a protein was found to have homology to the proteins shown in the BLASTP data in Table 25E.

Table 25E. Public BLASTP Results for NOV25a				
Protein Accession Number	**Protein/Organism/Length**	**NOV25a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8NGN0	Seven transmembrane helix receptor - Homo sapiens (Human), 318 aa.	1..318 1..318	318/318 (100%) 318/318 (100%)	0.0
Q8VFN1	Olfactory receptor MOR239-6 - Mus musculus (Mouse), 314 aa.	1..314 1..314	284/314 (90%) 296/314 (93%)	e-166
Q8VFU8	Olfactory receptor MOR239-5 - Mus musculus (Mouse), 314 aa.	1..317 1..313	200/317 (63%) 241/317 (75%)	e-113
Q8NGE8	Seven transmembrane helix receptor - Homo sapiens (Human), 314 aa.	1..304 1..304	190/304 (62%) 236/304 (77%)	e-110
Q8VF58	Olfactory receptor MOR240-1 - Mus musculus (Mouse), 311 aa.	1..304 1..304	186/304 (61%) 237/304 (77%)	e-110

PFam analysis indicates that the NOV25a protein contains the domains shown in the Table 25F.

Table 25F. Domain Analysis of NOV25a			
Pfam Domain	**NOV25a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	41..287	57/276 (21%) 172/276 (62%)	3.3e-28

Example 26.

The NOV26 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 26A.

Table 26A. NOV26 Sequence Analysis		
NOV26a, CG53563-03 DNA Sequence	SEQ ID NO: 329	1028 bp
	ORF Start: ATG at 28	ORF Stop: TAG at 961

AGGGAGAGAGACCAAGGGTGAGAAGAAATGTCCAACGCCAGCCTACTGACAGCGTTCATCCTCATGGG
CCTTCCCCATGCCCCAGCGCTGGACGCCCCCCTCTTTGGAGTCTTCCTGGTGGTTTACGTGCTCACTG
TGCTGGGGAACCTCCTCATCCTGCTGGTGATCAGGGTGGATTCTCACCTCCACACCACCATGTACTAC
TTCCTCACCAACCTGTCGTTCATTGACATGTGGTTCTCCACTGTCACGGTGCCCAAATTGCTGATGAC
TTTGGTGTTCCCAAGTGGCAGGGCTATCTCCTTCCACAGCTGCATGGCTCAGCTCTATTTCTTTCACT

TCCTAGGGGGCACCGAGTGTTTCCTCTACAGGGTCATGTCCTGTGATCGCTACCTGGCCATCAGTTAC
CCGCTCAGGTACACCAGCATGATGACTGGGCGCTCGTGTACTCTTCTGGCCACCAGCACTTGGCTCAG
TGGCTCTCTGCACTCTGCTGTCCAGGCCATATTGACTTTCCATTTGCCCTACTGTGGACCCAACTGGA
TCCAGCACTATTTGTGTGATGCACCGCCCATCCTGAAACTGGCCTGTGCAGACACCTCAGCCATAGAG
ACTGTCATTTTTGTGACTGTTGGAATAGTGGCCTCGGGCTGCTTTGTCCTGATAGTGCTGTCCTATGT
GTCCATCGTCTGTTCCATCCTGCGGATCCGCACCTCAGAGGGGAAGCACAGAGCCTTTCAGACCTGTG
CCTCCCACTGTATCGTGGTCCTTTGCTTCTTTGGCCCTGGTCTTTTCATTTACCTGAGGCCAGGCTCC
AGGAAAGCTGTGGATGGAGTTGTGGCCGTTTTCTACACTGTGCTGACGCCCCTTCTCAACCCTGTTGT
GTACACCCTGAGGAACAAGGAGGTGAAGAAAGCTCTGTTGAAGCTGAAAGACAAAGTAGCACATTCTC
AGAGCAAATAGACACTAGGGAAGATTACATATCTTAGCTCTTGTGAATAGTGCTGTGAAAAACATACA
GGGGCAGG

NOV26a, CG53563-03 Protein Sequence	SEQ ID NO: 330	311 aa	MW at 34518.7kD

MSNASLLTAFILMGLPHAPALDAPLFGVFLVVYVLTVLGNLLILLVIRVDSHLHTTMYYFLTNLSFID
MWFSTVTVPKLLMTLVFPSGRAISFHSCMAQLYFFHFLGGTECFLYRVMSCDRYLAISYPLRYTSMMT
GRSCTLLATSTWLSGSLHSAVQAILTFHLPYCGPNWIQHYLCDAPPILKLACADTSAIETVIFVTVGI
VASGCFVLIVLSYVSIVCSILRIRTSEGKHRAFQTCASHCIVVLCFFGPGLFIYLRPGSRKAVDGVVA
VFYTVLTPLLNPVVYTLRNKEVKKALLKLKDKVAHSQSK

NOV26b, CG53563-01 DNA Sequence	SEQ ID NO: 331	1038 bp
	ORF Start: ATG at 28	ORF Stop: TAG at 961

AGGGAGAGAGACCAAGGGTGAGAAGAA**ATG**TCCAACGCCAGCCTACTGACAGCGTTCATCCTCATGGG
CCTTCCCCATGCCCCAGCGCTGGACGCCCCCCTCTTTGGAGTCTTCCTGGTGGTTTACGTGCTCACTG
TGCTGGGGAACCTCCTCATCCTGCTGGTGATCAGGGTGGATTCTCACCTCCACACCACCATGTACTAC
TTCCTCACCAACCTGTCGTTCATTGACATGTGGTTCTCCACTGTCACGGTGCCCAAATTGCTGATGAC
TTTGGTGTTCCCAAGTGGCAGGGCTATCTCCTTCCACAGCTGCATGGCTCAGCTCTATTTCTTTCACT
TCCTAGGGGGCACCGAGTGTTTCCTCTACAGGGTCATGTCCTGTGATCGCTACCTGGCCATCAGTTAC
CCGCTCAGGTACACCAGCATGATGACTGGGCGCTCGTGTACTCTTCTGGCCACCAGCACTTGGCTCAG
TGGCTCTCTGCACTCTGCTGTCCAGGCCATATTGACTTTCCATTTGCCCTACTGTGGACCCAACTGGA
TCCAGCACTATTTGTGTGATGCACCGCCCATCCTGAAACTGGCCTGTGCAGACACCTCAGCCATAGAG
ACTGTCATTTTTGTGACTGTTGGAATAGTGGCCTCGGGCTGCTTTGTCCTGATAGTGCTGTCCTATGT
GTCCATCGTCTGTTCCATCCTGCGGATCCGCACCTCAGAGGGGAAGCACAGAGCCTTTCAGACCTGTG
CCTCCCACTGTATCGTGGTCCTTTGCTTCTTTGGCCCTGGTCTTTTCATTTACCTGAGGCCAGGCTCC
AGGAAAGCTGTGGATGGAGTTGTGGCCGTTTTCTACACTGTGCTGACGCCCCTTCTCAACCCTGTTGT
GTACACCCTGAGGAACAAGGAGGTGAAGAAAGCTCTGTTGAAGCTGAAAGACAAAGTAGCACATTCTC
AGAGCAAA**TAG**ACACTAGGGAAGATTACATATCTTAGCTCTTGTGAATAGTGCTGTGAAAAACATACA
GGGGCAGGTATCTTTTGG

NOV26b, CG53563-01 Protein Sequence	SEQ ID NO: 332	311 aa	MW at 34518.7kD

MSNASLLTAFILMGLPHAPALDAPLFGVFLVVYVLTVLGNLLILLVIRVDSHLHTTMYYFLTNLSFID
MWFSTVTVPKLLMTLVFPSGRAISFHSCMAQLYFFHFLGGTECFLYRVMSCDRYLAISYPLRYTSMMT
GRSCTLLATSTWLSGSLHSAVQAILTFHLPYCGPNWIQHYLCDAPPILKLACADTSAIETVIFVTVGI
VASGCFVLIVLSYVSIVCSILRIRTSEGKHRAFQTCASHCIVVLCFFGPGLFIYLRPGSRKAVDGVVA
VFYTVLTPLLNPVVYTLRNKEVKKALLKLKDKVAHSQSK

NOV26c, CG53563-02 DNA Sequence	SEQ ID NO: 333	1040 bp
	ORF Start: ATG at 36	ORF Stop: TAG at 969

TTTCTGTAAGAACAGAGCCCCATATATGAGAAGAA**ATG**TCCAACGCCACCCTACTGACAGCGTTCATC
CTCACGGGCCTTCCCCATGCCCCAGGGCTGGACGCCCCCCTCTTTGGAATCTTCCTGGTGGTTTACGT
GCTCACTGTGCTGGGGAACCTCCTCATCCTGCTGGTGATCAGGGTGGATTCTCACCTCCACACCCCCA
TGTACTACTTCCTCACCAACCTGTCCTTCATTGACATGTGGTTCTCCACTGTCACGGTGCCCAAAATG
CTGATGACCTTGGTGTCCCCAAGCGGCAGGACTATCTCCTTCCACAGCTGCGTGGCTCAGCTCTATTT
TTTCCACTTCCTGGGGAGCACCGAGTGTTTCCTCTACACAGTCATGTCCTATGATCGCTACCTGGCCA
TCAGTTACCCGCTCAGGTACACCAACATGATGACTGGGCGCTCGTGTGCCCTCCTGGCCACCGGCACT
TGGCTCAGTGGCTCTCTGCACTCTGCTGTCCAGACCATATTGACTTTCCATTTGCCCTACTGTGGACC
CAACCAGATCCAGCACTACTTCTGTGACGCACCGCCCATCCTGAAACTGGCCTGTGCAGACACCTCAG

CCAACGAGATGGTCATCTTTGTGAATATTGGGCTAGTGGCCTCGGGCTGCTTTGTCCTGATAGTGCTG
TCCTATGTGTCCATCGTCTGTTCCATCCTGCGGATCCGCACCTCAGAGGGGAGGCACAGAGCCTTTCA
GACCTGTGCCTCCCACTGTATCGTGGTCCTTTGCTTCTTTGGCCCTGGTCTTTTCATTTACCTGAGGC
CAGGCTCCAGGGACGCCTTGCATGGGGTTGTGGCCGTTTTCTACACCACGCTGACTCCTCTTTTCAAC
CCTGTTGTGTACACCCTGAGAAACAAGGAGGTAAAGAAAGCTCTGTTGAAGCTGAAAAATGGGTCAGT
ATTTGCTCAGGGTGAATAGTTAAGAAAGGCCATATATGGCTAACTTTTCTTTTTTTATTTGTAATTAA
ATTAAACCTTCAACATAAGC

NOV26c, CG53563-02 Protein Sequence	SEQ ID NO: 334	311 aa	MW at 34516.4kD

MSNATLLTAFILTGLPHAPGLDAPLFGIFLVVYVLTVLGNLLILLVIRVDSHLHTPMYYFLTNLSFID
MWFSTVTVPKMLMTLVSPSGRTISFHSCVAQLYFFHFLGSTECFLYTVMSYDRYLAISYPLRYTNMMT
GRSCALLATGTWLSGSLHSAVQTILTFHLPYCGPNQIQHYFCDAPPILKLACADTSANEMVIFVNIGL
VASGCFVLIVLSYVSIVCSILRIRTSEGRHRAFQTCASHCIVVLCFFGPGLFIYLRPGSRDALHGVVA
VFYTTLTPLFNPVVYTLRNKEVKKALLKLKNGSVFAQGE

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 26B.

Table 26B. Comparison of the NOV26 protein sequences.

```
NOV26a  MSNASLLTAFILMGLPHAPALDAPLFGVFLVVYVLTVLGNLLILLVIRVDSHLHTTMYYF
NOV26b  MSNASLLTAFILMGLPHAPALDAPLFGVFLVVYVLTVLGNLLILLVIRVDSHLHTTMYYF
NOV26c  MSNATLLTAFILTGLPHAPGLDAPLFGIFLVVYVLTVLGNLLILLVIRVDSHLHTPMYYF

NOV26a  LTNLSFIDMWFSTVTVPKLLMTLVFPSGRAISFHSCMAQLYFFHFLGGTECFLYRVMSCD
NOV26b  LTNLSFIDMWFSTVTVPKLLMTLVFPSGRAISFHSCMAQLYFFHFLGGTECFLYRVMSCD
NOV26c  LTNLSFIDMWFSTVTVPKMLMTLVSPSGRTISFHSCVAQLYFFHFLGSTECFLYTVMSYD

NOV26a  RYLAISYPLRYTSMMTGRSCTLLATSTWLSGSLHSAVQAILTFHLPYCGPNWIQHYLCDA
NOV26b  RYLAISYPLRYTSMMTGRSCTLLATSTWLSGSLHSAVQAILTFHLPYCGPNWIQHYLCDA
NOV26c  RYLAISYPLRYTNMMTGRSCALLATGTWLSGSLHSAVQTILTFHLPYCGPNQIQHYFCDA

NOV26a  PPILKLACADTSAIETVIFVTVGIVASGCFVLIVLSYVSIVCSILRIRTSEGKHRAFQTC
NOV26b  PPILKLACADTSAIETVIFVTVGIVASGCFVLIVLSYVSIVCSILRIRTSEGKHRAFQTC
NOV26c  PPILKLACADTSANEMVIFVNIGLVASGCFVLIVLSYVSIVCSILRIRTSEGRHRAFQTC

NOV26a  ASHCIVVLCFFGPGLFIYLRPGSRKAVDGVVAVFYTVLTPLLNPVVYTLRNKEVKKALLK
NOV26b  ASHCIVVLCFFGPGLFIYLRPGSRKAVDGVVAVFYTVLTPLLNPVVYTLRNKEVKKALLK
NOV26c  ASHCIVVLCFFGPGLFIYLRPGSRDALHGVVAVFYTTLTPLFNPVVYTLRNKEVKKALLK

NOV26a  LKDKVAHSQSK
NOV26b  LKDKVAHSQSK
NOV26c  LKNGSVFAQGE

NOV26a  (SEQ ID NO:  330)
NOV26b  (SEQ ID NO:  332)
NOV26c  (SEQ ID NO:  334)
```

Further analysis of the NOV26a protein yielded the following properties shown in Table 26C.

Table 26C. Protein Sequence Properties NOV26a

SignalP analysis:	Cleavage site between residues 52 and 53

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 21;  peak value  11.19
      PSG score:   6.79

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.58
      possible cleavage site: between 51 and 52

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   4
      INTEGRAL    Likelihood =-10.61   Transmembrane   31 -  47
      INTEGRAL    Likelihood =-10.08   Transmembrane  197 - 213
      INTEGRAL    Likelihood = -4.46   Transmembrane  243 - 259
      INTEGRAL    Likelihood = -2.60   Transmembrane  270 - 286
      PERIPHERAL  Likelihood =  0.85 (at 5)
      ALOM score: -10.61  (number of TMSs: 4)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 38
      Charge difference:  1.5   C( 1.0) - N(-0.5)
      C > N: C-terminal side will be inside

>>>Caution: Inconsistent mtop result with signal peptide
>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  4.41
      Hyd Moment(95):  3.21    G content:       1
      D/E content:     1       S/T content:     3
      Score: -5.14

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 99  GRA|IS

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.4%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none
```

```
SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       44.4 %: endoplasmic reticulum
       22.2 %: vacuolar
       11.1 %: Golgi
       11.1 %: vesicles of secretory system
       11.1 %: mitochondrial

>> prediction for CG53563-03 is end (k=9)
```

A search of the NOV26a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 26D.

Table 26D. Geneseq Results for NOV26a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV26a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU85273	G-coupled olfactory receptor #134 - Homo sapiens, 311 aa. [WO200198526-A2, 27-DEC-2001]	1..311 1..311	311/311 (100%) 311/311 (100%)	0.0
AAU95627	Human olfactory and pheromone G protein-coupled receptor #114 - Homo sapiens, 311 aa. [WO200224726-A2, 28-MAR-2002]	1..311 1..311	311/311 (100%) 311/311 (100%)	0.0
ABP95802	Human GPCR polypeptide SEQ ID NO 414 - Homo sapiens, 311 aa. [WO200216548-A2, 28-FEB-2002]	1..311 1..311	311/311 (100%) 311/311 (100%)	0.0
AAG66376	Partial NOV 15 protein sequence #2 - Unidentified, 311 aa. [WO200155179-A2, 02-AUG-2001]	1..311 1..311	311/311 (100%) 311/311 (100%)	0.0
AAG66335	Human NOV 15 protein sequence - Homo sapiens, 311 aa. [WO200155179-A2, 02-AUG-2001]	1..311 1..311	311/311 (100%) 311/311 (100%)	0.0

In a BLAST search of public sequence databases, the NOV26a protein was found to have homology to the proteins shown in the BLASTP data in Table 26E.

Table 26E. Public BLASTP Results for NOV26a				
Protein Accession Number	**Protein/Organism/Length**	**NOV26a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8NGN5	Seven transmembrane helix receptor - Homo sapiens (Human), 311 aa.	1..311 1..311	311/311 (100%) 311/311 (100%)	0.0
CAC69628	Sequence 13 from Patent WO0155179 - Homo sapiens (Human), 311 aa.	1..311 1..311	308/311 (99%) 308/311 (99%)	e-179
CAC69627	Sequence 11 from Patent WO0155179 - Homo sapiens (Human), 311 aa.	1..311 1..311	308/311 (99%) 308/311 (99%)	e-178
Q8NGN4	Seven transmembrane helix receptor - Homo sapiens (Human), 311 aa.	1..311 1..311	274/311 (88%) 288/311 (92%)	e-159
Q8NGN6	Seven transmembrane helix receptor - Homo sapiens (Human), 311 aa.	1..311 1..311	274/311 (88%) 286/311 (91%)	e-159

PFam analysis indicates that the NOV26a protein contains the domains shown in the Table 26F.

Table 26F. Domain Analysis of NOV26a			
Pfam Domain	**NOV26a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	39..287	52/276 (19%) 165/276 (60%)	2.4e-26

Example 27.

The NOV27 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 27A.

Table 27A. NOV27 Sequence Analysis		
NOV27a, CG53719-02 DNA Sequence	SEQ ID NO: 335	993 bp
	ORF Start: at 256	ORF Stop: end of sequence

TATTAGAAGAAGAGTGTCATAGAAGCCAATAACATTTCTGGGCCTGTGAGTGAATTTATCCTCCTGGG
CTTCCCCTGCCTGCTGCAGGGAGACCAAGATCCTCCTCTTTGTGGTCTTCTCCCTCATCTACCTTCTG
ACCCTCATGGGTAACACATCCATCATCTGCGCTGTGTGGTCAAGCCAGAAACTCCACACACCTATGTA
CATCCTCTTGGCTAATTTCTCTTTCCTGGAGATCTGCTGCATTAGTTCTGATGTCCCAAAATGTTGGC
CAATCTCATCTCCCATATCAAGAGCATCTCCTATGCTGGCTGCCTGCTCCAGTTCTTCTACTTCTCCA
TGTGTGCTGCAGAGGGCTATTTTCTGTCTGTGATGTCCTTTGATCGGTTCCTTACCATCTGTCGACCT
TTGCATTATCCCACAGTCATGACTCACCACCTGTGTGTCCGATTAGTGGCCTTCTGCAGGGCAGGTGG
TTTTCTATCCATACTGATGCCTGCAGTGCTTATGTCCCGAGTGCCTTTCTGTGGCCCTAACATCACTG
ACCATTTTTTCTGTAACCTGGGACCATTGCTGGCACTGTCCTGTGCCCCAGTTCCCAAAACTACTCTG
ACTTGTGCTACAGTAAGCTCTCTCATCATCTTCATCACCTTCCTCTACATTCTTGGGTCCCATATCTT
AGTTTTGCGAGCTGTTCTGTGGGTCCCAGCTGGCTCAGGCAGGAACAAAGCTTTCTCTACATGTGCTT
CCCATTTCTTGGTTGTTTCTTTCTTCTATGGCTCAGTCATGGTGATGTATGTGAGTCCAGGCTCCAGG
AGCCGTCCTGGGACACAGAAATTTGTGACATTGTTTTACTGCACAGCAAACCCATTCTTTAATCCCCT
GACCTACAGTCTCTGGAACAAAGATATGACAGATGCCCTTAAAAAAGTGCTGGGAGTGCCATCAAAAG
AAATATCTTGGAACACACTGAAATGATATACATTCTTCTAC

NOV27a, CG53719-02 Protein Sequence	SEQ ID NO: 336	321 aa	MW at 35567.8kD

MSVIEANNISGPVSELSSWASPACCRETKILLFVVFSLIYLLTLMGNTSIICAVWSSQKLHTPMYILL
ANFSFLEICCISSDVPMLANLISHIKSISYAGCLLQFFYFSMCAAEGYFLSVMSFDRFLTICRPLHYP
TVMTHHLCVRLVAFCRAGGFLSILMPAVLMSRVPFCGPNITDHFFCNLGPLLALSCAPVPKTTLTCAT
VSSLIIFITFLYILGSHILVLRAVLWVPAGSGRNKAFSTCASHFLVVSFFYGSVMVMYVSPGSRSRPG
TQKFVTLFYCTANPFFNPLTYSLWNKDMTDALKKVLGVPSKEISWNTLK

NOV27b, CG53719-01 DNA Sequence	SEQ ID NO: 337	1012 bp
	ORF Start: ATG at 25	ORF Stop: TGA at 988

TCATTTCCTTCATAGATTAGAAGAATGAGTGTCATAGAAGCCAATAACATTTCTGGGCCTGTGAGTGA
ATTTATCCTCCTGGGCTTCCCTGCCTGCTGCAGGGAGACCAAGATCCTCCTCTTTGTGGTCTTCTCCC
TCATCTACCTTCTGACCCTCATGGGTAACACATCCATCATCTGCGCTGTGTGGTCAAGCCAGAAACTC
CACACACCTATGTACATCCTCTTGGCTAATTTCTCTTTCCTGGAGATCTGCTGCATTAGTTCTGATGT
CCCAATGTTGGCCAATCTCATCTCCCATATCAAGAGCATCTCCTATGCTGGCTGCCTGCTCCAGTTCT
TCTACTTCTCCATGTGTGCTGCAGAAGGCTACTTTCTGTCTGTGATGTCCTTTGATCGGTTCCTTACC
ATCTGTCGACCTTTGCATTATCCCACAGTCATGACTCACCACCTGTGTGTCCGATTAGTGGCCTTCTG
CAGGGCAGGTGGTTTTCTATCCATACTGATGCCTGCAGTGCTTATGTCCCGAGTGCCTTTCTGTGGCC
CTAACATCACTGACCATTTTTTCTGTAACCTGGGACCATTGCTGGCACTGTCCTGTGCCCCAGTTCCC
AAAACTACTCTGACTTGTGCTACAGTAAGCTCTCTCATCATCTTCATCACCTTCCTCTACATTCTTGG
GTCCCATATCTTAGTTTTGCGAGCTGTTCTGTGGGTCCCAGCTGGCTCAGGCAGGAACAAAGCTTTCT
CTACATGTGCTTCCCATTTCTTGGTTGTTTCTTTCTTCTATGGCTCAGTCATGGTGATGTATGTGAGT
CCAGGCTCCAGGAGCCGCCCTGGGACACAGAAATTTGTGACATTGTTTTACTGCACAGCAACCCCATT
CTTTAATCCCCTGACCTACAGTCTCTGGAACAAAGATATGACAGATGCCCTTAAAAAAGTGCTGGGAG
TGCCATCAAAAGAAATATATTGGAACACACTGAAATGATATACATTCTTCTACAATTATT

NOV27b, CG53719-01 Protein Sequence	SEQ ID NO: 338	321 aa	MW at 35690.1kD

MSVIEANNISGPVSEFILLGFPACCRETKILLFVVFSLIYLLTLMGNTSIICAVWSSQKLHTPMYILL
ANFSFLEICCISSDVPMLANLISHIKSISYAGCLLQFFYFSMCAAEGYFLSVMSFDRFLTICRPLHYP
TVMTHHLCVRLVAFCRAGGFLSILMPAVLMSRVPFCGPNITDHFFCNLGPLLALSCAPVPKTTLTCAT
VSSLIIFITFLYILGSHILVLRAVLWVPAGSGRNKAFSTCASHFLVVSFFYGSVMVMYVSPGSRSRPG
TQKFVTLFYCTATPFFNPLTYSLWNKDMTDALKKVLGVPSKEIYWNTLK

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 27B.

Table 27B. Comparison of the NOV27 protein sequences.

```
NOV27a  MSVIEANNISGPVSELSSWASPACCRETKILLFVVFSLIYLLTLMGNTSIICAVWSSQKL
NOV27b  MSVIEANNISGPVSEFILLGFPACCRETKILLFVVFSLIYLLTLMGNTSIICAVWSSQKL

NOV27a  HTPMYILLANFSFLEICCISSDVPMLANLISHIKSISYAGCLLQFFYFSMCAAEGYFLSV
NOV27b  HTPMYILLANFSFLEICCISSDVPMLANLISHIKSISYAGCLLQFFYFSMCAAEGYFLSV

NOV27a  MSFDRFLTICRPLHYPTVMTHHLCVRLVAFCRAGGFLSILMPAVLMSRVPFCGPNITDHF
NOV27b  MSFDRFLTICRPLHYPTVMTHHLCVRLVAFCRAGGFLSILMPAVLMSRVPFCGPNITDHF

NOV27a  FCNLGPLLALSCAPVPKTTLTCATVSSLIIFITFLYILGSHILVLRAVLWVPAGSGRNKA
NOV27b  FCNLGPLLALSCAPVPKTTLTCATVSSLIIFITFLYILGSHILVLRAVLWVPAGSGRNKA

NOV27a  FSTCASHFLVVSFFYGSVMVMYVSPGSRSRPGTQKFVTLFYCTANPFFNPLTYSLWNKDM
NOV27b  FSTCASHFLVVSFFYGSVMVMYVSPGSRSRPGTQKFVTLFYCTATPFFNPLTYSLWNKDM

NOV27a  TDALKKVLGVPSKEISWNTLK
NOV27b  TDALKKVLGVPSKEIYWNTLK

NOV27a  (SEQ ID NO:  336)
NOV27b  (SEQ ID NO:  338)
```

Further analysis of the NOV27a protein yielded the following properties shown in Table 27C.

Table 27C. Protein Sequence Properties NOV27a

SignalP analysis:	Cleavage site between residues 47 and 48

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 5;  pos.chg 0;  neg.chg 1
      H-region:  length 9;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.31
      possible cleavage site: between 43 and 44

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   5
      INTEGRAL    Likelihood = -9.34   Transmembrane   30 -  46
      INTEGRAL    Likelihood = -1.28   Transmembrane   64 -  80
      INTEGRAL    Likelihood = -0.96   Transmembrane  156 - 172
```

```
      INTEGRAL    Likelihood = -7.54   Transmembrane  208 - 224
      INTEGRAL    Likelihood = -2.50   Transmembrane  248 - 264
      PERIPHERAL  Likelihood =  1.48 (at 180)
      ALOM score:  -9.34  (number of TMSs: 5)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 37
      Charge difference:  1.5   C( 1.5) - N( 0.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  0.86
      Hyd Moment(95):  3.03     G content:       1
      D/E content:     2        S/T content:     3
      Score: -7.70

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   6.5%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none
```

```
checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

---------------------------

Final Results (k = 9/23):

       44.4 %: endoplasmic reticulum
       22.2 %: vacuolar
       11.1 %: Golgi
       11.1 %: vesicles of secretory system
       11.1 %: mitochondrial

>> prediction for CG53719-02 is end (k=9)
```

A search of the NOV27a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 27D.

Table 27D. Geneseq Results for NOV27a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV27a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU07092	Human odorant receptor (OR) polypeptide #9 - Homo sapiens, 321 aa. [WO200157215-A2, 09-AUG-2001]	1..321 1..321	313/321 (97%) 313/321 (97%)	0.0
ABP72226	Human G-protein coupled receptor GCREC-20 (odorant receptor) - Homo sapiens, 322 aa. [WO2003000859-A2, 03-JAN-2003]	1..321 1..322	314/322 (97%) 314/322 (97%)	0.0
AAG72445	Human OR-like polypeptide query sequence, SEQ ID NO: 2126 - Homo sapiens, 322 aa. [WO200127158-A2, 19-APR-2001]	1..321 1..322	310/322 (96%) 311/322 (96%)	0.0
AAG72079	Human olfactory receptor polypeptide, SEQ ID NO: 1760 - Homo sapiens, 322 aa. [WO200127158-A2, 19-APR-2001]	1..321 1..322	309/322 (95%) 310/322 (95%)	e-180
ABP72218	Human G-protein coupled receptor GCREC-12 (olfactory receptor) - Homo sapiens, 333 aa. [WO2003000859-A2, 03-JAN-2003]	29..321 39..333	255/295 (86%) 270/295 (91%)	e-149

In a BLAST search of public sequence databases, the NOV27a protein was found to have homology to the proteins shown in the BLASTP data in Table 27E.

Table 27E. Public BLASTP Results for NOV27a				
Protein Accession Number	**Protein/Organism/Length**	**NOV27a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8VFT6	Olfactory receptor MOR106-5 - Mus musculus (Mouse), 312 aa.	6..309 4..310	181/308 (58%) 224/308 (71%)	1e-98
Q8NGC7	Seven transmembrane helix receptor - Homo sapiens (Human), 330 aa.	26..310 39..325	169/287 (58%) 212/287 (72%)	2e-93
CAD57956	Sequence 107 from Patent WO0250276 - Homo sapiens (Human), 307 aa.	26..310 16..302	169/287 (58%) 211/287 (72%)	8e-93
Q8VFT9	Olfactory receptor MOR106-2 - Mus musculus (Mouse), 309 aa.	24..308 21..307	168/287 (58%) 210/287 (72%)	2e-90
Q8VFT8	Olfactory receptor MOR106-3 - Mus musculus (Mouse), 311 aa.	24..308 23..309	166/287 (57%) 212/287 (73%)	4e-90

PFam analysis indicates that the NOV27a protein contains the domains shown in the Table 27F.

Table 27F. Domain Analysis of NOV27a			
Pfam Domain	**NOV27a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	46..293	47/276 (17%) 166/276 (60%)	4.5e-13

Example 28.

The NOV28 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 28A.

Table 28A. NOV28 Sequence Analysis		
NOV28a, CG53746-04 DNA Sequence	SEQ ID NO: 339	1604 bp
	ORF Start: ATG at 648	ORF Stop: TGA at 1581

ACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTTGGATATTTGGTGTACTTTTCCTCAATTGTGA
AATCTCTGGGTGGGGACCACAGCTCAGTGTTGAGTTACTGACCTCTTCTTGGTCCTGTGGATTAGCAT
GCAGCTAATCTGTTTGACCTCTGTTTGGAATTCGGAATTCTTAATGACTACACATCTTTGATACAATA
GATGATACCTCAAACATCCTTTTGAACAGCTGTTCTTTTCCATGAGTCTTGGTCTATTCTGACATTTA
TGTCTCCCTTCAATCACGTCTTTGGCCTTAGAAGATTGAGTTACTGGATTCTTTATATATTCTAGTGG
TCATCTCTGAAATGTGCTCAGAGAGCACCTAAATTAACCATCCAATACGAGTTGAGTGTGTTAAGTTA

AAAAAAAAAAAAAGATTTTTCTGAGTATTCCTGACCTTACATCAGTGAACATTTATGTTTTAAAGTCT
TACATAAGATACTGTGTGTGAAAGCATTTTCTTCCCAAATTTACATGAGTGCCTAAATTGTTATACTT
TTTGGTTAAATAGATATTGGAAAAATAAGTGTGCAATTATAGCATTTAATCCCATTATAATATTCATT
TGTTTTTCTTTCAGTTGGCACCATTCCCAATTTAG**ATG**GGGACTGGAAATGACACCACTGTGGTAGAG
TTTACTCTTTTGGGATTATCCGAGGATACTACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAAT
TTATGTTGTCACCTTAATGGGTAATATCAGCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATA
CACCCATGTACATTTTCCTCTGCCATTTGGCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCT
GTCATGCTCATGAGCTTCCTAAGGAAAGAAACCTCTCTCCCTATTGCTGGTTGTGTGGCCCAGCTCTG
TTCTGTAGTGACGTTTGGTACGGCCGAGTGCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGG
CCATCTGCTCACCCCTGCTCTACTCTACCTGCATGTCCCCTGGAGTCTGCATCATCTTAGTGGGCATG
TCCTACCTGGGTGGATGTGTGAATGCTTGGACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGG
GCCAAATAAAGTCAATCACTTTTTCTGTGACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATT
TTACTTTTGAAATAATTCCAGCTATCTCTTCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCC
ATATCCTACATCTATATCCTCATCACCATCCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTT
CTCCACCTGCACCTCCCACCTCACTGCAGTCACTCTGTTCTATGGGACCATTACCTTCATTTATGTGA
TGCCCAAGTCCAGCTACTCAACTGACCAGAACAAGGTGGTGTCTGTGTTCTACACCGTGGTGATTCCC
ATGTTGAACCCCCTGATCTACAGCCTCAGGAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAG
AATAAAAATATTTTCT**TGA**TGAAACTAGTTAGTTTGAAGA

NOV28a, CG53746-04 Protein Sequence	SEQ ID NO: 340	311 aa	MW at 34309.4kD

MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF
VDIGYSSSVTPVMLMSFLRKETSLPIAGCVAQLCSVVTFGTAECFLLAAMAYDRYVAICSPLLYSTCM
SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG
SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK
VVSVFYTVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS

NOV28b, CG53746-01 DNA Sequence	SEQ ID NO: 341	984 bp
	ORF Start: ATG at 24	ORF Stop: TGA at 957

AGTTGGCACCATTCCCAATATAG**ATG**GGGACTGGAAATGACACCACTGTGGTAGAGTTTACTCTTTTG
GGGTTATCTGAGGATACTACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAATTTATGTTGTCAC
CTTAATGGGTAATATCAGCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATACACCCATGTACA
TTTTCCTCTGCCATTTGGCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCTGTCATGCTCATG
AGCTTCCTAAGGAAAGAAACCTCTCTCCCTGTTGCTGGTTGTGTGGCCCAGCTCTGTTCTGTAGTGAC
GTTTGGTACGGCCGAGTGCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGGCCATCTGCTCAC
CCCTGCTCTACTCTACCTGCATGTCCCCTGGAGTCTGCATCATCTTAGTGGGCATGTCCTACCTGGGT
GGATGTGTGAATGCTTGGACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGGGCCAAATAAAGT
CAATCACTTTTTCTGTGACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATTTTACTTTTGAAA
TAATTCCAGCTATCTCTTCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCCATATCCTACATC
TATATCCTCATCACCATCCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTTCTCCACCTGCAC
CTCCCACCTCACTGCAGTCACTCTGTTCTATGGGACCATTACCTTCATTTATGTGATGCCCAAGTCCA
GCTACTCAACTGACCAGAACAAGGTGGTGTCTGTGTTCTACACCGTGGTGATTCCCATGTTGAACCCC
CTGATCTACAGCCTCAGGAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAGAATAAAAATATT
TTCTTGATGAAACTAGTTAGTTTGAAGAATCT

NOV28b, CG53746-01 Protein Sequence	SEQ ID NO: 342	311 aa	MW at 34295.4kD

MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF
VDIGYSSSVTPVMLMSFLRKETSLPVAGCVAQLCSVVTFGTAECFLLAAMAYDRYVAICSPLLYSTCM
SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG
SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK
VVSVFYTVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS

NOV28c, CG53746-02 DNA Sequence	SEQ ID NO: 343	952 bp
	ORF Start: ATG at 7	ORF Stop: TGA at 940

ATATAG**ATG**GGGACTGGAAATGACACCACTGTGGTAGAGTTTACTCTTTTGGGGTTATCCGAGGATAC
TACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAATTTATGTTGTCACCTTAATGGGTAATATCA
GCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATACACCCATGTACATTTTCCTCTGCCATTTG

GCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCTGTCATGCTCATGAGCTTCCTAAGGAAAGA
AACCTCTCTCCCTGTTGCTGGTTGTGTGGCCCAGCTCTGTTCTGTGGTGACGTTTGGTACGGCCGAGT
GCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGGCCATCTGCTCACCCCTGCTCTACTCTACC
TGCATGTCCCCTGGAGTCTGCATCATCTTAGTGGGCATGTCCTACCTGGGTGGATGTGTGAATGCTTG
GACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGGGCCAAATAAAGTCAATCACTTTTTCTGTG
ACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATTTTACTTTTGAAATAATTCCAGCTATCTCT
TCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCCATATCCTACATCTATATCCTCATCACCAT
CCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTTCTCCACCTGCACCTCCCACCTCACTGCAG
TCACTCTGTTTTATGGGACCATTACCTTCATTTATGTGATGCCCAAGTCCAGCTACTCAACTGACCAG
AACAAGGTGGTGTCTGTGTTCTACACCGTGGTGATTCCCATGTTGAACCCCCTGATCTACAGCCTCAG
GAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAGAATAAAAATATTTTCTTGATGAAACTAGT

NOV28c, CG53746-02 Protein Sequence	SEQ ID NO: 344	311 aa	MW at 34295.4kD

MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF
VDIGYSSSVTPVMLMSFLRKETSLPVAGCVAQLCSVVTFGTAECFLLAAMAYDRYVAICSPLLYSTCM
SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG
SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK
VVSVFYTVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS

NOV28d, CG53746-03 DNA Sequence	SEQ ID NO: 345	964 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 941

TATATAGATGGGGACTGGAAATGACACCACTGTGGTAGAGTTTACTCTTTTGGGGTTATCCGAGGATA
CTACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAATTTATGTTGTCACCTTAATGGGTAATATC
AGCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATACACCCATGTACATTTTCCTCTGCCATTT
GGCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCTGTCATGCTCATGAGCTTCCTAAGGAAAG
AAACCTCTCTCCCTGTTGCTGGTTGTGTGGCCCAGCTCTGTTCTGTGGTGACGTTTGGTACGGCCGAG
TGCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGGCCATCTGCTCACCCCTGCTCTACTCTAC
CTGCATGTCCCCTGGAGTCTGCATCATCTTAGTGGGCATGTCCTACCTGGGTGGATGTGTGAATGCTT
GGACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGGGCCAAATAAAGTCAATCACTTTTTCTGT
GACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATTTTACTTTTGAAATAATTCCAGCTATCTC
TTCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCCATATCCTACATCTATATCCTCATCACCA
TCCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTTCTCCACCTGCACCTCCCACCTCACTGCA
GTCACTCTGTTTTATGGGACCATTACCTTCATTTATGTGATGCCCAAGTCCAGCTACTCAACTGACCA
GAACAAGGTGGTGTCTGTGTTCTACACCGTGGTGATTCCCATGTTGAACCCCCTGATCTACAGCCTCA
GGAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAGAATAAAAATATTTTCTTGATGAAACTAG
TTAGTTTGAAGA

NOV28d, CG53746-03 Protein Sequence	SEQ ID NO: 346	311 aa	MW at 34295.4kD

MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF
VDIGYSSSVTPVMLMSFLRKETSLPVAGCVAQLCSVVTFGTAECFLLAAMAYDRYVAICSPLLYSTCM
SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG
SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK
VVSVFYTVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS

NOV28e, SNP13373967 of CG53746-04, DNA Sequence	SEQ ID NO: 347	1604 bp
	ORF Start: ATG at 648	ORF Stop: TGA at 1581
	SNP Pos: 466	SNP Change: T to C

ACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTTGGATATTTGGTGTACTTTTCCTCAATTGTGA
AATCTCTGGGTGGGGACCACAGCTCAGTGTTGAGTTACTGACCTCTTCTTGGTCCTGTGGATTAGCAT
GCAGCTAATCTGTTTGACCTCTGTTTGGAATTCGGAATTCTTAATGACTACACATCTTTGATACAATA
GATGATACCTCAAACATCCTTTTGAACAGCTGTTCTTTTCCATGAGTCTTGGTCTATTCTGACATTTA
TGTCTCCCTTCAATCACGTCTTTGGCCTTAGAAGATTGAGTTACTGGATTCTTTATATATTCTAGTGG
TCATCTCTGAAATGTGCTCAGAGAGCACCTAAATTAACCATCCAATACGAGTTGAGTGTGTTAAGTTA
AAAAAAAAAAAAAGATTTTTCTGAGTATTCCTGACCTTACATCAGTGAACATTTATGCTTTAAAGTCT
TACATAAGATACTGTGTGTGAAAGCATTTTCTTCCCAAATTTACATGAGTGCCTAAATTGTTATACTT
TTTGGTTAAATAGATATTGGAAAAATAAGTGTGCAATTATAGCATTTAATCCCATTATAATATTCATT

TGTTTTTCTTTCAGTTGGCACCATTCCCAATTTAGATGGGGACTGGAAATGACACCACTGTGGTAGAG
TTTACTCTTTTGGGATTATCCGAGGATACTACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAAT
TTATGTTGTCACCTTAATGGGTAATATCAGCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATA
CACCCATGTACATTTTCCTCTGCCATTTGGCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCT
GTCATGCTCATGAGCTTCCTAAGGAAAGAAACCTCTCTCCCTATTGCTGGTTGTGTGGCCCAGCTCTG
TTCTGTAGTGACGTTTGGTACGGCCGAGTGCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGG
CCATCTGCTCACCCCTGCTCTACTCTACCTGCATGTCCCCTGGAGTCTGCATCATCTTAGTGGGCATG
TCCTACCTGGGTGGATGTGTGAATGCTTGGACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGG
GCCAAATAAAGTCAATCACTTTTTCTGTGACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATT
TTACTTTTGAAATAATTCCAGCTATCTCTTCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCC
ATATCCTACATCTATATCCTCATCACCATCCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTT
CTCCACCTGCACCTCCCACCTCACTGCAGTCACTCTGTTCTATGGGACCATTACCTTCATTTATGTGA
TGCCCAAGTCCAGCTACTCAACTGACCAGAACAAGGTGGTGTCTGTGTTCTACACCGTGGTGATTCCC
ATGTTGAACCCCCTGATCTACAGCCTCAGGAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAG
AATAAAAATATTTTCTTGATGAAACTAGTTAGTTTGAAGA

NOV28e, SNP13373967 of CG53746-04, Protein Sequence	SEQ ID NO: 348	311 aa	MW at 34309.4kD
			SNP Change: no change

MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF
VDIGYSSSVTPVMLMSFLRKETSLPIAGCVAQLCSVVTFGTAECFLLAAMAYDRYVAICSPLLYSTCM
SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG
SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK
VVSVFYTVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS

NOV28f, SNP13373968 of CG53746-04, DNA Sequence	SEQ ID NO: 349	1604 bp
	ORF Start: ATG at 648	ORF Stop: TGA at 1581
	SNP Pos: 550	SNP Change: T to C

ACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTTGGATATTTGGTGTACTTTTCCTCAATTGTGA
AATCTCTGGGTGGGGACCACAGCTCAGTGTTGAGTTACTGACCTCTTCTTGGTCCTGTGGATTAGCAT
GCAGCTAATCTGTTTGACCTCTGTTTGGAATTCGGAATTCTTAATGACTACACATCTTTGATACAATA
GATGATACCTCAAACATCCTTTTGAACAGCTGTTCTTTTCCATGAGTCTTGGTCTATTCTGACATTTA
TGTCTCCCTTCAATCACGTCTTTGGCCTTAGAAGATTGAGTTACTGGATTCTTTATATATTCTAGTGG
TCATCTCTGAAATGTGCTCAGAGAGCACCTAAATTAACCATCCAATACGAGTTGAGTGTGTTAAGTTA
AAAAAAAAAAAAAAGATTTTTCTGAGTATTCCTGACCTTACATCAGTGAACATTTATGTTTTAAAGTCT
TACATAAGATACTGTGTGTGAAAGCATTTTCTTCCCAAATTTACATGAGTGCCTAAATTGTTATACTT
TTTGGCTAAATAGATATTGGAAAAATAAGTGTGCAATTATAGCATTTAATCCCATTATAATATTCATT
TGTTTTTCTTTCAGTTGGCACCATTCCCAATTTAGATGGGGACTGGAAATGACACCACTGTGGTAGAG
TTTACTCTTTTGGGATTATCCGAGGATACTACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAAT
TTATGTTGTCACCTTAATGGGTAATATCAGCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATA
CACCCATGTACATTTTCCTCTGCCATTTGGCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCT
GTCATGCTCATGAGCTTCCTAAGGAAAGAAACCTCTCTCCCTATTGCTGGTTGTGTGGCCCAGCTCTG
TTCTGTAGTGACGTTTGGTACGGCCGAGTGCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGG
CCATCTGCTCACCCCTGCTCTACTCTACCTGCATGTCCCCTGGAGTCTGCATCATCTTAGTGGGCATG
TCCTACCTGGGTGGATGTGTGAATGCTTGGACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGG
GCCAAATAAAGTCAATCACTTTTTCTGTGACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATT
TTACTTTTGAAATAATTCCAGCTATCTCTTCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCC
ATATCCTACATCTATATCCTCATCACCATCCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTT
CTCCACCTGCACCTCCCACCTCACTGCAGTCACTCTGTTCTATGGGACCATTACCTTCATTTATGTGA
TGCCCAAGTCCAGCTACTCAACTGACCAGAACAAGGTGGTGTCTGTGTTCTACACCGTGGTGATTCCC
ATGTTGAACCCCCTGATCTACAGCCTCAGGAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAG
AATAAAAATATTTTCTTGATGAAACTAGTTAGTTTGAAGA

NOV28f, SNP13373968 of CG53746-04, Protein Sequence	SEQ ID NO: 350	311 aa	MW at 34309.4kD
			SNP Change: no change

MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF
VDIGYSSSVTPVMLMSFLRKETSLPIAGCVAQLCSVVTFGTAECFLLAAMAYDRYVAICSPLLYSTCM
SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG
SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK

VVSVFYTVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS

NOV28g, SNP13382432 of CG53746-04, DNA Sequence	SEQ ID NO: 351	1604 bp
	ORF Start: ATG at 648	ORF Stop: TGA at 1581
	SNP Pos: 644	SNP Change: T to A

ACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTTGGATATTTGGTGTACTTTTCCTCAATTGTGA
AATCTCTGGGTGGGGACCACAGCTCAGTGTTGAGTTACTGACCTCTTCTTGGTCCTGTGGATTAGCAT
GCAGCTAATCTGTTTGACCTCTGTTTGGAATTCGGAATTCTTAATGACTACACATCTTTGATACAATA
GATGATACCTCAAACATCCTTTTGAACAGCTGTTCTTTTCCATGAGTCTTGGTCTATTCTGACATTTA
TGTCTCCCTTCAATCACGTCTTTGGCCTTAGAAGATTGAGTTACTGGATTCTTTATATATTCTAGTGG
TCATCTCTGAAATGTGCTCAGAGAGCACCTAAATTAACCATCCAATACGAGTTGAGTGTGTTAAGTTA
AAAAAAAAAAAAAGATTTTTCTGAGTATTCCTGACCTTACATCAGTGAACATTTATGTTTTAAAGTCT
TACATAAGATACTGTGTGTGAAAGCATTTTCTTCCCAAATTTACATGAGTGCCTAAATTGTTATACTT
TTTGGTTAAATAGATATTGGAAAAATAAGTGTGCAATTATAGCATTTAATCCCATTATAATATTCATT
TGTTTTTCTTTCAGTTGGCACCATTCCCAATATAGATGGGGACTGGAAATGACACCACTGTGGTAGAG
TTTACTCTTTTGGGATTATCCGAGGATACTACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAAT
TTATGTTGTCACCTTAATGGGTAATATCAGCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATA
CACCCATGTACATTTTCCTCTGCCATTTGGCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCT
GTCATGCTCATGAGCTTCCTAAGGAAAGAAACCTCTCTCCCTATTGCTGGTTGTGTGGCCCAGCTCTG
TTCTGTAGTGACGTTTGGTACGGCCGAGTGCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGG
CCATCTGCTCACCCCTGCTCTACTCTACCTGCATGTCCCCTGGAGTCTGCATCATCTTAGTGGGCATG
TCCTACCTGGGTGGATGTGTGAATGCTTGGACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGG
GCCAAATAAAGTCAATCACTTTTTCTGTGACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATT
TTACTTTTGAAATAATTCCAGCTATCTCTTCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCC
ATATCCTACATCTATATCCTCATCACCATCCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTT
CTCCACCTGCACCTCCCACCTCACTGCAGTCACTCTGTTCTATGGGACCATTACCTTCATTTATGTGA
TGCCCAAGTCCAGCTACTCAACTGACCAGAACAAGGTGGTGTCTGTGTTCTACACCGTGGTGATTCCC
ATGTTGAACCCCCTGATCTACAGCCTCAGGAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAG
AATAAAAATATTTTCTTGATGAAACTAGTTAGTTTGAAGA

NOV28g, SNP13382432 of CG53746-04, Protein Sequence	SEQ ID NO: 352	311 aa	MW at 34309.4kD
			SNP Change: no change

MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF
VDIGYSSSVTPVMLMSFLRKETSLPIAGCVAQLCSVVTFGTAECFLLAAMAYDRYVAICSPLLYSTCM
SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG
SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK
VVSVFYTVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS

NOV28h, SNP13373969 of CG53746-04, DNA Sequence	SEQ ID NO: 353	1604 bp
	ORF Start: ATG at 648	ORF Stop: TGA at 1581
	SNP Pos: 695	SNP Change: A to G

ACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTTGGATATTTGGTGTACTTTTCCTCAATTGTGA
AATCTCTGGGTGGGGACCACAGCTCAGTGTTGAGTTACTGACCTCTTCTTGGTCCTGTGGATTAGCAT
GCAGCTAATCTGTTTGACCTCTGTTTGGAATTCGGAATTCTTAATGACTACACATCTTTGATACAATA
GATGATACCTCAAACATCCTTTTGAACAGCTGTTCTTTTCCATGAGTCTTGGTCTATTCTGACATTTA
TGTCTCCCTTCAATCACGTCTTTGGCCTTAGAAGATTGAGTTACTGGATTCTTTATATATTCTAGTGG
TCATCTCTGAAATGTGCTCAGAGAGCACCTAAATTAACCATCCAATACGAGTTGAGTGTGTTAAGTTA
AAAAAAAAAAAAAAGATTTTTCTGAGTATTCCTGACCTTACATCAGTGAACATTTATGTTTTAAAGTCT
TACATAAGATACTGTGTGTGAAAGCATTTTCTTCCCAAATTTACATGAGTGCCTAAATTGTTATACTT
TTTGGTTAAATAGATATTGGAAAAATAAGTGTGCAATTATAGCATTTAATCCCATTATAATATTCATT
TGTTTTTCTTTCAGTTGGCACCATTCCCAATTTAGATGGGGACTGGAAATGACACCACTGTGGTAGAG
TTTACTCTTTTGGGGTTATCCGAGGATACTACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAAT
TTATGTTGTCACCTTAATGGGTAATATCAGCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATA
CACCCATGTACATTTTCCTCTGCCATTTGGCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCT
GTCATGCTCATGAGCTTCCTAAGGAAAGAAACCTCTCTCCCTATTGCTGGTTGTGTGGCCCAGCTCTG
TTCTGTAGTGACGTTTGGTACGGCCGAGTGCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGG
CCATCTGCTCACCCCTGCTCTACTCTACCTGCATGTCCCCTGGAGTCTGCATCATCTTAGTGGGCATG

TCCTACCTGGGTGGATGTGTGAATGCTTGGACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGG
GCCAAATAAAGTCAATCACTTTTTCTGTGACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATT
TTACTTTTGAAATAATTCCAGCTATCTCTTCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCC
ATATCCTACATCTATATCCTCATCACCATCCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTT
CTCCACCTGCACCTCCCACCTCACTGCAGTCACTCTGTTCTATGGGACCATTACCTTCATTTATGTGA
TGCCCAAGTCCAGCTACTCAACTGACCAGAACAAGGTGGTGTCTGTGTTCTACACCGTGGTGATTCCC
ATGTTGAACCCCCTGATCTACAGCCTCAGGAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAG
AATAAAAATATTTTCTTGATGAAACTAGTTAGTTTGAAGA

NOV28h, SNP13373969 of CG53746-04, Protein Sequence	SEQ ID NO: 354	311 aa	MW at 34309.4kD
	SNP Pos: 16		SNP Change: Gly to Gly

MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF
VDIGYSSSVTPVMLMSFLRKETSLPIAGCVAQLCSVVTFGTAECFLLAAMAYDRYVAICSPLLYSTCM
SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG
SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK
VVSVFYTVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS

NOV28i, SNP13373970 of CG53746-04, DNA Sequence	SEQ ID NO: 355	1604 bp
	ORF Start: ATG at 648	ORF Stop: TGA at 1581
	SNP Pos: 927	SNP Change: A to G

ACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTTGGATATTTGGTGTACTTTTCCTCAATTGTGA
AATCTCTGGGTGGGGACCACAGCTCAGTGTTGAGTTACTGACCTCTTCTTGGTCCTGTGGATTAGCAT
GCAGCTAATCTGTTTGACCTCTGTTTGGAATTCGGAATTCTTAATGACTACACATCTTTGATACAATA
GATGATACCTCAAACATCCTTTTGAACAGCTGTTCTTTTCCATGAGTCTTGGTCTATTCTGACATTTA
TGTCTCCCTTCAATCACGTCTTTGGCCTTAGAAGATTGAGTTACTGGATTCTTTATATATTCTAGTGG
TCATCTCTGAAATGTGCTCAGAGAGCACCTAAATTAACCATCCAATACGAGTTGAGTGTGTTAAGTTA
AAAAAAAAAAAAAAGATTTTTCTGAGTATTCCTGACCTTACATCAGTGAACATTTATGTTTTAAAGTCT
TACATAAGATACTGTGTGTGAAAGCATTTTCTTCCCAAATTTACATGAGTGCCTAAATTGTTATACTT
TTTGGTTAAATAGATATTGGAAAAATAAGTGTGCAATTATAGCATTTAATCCCATTATAATATTCATT
TGTTTTTCTTTCAGTTGGCACCATTCCCAATTTAGATGGGGACTGGAAATGACACCACTGTGGTAGAG
TTTACTCTTTTGGGATTATCCGAGGATACTACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAAT
TTATGTTGTCACCTTAATGGGTAATATCAGCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATA
CACCCATGTACATTTTCCTCTGCCATTTGGCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCT
GTCATGCTCATGAGCTTCCTAAGGAAAGAAACCTCTCTCCCTGTTGCTGGTTGTGTGGCCCAGCTCTG
TTCTGTAGTGACGTTTGGTACGGCCGAGTGCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGG
CCATCTGCTCACCCCTGCTCTACTCTACCTGCATGTCCCCTGGAGTCTGCATCATCTTAGTGGGCATG
TCCTACCTGGGTGGATGTGTGAATGCTTGGACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGG
GCCAAATAAAGTCAATCACTTTTTCTGTGACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATT
TTACTTTTGAAATAATTCCAGCTATCTCTTCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCC
ATATCCTACATCTATATCCTCATCACCATCCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTT
CTCCACCTGCACCTCCCACCTCACTGCAGTCACTCTGTTCTATGGGACCATTACCTTCATTTATGTGA
TGCCCAAGTCCAGCTACTCAACTGACCAGAACAAGGTGGTGTCTGTGTTCTACACCGTGGTGATTCCC
ATGTTGAACCCCCTGATCTACAGCCTCAGGAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAG
AATAAAAATATTTTCTTGATGAAACTAGTTAGTTTGAAGA

NOV28i, SNP13373970 of CG53746-04, Protein Sequence	SEQ ID NO: 356	311 aa	MW at 34295.4kD
	SNP Pos: 94		SNP Change: Ile to Val

MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF
VDIGYSSSVTPVMLMSFLRKETSLPVAGCVAQLCSVVTFGTAECFLLAAMAYDRYVAICSPLLYSTCM
SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG
SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK
VVSVFYTVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS

NOV28j, SNP13373971 of CG53746-04, DNA Sequence	SEQ ID NO: 357	1604 bp
	ORF Start: ATG at 648	ORF Stop: TGA at 1581
	SNP Pos: 959	SNP Change: A to G

ACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTTGGATATTTGGTGTACTTTTCCTCAATTGTGA

AATCTCTGGGTGGGGACCACAGCTCAGTGTTGAGTTACTGACCTCTTCTTGGTCCTGTGGATTAGCAT
GCAGCTAATCTGTTTGACCTCTGTTTGGAATTCGGAATTCTTAATGACTACACATCTTTGATACAATA
GATGATACCTCAAACATCCTTTTGAACAGCTGTTCTTTTCCATGAGTCTTGGTCTATTCTGACATTTA
TGTCTCCCTTCAATCACGTCTTTGGCCTTAGAAGATTGAGTTACTGGATTCTTTATATATTCTAGTGG
TCATCTCTGAAATGTGCTCAGAGAGCACCTAAATTAACCATCCAATACGAGTTGAGTGTGTTAAGTTA
AAAAAAAAAAAAAGATTTTTCTGAGTATTCCTGACCTTACATCAGTGAACATTTATGTTTTAAAGTCT
TACATAAGATACTGTGTGTGAAAGCATTTTCTTCCCAAATTTACATGAGTGCCTAAATTGTTATACTT
TTTGGTTAAATAGATATTGGAAAAATAAGTGTGCAATTATAGCATTTAATCCCATTATAATATTCATT
TGTTTTTCTTTCAGTTGGCACCATTCCCAATTTAG**ATG**GGGACTGGAAATGACACCACTGTGGTAGAG
TTTACTCTTTTGGGATTATCCGAGGATACTACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAAT
TTATGTTGTCACCTTAATGGGTAATATCAGCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATA
CACCCATGTACATTTTCCTCTGCCATTTGGCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCT
GTCATGCTCATGAGCTTCCTAAGGAAAGAAACCTCTCTCCCTATTGCTGGTTGTGTGGCCCAGCTCTG
TTCTGTGGTGACGTTTGGTACGGCCGAGTGCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGG
CCATCTGCTCACCCCTGCTCTACTCTACCTGCATGTCCCCTGGAGTCTGCATCATCTTAGTGGGCATG
TCCTACCTGGGTGGATGTGTGAATGCTTGGACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGG
GCCAAATAAAGTCAATCACTTTTTCTGTGACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATT
TTACTTTTGAAATAATTCCAGCTATCTCTTCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCC
ATATCCTACATCTATATCCTCATCACCATCCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTT
CTCCACCTGCACCTCCCACCTCACTGCAGTCACTCTGTTCTATGGGACCATTACCTTCATTTATGTGA
TGCCCAAGTCCAGCTACTCAACTGACCAGAACAAGGTGGTGTCTGTGTTCTACACCGTGGTGATTCCC
ATGTTGAACCCCCTGATCTACAGCCTCAGGAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAG
AATAAAAATATTTTCT**TGA**TGAAACTAGTTAGTTTGAAGA

NOV28j, SNP13373971 of CG53746-04, Protein Sequence	SEQ ID NO: 358	311 aa	MW at 34309.4kD
	SNP Pos: 104		SNP Change: Val to Val

MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF
VDIGYSSSVTPVMLMSFLRKETSLPIAGCVAQLCS**V**VTFGTAECFLLAAMAYDRYVAICSPLLYSTCM
SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG
SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK
VVSVFYTVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS

NOV28k, SNP13373972 of CG53746-04, DNA Sequence	SEQ ID NO: 359	1604 bp
	ORF Start: ATG at 648	ORF Stop: TGA at 1581
	SNP Pos: 1058	SNP Change: C to T

ACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTTGGATATTTGGTGTACTTTTCCTCAATTGTGA
AATCTCTGGGTGGGGACCACAGCTCAGTGTTGAGTTACTGACCTCTTCTTGGTCCTGTGGATTAGCAT
GCAGCTAATCTGTTTGACCTCTGTTTGGAATTCGGAATTCTTAATGACTACACATCTTTGATACAATA
GATGATACCTCAAACATCCTTTTGAACAGCTGTTCTTTTCCATGAGTCTTGGTCTATTCTGACATTTA
TGTCTCCCTTCAATCACGTCTTTGGCCTTAGAAGATTGAGTTACTGGATTCTTTATATATTCTAGTGG
TCATCTCTGAAATGTGCTCAGAGAGCACCTAAATTAACCATCCAATACGAGTTGAGTGTGTTAAGTTA
AAAAAAAAAAAAAAGATTTTTCTGAGTATTCCTGACCTTACATCAGTGAACATTTATGTTTTAAAGTCT
TACATAAGATACTGTGTGTGAAAGCATTTTCTTCCCAAATTTACATGAGTGCCTAAATTGTTATACTT
TTTGGTTAAATAGATATTGGAAAAATAAGTGTGCAATTATAGCATTTAATCCCATTATAATATTCATT
TGTTTTTCTTTCAGTTGGCACCATTCCCAATTTAG**ATG**GGGACTGGAAATGACACCACTGTGGTAGAG
TTTACTCTTTTGGGATTATCCGAGGATACTACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAAT
TTATGTTGTCACCTTAATGGGTAATATCAGCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATA
CACCCATGTACATTTTCCTCTGCCATTTGGCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCT
GTCATGCTCATGAGCTTCCTAAGGAAAGAAACCTCTCTCCCTATTGCTGGTTGTGTGGCCCAGCTCTG
TTCTGTAGTGACGTTTGGTACGGCCGAGTGCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGG
CCATCTGCTCACCCCTGCTCTACTCTACCTGCATGTC**T**CCTGGAGTCTGCATCATCTTAGTGGGCATG
TCCTACCTGGGTGGATGTGTGAATGCTTGGACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGG
GCCAAATAAAGTCAATCACTTTTTCTGTGACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATT
TTACTTTTGAAATAATTCCAGCTATCTCTTCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCC
ATATCCTACATCTATATCCTCATCACCATCCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTT
CTCCACCTGCACCTCCCACCTCACTGCAGTCACTCTGTTCTATGGGACCATTACCTTCATTTATGTGA
TGCCCAAGTCCAGCTACTCAACTGACCAGAACAAGGTGGTGTCTGTGTTCTACACCGTGGTGATTCCC
ATGTTGAACCCCCTGATCTACAGCCTCAGGAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAG

AATAAAAATATTTTCTTGATGAAACTAGTTAGTTTGAAGA			
NOV28k, SNP13373972 of CG53746-04, Protein Sequence	SEQ ID NO: 360	311 aa	MW at 34309.4kD
	SNP Pos: 137		SNP Change: Ser to Ser
MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF VDIGYSSSVTPVMLMSFLRKETSLPIAGCVAQLCSVVTFGTAECFLLAAMAYDRYVAICSPLLYSTCM SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK VVSVFYTVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS			
NOV28l, SNP13382433 of CG53746-04, DNA Sequence	SEQ ID NO: 361		1604 bp
	ORF Start: ATG at 648		ORF Stop: TGA at 1581
	SNP Pos: 1482		SNP Change: A to G
ACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTTGGATATTTGGTGTACTTTTCCTCAATTGTGA AATCTCTGGGTGGGGACCACAGCTCAGTGTTGAGTTACTGACCTCTTCTTGGTCCTGTGGATTAGCAT GCAGCTAATCTGTTTGACCTCTGTTTGGAATTCGGAATTCTTAATGACTACACATCTTTGATACAATA GATGATACCTCAAACATCCTTTTGAACAGCTGTTCTTTTCCATGAGTCTTGGTCTATTCTGACATTTA TGTCTCCCTTCAATCACGTCTTTGGCCTTAGAAGATTGAGTTACTGGATTCTTTATATATTCTAGTGG TCATCTCTGAAATGTGCTCAGAGAGCACCTAAATTAACCATCCAATACGAGTTGAGTGTGTTAAGTTA AAAAAAAAAAAAAGATTTTTCTGAGTATTCCTGACCTTACATCAGTGAACATTTATGTTTTAAAGTCT TACATAAGATACTGTGTGTGAAAGCATTTTCTTCCCAAATTTACATGAGTGCCTAAATTGTTATACTT TTTGGTTAAATAGATATTGGAAAAATAAGTGTGCAATTATAGCATTTAATCCCATTATAATATTCATT TGTTTTTCTTTCAGTTGGCACCATTCCCAATTTAGATGGGGACTGGAAATGACACCACTGTGGTAGAG TTTACTCTTTTGGGATTATCCGAGGATACTACAGTTTGTGCTATTTTATTTCTTGTGTTTCTAGGAAT TTATGTTGTCACCTTAATGGGTAATATCAGCATAATTGTATTGATCAGAAGAAGTCATCATCTTCATA CACCCATGTACATTTTCCTCTGCCATTTGGCCTTTGTAGACATTGGGTACTCCTCATCAGTCACACCT GTCATGCTCATGAGCTTCCTAAGGAAAGAAACCTCTCTCCCTATTGCTGGTTGTGTGGCCCAGCTCTG TTCTGTAGTGACGTTTGGTACGGCCGAGTGCTTCCTGCTGGCTGCCATGGCCTATGATCGCTATGTGG CCATCTGCTCACCCCTGCTCTACTCTACCTGCATGTCCCCTGGAGTCTGCATCATCTTAGTGGGCATG TCCTACCTGGGTGGATGTGTGAATGCTTGGACATTCATTGGCTGCTTATTAAGACTGTCCTTCTGTGG GCCAAATAAAGTCAATCACTTTTTCTGTGACTATTCACCACTTTTGAAGCTTGCTTGTTCCCATGATT TTACTTTTGAAATAATTCCAGCTATCTCTTCTGGATCTATCATTGTGGCCACTGTGTGTGTCATAGCC ATATCCTACATCTATATCCTCATCACCATCCTGAAGATGCACTCCACCAAGGGCCGCCACAAGGCCTT CTCCACCTGCACCTCCCACCTCACTGCAGTCACTCTGTTCTATGGGACCATTACCTTCATTTATGTGA TGCCCAAGTCCAGCTACTCAACTGACCAGAACAAGGTGGTGTCTGTGTTCTACGCCGTGGTGATTCCC ATGTTGAACCCCCTGATCTACAGCCTCAGGAACAAGGAGATTAAGGGGGCTCTGAAGAGAGAGCTTAG AATAAAAATATTTTCTTGATGAAACTAGTTAGTTTGAAGA			
NOV28l, SNP13382433 of CG53746-04, Protein Sequence	SEQ ID NO: 362	311 aa	MW at 34279.4kD
	SNP Pos: 279		SNP Change: Thr to Ala
MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMYIFLCHLAF VDIGYSSSVTPVMLMSFLRKETSLPIAGCVAQLCSVVTFGTAECFLLAAMAYDRYVAICSPLLYSTCM SPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCDYSPLLKLACSHDFTFEIIPAISSG SIIVATVCVIAISYIYILITILKMHSTKGRHKAFSTCTSHLTAVTLFYGTITFIYVMPKSSYSTDQNK VVSVFYAVVIPMLNPLIYSLRNKEIKGALKRELRIKIFS			

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 28B.

Table 28B. Comparison of the NOV28 protein sequences.

```
NOV28a  MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMY
NOV28b  MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMY
NOV28c  MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMY
NOV28d  MGTGNDTTVVEFTLLGLSEDTTVCAILFLVFLGIYVVTLMGNISIIVLIRRSHHLHTPMY
```

```
NOV28a  IFLCHLAFVDIGYSSSVTPVMLMSFLRKETSLPIAGCVAQLCSVVTFGTAECFLLAAMAY
NOV28b  IFLCHLAFVDIGYSSSVTPVMLMSFLRKETSLPVAGCVAQLCSVVTFGTAECFLLAAMAY
NOV28c  IFLCHLAFVDIGYSSSVTPVMLMSFLRKETSLPVAGCVAQLCSVVTFGTAECFLLAAMAY
NOV28d  IFLCHLAFVDIGYSSSVTPVMLMSFLRKETSLPVAGCVAQLCSVVTFGTAECFLLAAMAY

NOV28a  DRYVAICSPLLYSTCMSPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCD
NOV28b  DRYVAICSPLLYSTCMSPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCD
NOV28c  DRYVAICSPLLYSTCMSPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCD
NOV28d  DRYVAICSPLLYSTCMSPGVCIILVGMSYLGGCVNAWTFIGCLLRLSFCGPNKVNHFFCD

NOV28a  YSPLLKLACSHDFTFEIIPAISSGSIIVATVCVIAISYIYILITILKMHSTKGRHKAFST
NOV28b  YSPLLKLACSHDFTFEIIPAISSGSIIVATVCVIAISYIYILITILKMHSTKGRHKAFST
NOV28c  YSPLLKLACSHDFTFEIIPAISSGSIIVATVCVIAISYIYILITILKMHSTKGRHKAFST
NOV28d  YSPLLKLACSHDFTFEIIPAISSGSIIVATVCVIAISYIYILITILKMHSTKGRHKAFST

NOV28a  CTSHLTAVTLFYGTITFIYVMPKSSYSTDQNKVVSVFYTVVIPMLNPLIYSLRNKEIKGA
NOV28b  CTSHLTAVTLFYGTITFIYVMPKSSYSTDQNKVVSVFYTVVIPMLNPLIYSLRNKEIKGA
NOV28c  CTSHLTAVTLFYGTITFIYVMPKSSYSTDQNKVVSVFYTVVIPMLNPLIYSLRNKEIKGA
NOV28d  CTSHLTAVTLFYGTITFIYVMPKSSYSTDQNKVVSVFYTVVIPMLNPLIYSLRNKEIKGA

NOV28a  LKRELRIKIFS
NOV28b  LKRELRIKIFS
NOV28c  LKRELRIKIFS
NOV28d  LKRELRIKIFS

NOV28a  (SEQ ID NO:  340)
NOV28b  (SEQ ID NO:  342)
NOV28c  (SEQ ID NO:  344)
NOV28d  (SEQ ID NO:  346)
```

Further analysis of the NOV28a protein yielded the following properties shown in Table 28C.

Table 28C. Protein Sequence Properties NOV28a

SignalP analysis:	Cleavage site between residues 42 and 43

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 0;  neg.chg 2
      H-region:  length 7;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -5.11
      possible cleavage site: between 38 and 39

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood =-11.46   Transmembrane   23 -  39
```

```
      INTEGRAL    Likelihood = -1.49   Transmembrane  101 - 117
      INTEGRAL    Likelihood = -2.23   Transmembrane  140 - 156
      INTEGRAL    Likelihood = -9.82   Transmembrane  206 - 222
      INTEGRAL    Likelihood = -0.85   Transmembrane  245 - 261
      INTEGRAL    Likelihood = -3.13   Transmembrane  273 - 289
      PERIPHERAL  Likelihood =  1.48 (at 61)
      ALOM score: -11.46   (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 30
      Charge difference:  7.5   C( 3.5) - N(-4.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  5.92
      Hyd Moment(95):  6.18     G content:       2
      D/E content:     2        S/T content:     3
      Score: -7.11

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   6.8%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: IKIF

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none
```

```
Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       22.2 %: vacuolar
       11.1 %: Golgi
       11.1 %: mitochondrial

>> prediction for CG53746-04 is end (k=9)
```

A search of the NOV28a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 28D.

Table 28D. Geneseq Results for NOV28a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV28a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAG66704	Human GPCR2c polypeptide - Homo sapiens, 311 aa. [WO200160865-A2, 23-AUG-2001]	1..311 1..311	311/311 (100%) 311/311 (100%)	e-180
AAU95624	Human olfactory and pheromone G protein-coupled receptor #111 - Homo sapiens, 311 aa. [WO200224726-A2, 28-MAR-2002]	1..311 1..311	310/311 (99%) 311/311 (99%)	e-180
AAG66702	Human GPCR2a polypeptide - Homo sapiens, 311 aa. [WO200160865-A2, 23-AUG-2001]	1..311 1..311	310/311 (99%) 311/311 (99%)	e-180
AAE25068	Human G-protein coupled receptor (GCREC)-8 protein - Homo sapiens, 311 aa. [WO200246230-A2, 13-JUN-2002]	1..311 1..311	309/311 (99%) 311/311 (99%)	e-180
AAG66703	Human GPCR2b polypeptide - Homo sapiens, 311 aa. [WO200160865-A2, 23-AUG-2001]	1..311 1..311	309/311 (99%) 310/311 (99%)	e-180

In a BLAST search of public sequence databases, the NOV28a protein was found to have homology to the proteins shown in the BLASTP data in Table 28E.

Table 28E. Public BLASTP Results for NOV28a				
Protein Accession Number	**Protein/Organism/Length**	**NOV28a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8WZ94	Olfactory receptor 5P3 (Olfactory receptor-like protein JCG1) - Homo sapiens (Human), 311 aa.	1..311 1..311	310/311 (99%) 311/311 (99%)	e-180
CAD37573	Sequence 175 from Patent WO0224726 - Homo sapiens (Human), 311 aa.	1..311 1..311	304/311 (97%) 307/311 (97%)	e-175
Q8NGM2	Seven transmembrane helix receptor - Homo sapiens (Human), 253 aa.	59..311 1..253	252/253 (99%) 253/253 (99%)	e-145
Q8VG42	Olfactory receptor MOR204-6 - Mus musculus (Mouse), 310 aa.	1..303 1..303	240/303 (79%) 267/303 (87%)	e-138
Q8VEZ0	Olfactory receptor MOR204-32 - Mus musculus (Mouse), 312 aa.	1..303 1..303	236/303 (77%) 266/303 (86%)	e-136

PFam analysis indicates that the NOV28a protein contains the domains shown in the Table 28F.

Table 28F. Domain Analysis of NOV28a			
Pfam Domain	**NOV28a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	41..290	56/276 (20%) 179/276 (65%)	2.5e-34

5

Example 29.

The NOV29 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 29A.

Table 29A. NOV29 Sequence Analysis		
NOV29a, CG53767-02 DNA Sequence	SEQ ID NO: 363	1378 bp
	ORF Start: ATG at 473	ORF Stop: TAA at 1331
TGAAATTGGAGAGTCATACAAAATCCAGAATGGAGAAAAGAGAGATAAACAGTGCTTTTGGGGTAAAA		
TCCAGGTATAGGTATTTGGGCCCACTCTATGTCATCTATACATATTATCATATTAGAATCATTCTTCT		
CTTCACTCAGGCTCCTGACCTTGACCTCTTTCCAGGTGCATTTAACTCCCTTTCCTCAGGTTCTCTAA		

GAGTTGCCTGACACTTGCCCAGGTCCCCTGAATAAAATACAGGTAAATCAGTGCATAAGAAGCTTTAT
ACAGTGATCTCAGACTTGTGACCAAGAGCCCCTCTGCACCATTGATAGACAGTGGCCGTGCCCACCAG
GAGGATGGGCAATCACACTGCAGTGAGCCTATTCCTTCTGTGGGGATTTTCCAGTTTTTCAGACCTGC
AGAGTCTACTTTTTGTGGTGATTCTCTTCTACATGTGACCATCCTAGCTGCAAACGTGTCCATA**ATGG**
GGGCCATCAAGCTCAGCCACAACCTTCACACTCCTATGTACTTTTTCCTCTGTGGCCTGTCCTTTTCA
GAAACTTGTACCACTGTGGTAGTAATCCCTCGCATGTTGGTGGACTTTCTATCAGAGAGCAAGACCAT
TTCTCTTCCTGAGTGTGCCACACAGATGTTTTTCTTTCTGGGCTTTGCATCCAACAACTGTTTCATCA
TGGCCGCTATGTCCTACGACCGCTACACGGCCATCCACAACCCACTGCAGTACCACACCCTTATGACA
AGAAAGATCTGCTTGCAGATGATGATGGCTTCTTGGATGGTTGGGTTCCTGTTTTCTCTGTGCATCAT
CGTCACTGTATTCAACTTGTCTCTTTGCGACTTGAACACTATCCAGCACTATTTCTGTGATATCTCAC
CAGTGGTCTCCCTTGCTTGTAATTACACTTTCTATCATGAAATGGCTATTTTTGTGCTCTCTGCCTTT
GTGTTGGTGGGCAGCTGTATTTTAATTATGATTTCCTATGTCTTCATTGTGTTCATAGTCATAAAGAT
GCCCTCTGCAAAGGGGAGGTCTAAGGCCTTCTCAACTTGCTCCTCCCACCTCACTGTTGTGTCCATAC
ACTATGGATTTGCTTGCTTTGTCTATTTGAGGCCCAAGAACAGCAACTCCTTCGATGAAGACATGCCG
ACGGCCATGATATATACAATACTGATGCCTCTGCTTAACCCCATCGTGTACAGTCTGAGAAACAAAGA
AATGCAGATAGCCCTAAGAAAAACACTAGGCAGTGTATTTGGGGTTTTCCCTCAGAAGACAAAAAAAG
AGCCTGAACATTTAAAAAAATTACACAGCATTGATAAA**TAA**AGGTGAGAAAAGCGAAAAAAAAAAAAA
AAAAAAAAAAAAAAAGGG

NOV29a, CG53767-02 Protein Sequence	SEQ ID NO: 364	286 aa	MW at 32538.4kD

MGAIKLSHNLHTPMYFFLCGLSFSETCTTVVVIPRMLVDFLSESKTISLPECATQMFFFLGFASNNCF
IMAAMSYDRYTAIHNPLQYHTLMTRKICLQMMMASWMVGFLFSLCIIVTVFNLSLCDLNTIQHYFCDI
SPVVSLACNYTFYHEMAIFVLSAFVLVGSCILIMISYVFIVFIVIKMPSAKGRSKAFSTCSSHLTVVS
IHYGFACFVYLRPKNSNSFDEDMPTAMIYTILMPLLNPIVYSLRNKEMQIALRKTLGSVFGVFPQKTK
KEPEHLKKLHSIDK

NOV29b, CG53767-01 DNA Sequence	SEQ ID NO: 365	1028 bp
	ORF Start: ATG at 23	ORF Stop: TAA at 1007

AGTGGCCGTGCCCACCAGGAGG**ATG**GGCAATCACACTGCAGTGAGCCTATTCCTTCTGTGGGGATTTT
CCAGTTTTTCAGACCTGCAGAGTCTACTTTGTGGTGATTCTCTTCTACATGTGACCATCCTAGCTGCA
AACGTGTCCATAATGGGGGCCATCAAGCTCAGCCACAACCTTCACACTCCTATGTACTTTTTCCTCTG
TGGCCTGTCCTTTTCAGAAACTTGTACCACTGTGGTAGTAATCCCTCGCATGTTGGTGGACTTTCTAT
CAGAGAGCAAGACCATTTCTCTTCCTGAGTGTGCCACACAGATGTTTTTCTTTCTGGGCTTTGCATCC
AACAACTGTTTCATCATGGCCGCTATGTCCTACGACCGCTACACGGCCATCCACAACCCACTGCAGTA
CCACACCCTTATGACAAGAAAGATCTGCTTGCAGATGATGATGGCTTCTTGGATGGTTGGGTTCCTGT
TTTCTCTGTGCATCATCGTCACTGTATTCAACTTGTCTCTTTGCGACTTGAACACTATCCAGCACTAT
TTCTGTGATATCTCACCAGTGGTCTCCCTTGCTTGTAATTACACTTTCTATCATGAAATGGCTATTTT
TGTGCTCTCTGCCTTTGTGTTGGTGGGCAGCTGTATTTTAATTATGATTTCCTATGTCTTCATTGTGT
TCATAGTCATAAAGATGCCCTCTGCAAAGGGGAGGTCTAAGGCCTTCTCAACTTGCTCCTCCCACCTC
ACTGTTGTGTCCATACACTATGGATTTGCTTGCTTTGTCTATTTGAGGCCCAAGAACAGCAACTCCTT
CGATGAAGACATGCTGACGGCCATGATATATACAATACTGATGCCTCTGCTTAACCCCATCGTGTACA
GTCTGAGAAACAAAGAAATGCAGATAGCCCTAAGAAAAACACTAGGCAGTGTATTTGGGGTTTTCCCT
CAGAAGACAAAAAAAGAGCCTGAACATTTAAAAAAATTACACAGCATTGATAAA**TAA**AGGTGAGAAAA
GTGGAGTA

NOV29b, CG53767-01 Protein Sequence	SEQ ID NO: 366	328 aa	MW at 37015.5kD

MGNHTAVSLFLLWGFSSFSDLQSLLCGDSLLHVTILAANVSIMGAIKLSHNLHTPMYFFLCGLSFSET
CTTVVVIPRMLVDFLSESKTISLPECATQMFFFLGFASNNCFIMAAMSYDRYTAIHNPLQYHTLMTRK
ICLQMMMASWMVGFLFSLCIIVTVFNLSLCDLNTIQHYFCDISPVVSLACNYTFYHEMAIFVLSAFVL
VGSCILIMISYVFIVFIVIKMPSAKGRSKAFSTCSSHLTVVSIHYGFACFVYLRPKNSNSFDEDMLTA
MIYTILMPLLNPIVYSLRNKEMQIALRKTLGSVFGVFPQKTKKEPEHLKKLHSIDK

NOV29c, SNP13382437 of CG53767-02, DNA Sequence	SEQ ID NO: 367	1378 bp
	ORF Start: ATG at 473	ORF Stop: TAA at 1331
	SNP Pos: 961	SNP Change: G to A

TGAAATTGGAGAGTCATACAAAATCCAGAATGGAGAAAAGAGAGATAAACAGTGCTTTTGGGGTAAAA

TCCAGGTATAGGTATTTGGGCCCACTCTATGTCATCTATACATATTATCATATTAGAATCATTCTTCT
CTTCACTCAGGCTCCTGACCTTGACCTCTTTCCAGGTGCATTTAACTCCCTTTCCTCAGGTTCTCTAA
GAGTTGCCTGACACTTGCCCAGGTCCCCTGAATAAAATACAGGTAAATCAGTGCATAAGAAGCTTTAT
ACAGTGATCTCAGACTTGTGACCAAGAGCCCCTCTGCACCATTGATAGACAGTGGCCGTGCCCACCAG
GAGGATGGGCAATCACACTGCAGTGAGCCTATTCCTTCTGTGGGGATTTTCCAGTTTTTCAGACCTGC
AGAGTCTACTTTTTGTGGTGATTCTCTTCTACATGTGACCATCCTAGCTGCAAACGTGTCCATA**ATG**G
GGGCCATCAAGCTCAGCCACAACCTTCACACTCCTATGTACTTTTTCCTCTGTGGCCTGTCCTTTTCA
GAAACTTGTACCACTGTGGTAGTAATCCCTCGCATGTTGGTGGACTTTCTATCAGAGAGCAAGACCAT
TTCTCTTCCTGAGTGTGCCACACAGATGTTTTTCTTTCTGGGCTTTGCATCCAACAACTGTTTCATCA
TGGCCGCTATGTCCTACGACCGCTACACGGCCATCCACAACCCACTGCAGTACCACACCCTTATGACA
AGAAAGATCTGCTTGCAGATGATGATGGCTTCTTGGATGGTTGGGTTCCTGTTTTCTCTGTGCATCAT
CGTCACTGTATTCAACTTGTCTCTTTGCGACTTGAACACTATCCAGCACTATTTCTGTGATATCTCAC
CAGTGGTCTCCCTTGCTTGTAATTACACTTTCTATCATGAAATGGCTATTTTTGTGCTCTCTGCCTTT
GTGTTGGT**A**GGCAGCTGTATTTTAATTATGATTTCCTATGTCTTCATTGTGTTCATAGTCATAAAGAT
GCCCTCTGCAAAGGGGAGGTCTAAGGCCTTCTCAACTTGCTCCTCCCACCTCACTGTTGTGTCCATAC
ACTATGGATTTGCTTGCTTTGTCTATTTGAGGCCCAAGAACAGCAACTCCTTCGATGAAGACATGCCG
ACGGCCATGATATATACAATACTGATGCCTCTGCTTAACCCCATCGTGTACAGTCTGAGAAACAAAGA
AATGCAGATAGCCCTAAGAAAAACACTAGGCAGTGTATTTGGGGTTTTCCCTCAGAAGACAAAAAAAG
AGCCTGAACATTTAAAAAAATTACACAGCATTGATAAAT**AA**AGGTGAGAAAAGCGAAAAAAAAAAAAA
AAAAAAAAAAAAAAAGGG

NOV29c, SNP13382437 of CG53767-02, Protein Sequence	SEQ ID NO: 368	286 aa	MW at 32538.4kD
	SNP Pos: 163		SNP Change: Val to Val

MGAIKLSHNLHTPMYFFLCGLSFSETCTTVVVIPRMLVDFLSESKTISLPECATQMFFFLGFASNNCF
IMAAMSYDRYTAIHNPLQYHTLMTRKICLQMMMASWMVGFLFSLCIIVTVFNLSLCDLNTIQHYFCDI
SPVVSLACNYTFYHEMAIFVLSAFVL**V**GSCILIMISYVFIVFIVIKMPSAKGRSKAFSTCSSHLTVVS
IHYGFACFVYLRPKNSNSFDEDMPTA**M**IYTILMPLLNPIVYSLRNKEMQIALRKTLGSVFGVFPQKTK
KEPEHLKKLHSIDK

NOV29d, SNP13382436 of CG53767-02, DNA Sequence	SEQ ID NO: 369	1378 bp
	ORF Start: ATG at 473	ORF Stop: TAA at 1331
	SNP Pos: 1283	SNP Change: A to G

TGAAATTGGAGAGTCATACAAAATCCAGAATGGAGAAAAGAGAGATAAACAGTGCTTTTGGGGTAAAA
TCCAGGTATAGGTATTTGGGCCCACTCTATGTCATCTATACATATTATCATATTAGAATCATTCTTCT
CTTCACTCAGGCTCCTGACCTTGACCTCTTTCCAGGTGCATTTAACTCCCTTTCCTCAGGTTCTCTAA
GAGTTGCCTGACACTTGCCCAGGTCCCCTGAATAAAATACAGGTAAATCAGTGCATAAGAAGCTTTAT
ACAGTGATCTCAGACTTGTGACCAAGAGCCCCTCTGCACCATTGATAGACAGTGGCCGTGCCCACCAG
GAGGATGGGCAATCACACTGCAGTGAGCCTATTCCTTCTGTGGGGATTTTCCAGTTTTTCAGACCTGC
AGAGTCTACTTTTTGTGGTGATTCTCTTCTACATGTGACCATCCTAGCTGCAAACGTGTCCATA**ATG**G
GGGCCATCAAGCTCAGCCACAACCTTCACACTCCTATGTACTTTTTCCTCTGTGGCCTGTCCTTTTCA
GAAACTTGTACCACTGTGGTAGTAATCCCTCGCATGTTGGTGGACTTTCTATCAGAGAGCAAGACCAT
TTCTCTTCCTGAGTGTGCCACACAGATGTTTTTCTTTCTGGGCTTTGCATCCAACAACTGTTTCATCA
TGGCCGCTATGTCCTACGACCGCTACACGGCCATCCACAACCCACTGCAGTACCACACCCTTATGACA
AGAAAGATCTGCTTGCAGATGATGATGGCTTCTTGGATGGTTGGGTTCCTGTTTTCTCTGTGCATCAT
CGTCACTGTATTCAACTTGTCTCTTTGCGACTTGAACACTATCCAGCACTATTTCTGTGATATCTCAC
CAGTGGTCTCCCTTGCTTGTAATTACACTTTCTATCATGAAATGGCTATTTTTGTGCTCTCTGCCTTT
GTGTTGGTGGGCAGCTGTATTTTAATTATGATTTCCTATGTCTTCATTGTGTTCATAGTCATAAAGAT
GCCCTCTGCAAAGGGGAGGTCTAAGGCCTTCTCAACTTGCTCCTCCCACCTCACTGTTGTGTCCATAC
ACTATGGATTTGCTTGCTTTGTCTATTTGAGGCCCAAGAACAGCAACTCCTTCGATGAAGACATGCCG
ACGGCCATGATATATACAATACTGATGCCTCTGCTTAACCCCATCGTGTACAGTCTGAGAAACAAAGA
AATGCAGATAGCCCTAAGAAAAACACTAGGCAGTGTATTTGGGGTTTTCCCTCAGAAG**G**CAAAAAAAG
AGCCTGAACATTTAAAAAAATTACACAGCATTGATAAAT**AA**AGGTGAGAAAAGCGAAAAAAAAAAAAA
AAAAAAAAAAAAAAAGGG

NOV29d, SNP13382436 of CG53767-02, Protein Sequence	SEQ ID NO: 370	286 aa	MW at 32508.4kD
	SNP Pos: 271		SNP Change: Thr to Ala

MGAIKLSHNLHTPMYFFLCGLSFSETCTTVVVIPRMLVDFLSESKTISLPECATQMFFFLGFASNNCF
IMAAMSYDRYTAIHNPLQYHTLMTRKICLQMMMASWMVGFLFSLCIIVTVFNLSLCDLNTIQHYFCDI

SPVVSLACNYTFYHEMAIFVLSAFVLVGSCILIMISYVFIVFIVIKMPSAKGRSKAFSTCSSHLTVVS
IHYGFACFVYLRPKNSNSFDEDMPTAMIYTILMPLLNPIVYSLRNKEMQIALRKTLGSVFGVFPQKAK
KEPEHLKKLHSIDK

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 29B.

Table 29B. Comparison of the NOV29 protein sequences.

```
NOV29a  ----------------------------------------MGAIKLSHNLHTPMYFFL
NOV29b  MGNHTAVSLFLLWGFSSFSDLQSLLCGDSLLHVTILAANVSIMGAIKLSHNLHTPMYFFL

NOV29a  CGLSFSETCTTVVVIPRMLVDFLSESKTISLPECATQMFFFLGFASNNCFIMAAMSYDRY
NOV29b  CGLSFSETCTTVVVIPRMLVDFLSESKTISLPECATQMFFFLGFASNNCFIMAAMSYDRY

NOV29a  TAIHNPLQYHTLMTRKICLQMMMASWMVGFLFSLCIIVTVFNLSLCDLNTIQHYFCDISP
NOV29b  TAIHNPLQYHTLMTRKICLQMMMASWMVGFLFSLCIIVTVFNLSLCDLNTIQHYFCDISP

NOV29a  VVSLACNYTFYHEMAIFVLSAFVLVGSCILIMISYVFIVFIVIKMPSAKGRSKAFSTCSS
NOV29b  VVSLACNYTFYHEMAIFVLSAFVLVGSCILIMISYVFIVFIVIKMPSAKGRSKAFSTCSS

NOV29a  HLTVVSIHYGFACFVYLRPKNSNSFDEDMPTAMIYTILMPLLNPIVYSLRNKEMQIALRK
NOV29b  HLTVVSIHYGFACFVYLRPKNSNSFDEDMLTAMIYTILMPLLNPIVYSLRNKEMQIALRK

NOV29a  TLGSVFGVFPQKTKKEPEHLKKLHSIDK
NOV29b  TLGSVFGVFPQKTKKEPEHLKKLHSIDK

NOV29a  (SEQ ID NO:  364)
NOV29b  (SEQ ID NO:  366)
```

Further analysis of the NOV29a protein yielded the following properties shown in Table 29C.

Table 29C. Protein Sequence Properties NOV29a

SignalP analysis:	Cleavage site between residues 25 and 26

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 5;  pos.chg 1;  neg.chg 0
      H-region:  length 19;  peak value  10.63
      PSG score:   6.23

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -2.51
      possible cleavage site: between 29 and 30

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   4
```

```
      INTEGRAL    Likelihood = -0.43   Transmembrane   17 -  33
      INTEGRAL    Likelihood = -7.17   Transmembrane  105 - 121
      INTEGRAL    Likelihood = -1.33   Transmembrane  148 - 164
      INTEGRAL    Likelihood =-12.52   Transmembrane  165 - 181
      PERIPHERAL  Likelihood =  0.53 (at 231)
      ALOM score: -12.52  (number of TMSs: 4)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 24
      Charge difference: -0.5   C(-1.0) - N(-0.5)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  4.10
      Hyd Moment(95):  4.80     G content:       2
      D/E content:     1        S/T content:     4
      Score: -5.17

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 45  PRM|LV

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: PQKTKKE (4) at  268
      bipartite: RKTLGSVFGVFPQKTKK at  257
      content of basic residues:   7.7%
      NLS Score:  0.36

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none
```

```
Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       44.4 %: mitochondrial

>> prediction for CG53767-02 is end (k=9)
```

A search of the NOV29a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 29D.

Table 29D. Geneseq Results for NOV29a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV29a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABP72224	Human G-protein coupled receptor GCREC-18 (olfactory receptor) - Homo sapiens, 329 aa. [WO2003000859-A2, 03-JAN-2003]	1..286 44..329	285/286 (99%) 285/286 (99%)	e-166
ABP95890	Human GPCR polypeptide SEQ ID NO 590 - Homo sapiens, 286 aa. [WO200216548-A2, 28-FEB-2002]	1..286 1..286	285/286 (99%) 285/286 (99%)	e-166
AAG66705	Human GPCR3 polypeptide - Homo sapiens, 328 aa. [WO200160865-A2, 23-AUG-2001]	1..286 43..328	285/286 (99%) 285/286 (99%)	e-166
AAG71567	Human olfactory receptor polypeptide, SEQ ID NO: 1248 - Homo sapiens, 316 aa. [WO200127158-A2, 19-APR-2001]	1..271 46..316	270/271 (99%) 270/271 (99%)	e-156
ABG59965	Human DITHP polypeptide #23 - Homo sapiens, 328 aa. [WO200220754-A2, 14-MAR-2002]	1..258 71..328	257/258 (99%) 257/258 (99%)	e-149

In a BLAST search of public sequence databases, the NOV29a protein was found to have homology to the proteins shown in the BLASTP data in Table 29E.

Table 29E. Public BLASTP Results for NOV29a				
Protein Accession Number	**Protein/Organism/Length**	**NOV29a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8NGM4	Seven transmembrane helix receptor - Homo sapiens (Human), 339 aa.	1..286 54..339	285/286 (99%) 285/286 (99%)	e-165
Q8VF20	Olfactory receptor MOR267-14 - Mus musculus (Mouse), 321 aa.	1..278 44..321	187/278 (67%) 226/278 (81%)	e-109
CAD42418	Sequence 75 from Patent WO0212343 - Homo sapiens (Human), 314 aa.	1..261 46..306	136/261 (52%) 182/261 (69%)	3e-75
CAD42393	Sequence 23 from Patent WO0212343 - Homo sapiens (Human), 313 aa.	1..261 44..304	134/261 (51%) 177/261 (67%)	1e-73
Q8NGX3	Seven transmembrane helix receptor - Homo sapiens (Human), 314 aa.	1..261 47..306	132/261 (50%) 186/261 (70%)	4e-73

PFam analysis indicates that the NOV29a protein contains the domains shown in the Table 29F.

Table 29F. Domain Analysis of NOV29a			
Pfam Domain	**NOV29a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	3..245	47/278 (17%) 169/278 (61%)	3.1e-19

Example 30.

The NOV30 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 30A.

Table 30A. NOV30 Sequence Analysis		
NOV30a, CG53776-02 DNA Sequence	SEQ ID NO: 371	1072 bp
	ORF Start: ATG at 12	ORF Stop: TAG at 1047
TTAACATTTTAATGTGCTGTTCTCATTTGGGTTCATTTAGTCAGCAGCTACTTCGTCTCATGAATTCC		

CTGAAGGACGGGAATCACACCGCTCTGACGGGGTTCATCCTATTGGGCTTAACAGATGATCCAATCCT
TCGAGTCATCCTCTTCATGATCATCCTATGCATCTACCTGGTAACCATATCTGGTAATCTCAGCATAA
TTATTCTTATCAGAATTTCTTCTCAGCTCCATCATCCTATGTATTTCTTTCTGAGCCACTTGGCTTTT
GCTGACATGGCCTATTCATCTTCTGTCACACCCAACATGCTTGTAAACTTCCTGGTGGAGAGAAATAC
AGTCTCCTACCTTGGATGTGCCATCCAGCTTGGTTCAGCGGCTTTCTTTGCAACAGTCGAATGCGTCT
TTCTGGCTGCCATGGCCTATGACCGCTTTGTGGCAATTTGCAGTCCACTGCTTTATTCAACCAAAATG
TCCACACAAGTCAGTGTCCAGCTACTCTTAGTAGTTTACATAGCTGGTTTTCTCATTGCTGTCTCCTA
TACTACTTCCTTCTATTTTTTACTCTTCTGTGGACCAAATCAAGTCAATCATTTTTTCTGTGATTTCG
CTCCCTTACTTGAACTCTCCTGTTCTGATATCAGTGTCTCCACAGTTGTTCTCTCATTTTCTTCTGGA
TCCATCATTGTGGTCACTGTGTGTGTCATAGCCGTCTGCTACATCTATATCCTCATCACCATCCTGAA
GATGCGCTCCACTGAGGGGCACCACAAGGCCTTCTCCACCTGCACTTCCCACCTCACTGTGGTTACCC
TGTTCTATGGGACCATTACCTTCATTTATGTGATGCCCAATTTTAGCTACTCAACTGACCAGAACAAG
GTGGTGTCTGTGTTGTACACAGTGGTGATTCCCATGTTGAACCCCTTGATCTACAGCCTCAGGAACAA
GGAGATTAAGGGGGCTCTGAAGAGAGAGCTTGTTAGAAAAATACTTTCTCATGATGCTTGTTATTTTA
GTAGAACTTCAAATAATGATATTACATAGAACCCTATCTCTTCTCTTGAGAA

NOV30a, CG53776-02 Protein Sequence	SEQ ID NO: 372	345 aa	MW at 38430.9kD

MCCSHLGSFSQQLLRLMNSLKDGNHTALTGFILLGLTDDPILRVILFMIILCIYLVTISGNLSIIILI
RISSQLHHPMYFFLSHLAFADMAYSSSVTPNMLVNFLVERNTVSYLGCAIQLGSAAFFATVECVFLAA
MAYDRFVAICSPLLYSTKMSTQVSVQLLLVVYIAGFLIAVSYTTSFYFLLFCGPNQVNHFFCDFAPLL
ELSCSDISVSTVVLSFSSGSIIVVTVCVIAVCYIYILITILKMRSTEGHHKAFSTCTSHLTVVTLFYG
TITFIYVMPNFSYSTDQNKVVSVLYTVVIPMLNPLIYSLRNKEIKGALKRELVRKILSHDACYFSRTS
NNDIT

NOV30b, CG53776-03 DNA Sequence	SEQ ID NO: 373	1050 bp
	ORF Start: at 207	ORF Stop: end of sequence

TTCTCAAGAGAAGAGATAGGGTTCTATGTAATATCATTATTTGAAGTTCTACTAAAATAACAAGCATC
ATGAGAAAGTATTTTTCTAACAAGCTCTCTCTTCAGAGCCCCCTTAATCTCCTTGTTCCTGAGGCTGT
AGATCAGGGGGTTCAACATGGGAATCACCACTGTGTACAACACAGACACCACCTTGTTCTGGTCAGTT
GAGTAGCTAAAATTGGGCATCACATAAATGAAGGTAATGGTCCCATAGAACAGGGTAACCACAGTGAG
GTGGGAAGTGCAGGTGGAGAAGGCCTTGTGGTGCCCCTCAGTGGAGCGCATCTTCAGGATGGTGATGA
GGATATAGATGTAGCAGACGGCTATGACACACACAGTGACCACAATGATGGATCCAGAAGAAAATGAG
AGAACAACTGTGGAGACACTGATATCAGAACAGGAGAGTTCAAGTAAGGGAGCGAAATCACAGAAAAA
ATGATTGACTTGGTCCACAGAAGAGTAAAAATAGAAGGAAGTAGTATAGGAGACAGCAATGAG
AAAACCAGCTATGTAAACTACTAAGAGTAGCTGGACACTGACTTGTGTGGACATTTTGGTTGAATAAA
GCAGTGGACTGCAAATTGCCACAAAGCGGTCATAGGCCATGGCAGCCAGAAGGACGCATTCGACTGTT
GCAAAGAAAGCCGCTGAACCAAGCTGGATGGCACATCCAAGGTAGGAGACTGTATTTCTCTCCACCAG
GAAGTTTACAAGCATGTTGGGTGTGACAGAAGATGAATAGGCCATGTCAGCAAAAGCCAAGTGGCTCA
GAAAGAAATACATAGGATGATGGAGCTGAGAAGAAATTCTGATAAGAATAATTATGCTGAGATTACCA
GATAGGATGATCATGAAGAGGATGACTCGAAGGATTGGATCATCTGTTAAGCCCAATAGGATGAACCC
CGTCAGAGCGGTGTGATTCCCGTCCTTCAGGGAATTCATGAGACGAAGTAGCTGCTGACTAAATGAAC
CCTAATGAGAACAGCACATTAAAATGTTAA

NOV30b, CG53776-03 Protein Sequence	SEQ ID NO: 374	338 aa	MW at 37640.9kD

MCCSHLGSFSQQLLRLMNSLKDGNHTALTGFILLGLTDDPILRVILFMIILCGNLSIIILIRISSQLH
HPMYFFLSHLAFADMAYSSSVTPNMLVNFLVERNTVSYLGCAIQLGSAAFFATVECVFLAAMAYDRFV
AICSPLLYSTKMSTQVSVQLLLVVYIAGFLIAVSYTTSFYFLLFCGPNQVNHFFCDFAPLLELSCSDI
SVSTVVLSFSSGSIIVVTVCVIAVCYIYILITILKMRSTEGHHKAFSTCTSHLTVVTLFYGTITFIYV
MPNFSYSTDQNKVVSVLYTVVIPMLNPLIYSLRNKEIKGALKRELVRKILSHDACYFSRTSNNDIT

NOV30c, CG53776-01 DNA Sequence	SEQ ID NO: 375	1070 bp
	ORF Start: ATG at 27	ORF Stop: TAG at 1041

GAAGCCATCAAATTTATAACATTTTA**ATG**TGCTGTTCTCATTTGGGTTCATTTAGTCAGCAGCTACTT
CGTCTCATGAATTCCCTGAAGGACGGGAATCACACCGCTCTGACGGGGTTCATCCTATTGGGCTTAAC
AGATGATCCAATCCTTCGAGTCATCCTCTTCATGATCATCCTATCTGGTAATCTCAGCATAATTATTC
TTATCAGAATTTCTTCTCAGCTCCATCATCCTATGTATTTCTTTCTGAGCCACTTGGCTTTTGCTGAC

ATGGCCTATTCATCTTCTGTCACACCCAACATGCTTGTAAACTTCCTGGTGGAGAGAAATACAGTCTC
CTACCTTGGATGTGCCATCCAGCTTGGTTCAGCGGCTTTCTTTGCAACAGTCGAATGCGTCCTTCTGG
CTGCCATGGCCTATGACCGCTTTGTGGCAATTTGCAGTCCACTGCTTTATTCAACCAAAATGTCCACA
CAAGTCAGTGTCCAGCTACTCTTAGTAGTTTACATAGCTGGTTTTCTCATTGCTGTCTCCTATACTAC
TTCCTTCTATTTTTTACTCTTCTGTGGACCAAATCAAGTCAATCATTTTTTCTGTGATTTCGCTCCCT
TACTTGAACTCTCCTGTTCTGATATCAGTGTCTCCACAGTTGTTCTCTCATTTTCTTCTGGATCCATC
ATTGTGGTCACTGTGTGTGTCATAGCCGTCTGCTACATCTATATCCTCATCACCATCCTGAAGATGCG
CTCCACTGAGGGGCACCACAAGGCCTTCTCCACCTGCACTTCCCACCTCACTGTGGTTACCCTGTTCT
ATGGGACCATTACCTTCATTTATGTGATGCCCAATTTTAGCTACTCAACTGACCAGAACAAGGTGGTG
TCTGTGTTGTACACAGTGGTGATTCCCATGTTGAACCCCCTGATCTACAGCCTCAGGAACAAGGAGAT
TAAGGGGGCTCTGAAGAGAGAGCTTGTTAGAAAAATACTTTCTCATGATGCTTGTTATTTTAGTAGAA
CTTCAAATAATGATATTACATAGAACCCTATCTCTTCTCTTGAGAATACT

NOV30c, CG53776-01 Protein Sequence	SEQ ID NO: 376	338 aa	MW at 37590.8kD

MCCSHLGSFSQQLLRLMNSLKDGNHTALTGFILLGLTDDPILRVILFMIILSGNLSIIILIRISSQLH
HPMYFFLSHLAFADMAYSSSVTPNMLVNFLVERNTVSYLGCAIQLGSAAFFATVECVLLAAMAYDRFV
AICSPLLYSTKMSTQVSVQLLLVVYIAGFLIAVSYTTSFYFLLFCGPNQVNHFFCDFAPLLELSCSDI
SVSTVVLSFSSGSIIVVTVCVIAVCYIYILITILKMRSTEGHHKAFSTCTSHLTVVTLFYGTITFIYV
MPNFSYSTDQNKVVSVLYTVVIPMLNPLIYSLRNKEIKGALKRELVRKILSHDACYFSRTSNNDIT

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 30B.

Table 30B. Comparison of the NOV30 protein sequences.

```
NOV30a  MCCSHLGSFSQQLLRLMNSLKDGNHTALTGFILLGLTDDPILRVILFMIILCIYLVTISG
NOV30b  MCCSHLGSFSQQLLRLMNSLKDGNHTALTGFILLGLTDDPILRVILFMIILC-------G
NOV30c  MCCSHLGSFSQQLLRLMNSLKDGNHTALTGFILLGLTDDPILRVILFMIILS-------G

NOV30a  NLSIIILIRISSQLHHPMYFFLSHLAFADMAYSSSVTPNMLVNFLVERNTVSYLGCAIQL
NOV30b  NLSIIILIRISSQLHHPMYFFLSHLAFADMAYSSSVTPNMLVNFLVERNTVSYLGCAIQL
NOV30c  NLSIIILIRISSQLHHPMYFFLSHLAFADMAYSSSVTPNMLVNFLVERNTVSYLGCAIQL

NOV30a  GSAAFFATVECVFLAAMAYDRFVAICSPLLYSTKMSTQVSVQLLLVVYIAGFLIAVSYTT
NOV30b  GSAAFFATVECVFLAAMAYDRFVAICSPLLYSTKMSTQVSVQLLLVVYIAGFLIAVSYTT
NOV30c  GSAAFFATVECVLLAAMAYDRFVAICSPLLYSTKMSTQVSVQLLLVVYIAGFLIAVSYTT

NOV30a  SFYFLLFCGPNQVNHFFCDFAPLLELSCSDISVSTVVLSFSSGSIIVVTVCVIAVCYIYI
NOV30b  SFYFLLFCGPNQVNHFFCDFAPLLELSCSDISVSTVVLSFSSGSIIVVTVCVIAVCYIYI
NOV30c  SFYFLLFCGPNQVNHFFCDFAPLLELSCSDISVSTVVLSFSSGSIIVVTVCVIAVCYIYI

NOV30a  LITILKMRSTEGHHKAFSTCTSHLTVVTLFYGTITFIYVMPNFSYSTDQNKVVSVLYTVV
NOV30b  LITILKMRSTEGHHKAFSTCTSHLTVVTLFYGTITFIYVMPNFSYSTDQNKVVSVLYTVV
NOV30c  LITILKMRSTEGHHKAFSTCTSHLTVVTLFYGTITFIYVMPNFSYSTDQNKVVSVLYTVV

NOV30a  IPMLNPLIYSLRNKEIKGALKRELVRKILSHDACYFSRTSNNDIT
NOV30b  IPMLNPLIYSLRNKEIKGALKRELVRKILSHDACYFSRTSNNDIT
NOV30c  IPMLNPLIYSLRNKEIKGALKRELVRKILSHDACYFSRTSNNDIT

NOV30a  (SEQ ID NO:  372)
NOV30b  (SEQ ID NO:  374)
NOV30c  (SEQ ID NO:  376)
```

Further analysis of the NOV30a protein yielded the following properties shown in Table 30C.

Table 30C. Protein Sequence Properties NOV30a

SignalP analysis:	Cleavage site between residues 61 and 62

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 14;  peak value   5.07
      PSG score:   0.67

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -5.40
      possible cleavage site: between 57 and 58

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood =-10.56   Transmembrane   41 -  57
      INTEGRAL    Likelihood = -2.07   Transmembrane  122 - 138
      INTEGRAL    Likelihood = -7.70   Transmembrane  159 - 175
      INTEGRAL    Likelihood =-12.68   Transmembrane  225 - 241
      INTEGRAL    Likelihood = -2.13   Transmembrane  264 - 280
      INTEGRAL    Likelihood = -3.66   Transmembrane  292 - 308
      PERIPHERAL  Likelihood =  1.91 (at 58)
      ALOM score: -12.68  (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 48
      Charge difference:  3.0   C( 2.0) - N(-1.0)
      C > N: C-terminal side will be inside

>>>Caution: Inconsistent mtop result with signal peptide
>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:        1       Hyd Moment(75):  3.87
      Hyd Moment(95):  1.84     G content:       1
      D/E content:     1        S/T content:     4
      Score: -4.16

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 25  LRL|MN

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
```

```
      content of basic residues:   5.5%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       44.4 %: endoplasmic reticulum
       22.2 %: vacuolar
       11.1 %: Golgi
       11.1 %: vesicles of secretory system
       11.1 %: mitochondrial

>> prediction for CG53776-02 is end (k=9)
```

A search of the NOV30a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 30D.

Table 30D. Geneseq Results for NOV30a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG66706	Human GPCR4 polypeptide - Homo sapiens, 338 aa. [WO200160865-A2, 23-AUG-2001]	1..345 1..338	337/345 (97%) 337/345 (97%)	0.0
ABU11199	Human G-protein coupled receptor GCREC-52 - Homo sapiens, 322 aa. [WO200279448-A2, 10-OCT-2002]	17..345 1..322	321/329 (97%) 321/329 (97%)	e-180
AAU85166	G-coupled olfactory receptor #27 - Homo sapiens, 322 aa. [WO200198526-A2, 27-DEC-2001]	17..345 1..322	321/329 (97%) 321/329 (97%)	e-180
AAU95576	Human olfactory and pheromone G protein-coupled receptor #63 - Homo sapiens, 322 aa. [WO200224726-A2, 28-MAR-2002]	17..345 1..322	321/329 (97%) 321/329 (97%)	e-180
ABP95758	Human GPCR polypeptide SEQ ID NO 326 - Homo sapiens, 322 aa. [WO200216548-A2, 28-FEB-2002]	17..345 1..322	321/329 (97%) 321/329 (97%)	e-180

In a BLAST search of public sequence databases, the NOV30a protein was found to have homology to the proteins shown in the BLASTP data in Table 30E.

Table 30E. Public BLASTP Results for NOV30a				
Protein Accession Number	**Protein/Organism/Length**	**NOV30a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8WZ92	Olfactory receptor 5P2 (Olfactory receptor-like protein JCG3) - Homo sapiens (Human), 322 aa.	17..345 1..322	321/329 (97%) 321/329 (97%)	e-179
CAD37572	Sequence 173 from Patent WO0224726 - Homo sapiens (Human), 322 aa.	17..345 1..322	320/329 (97%) 320/329 (97%)	e-178
Q8VG09	Olfactory receptor MOR204-8 - Mus musculus (Mouse), 314 aa.	17..330 1..314	253/314 (80%) 279/314 (88%)	e-145
Q8VG04	Olfactory receptor MOR204-13 - Mus musculus (Mouse), 314 aa.	21..330 5..314	246/310 (79%) 272/310 (87%)	e-141
Q8VFD3	Olfactory receptor MOR204-16 - Mus musculus (Mouse), 321 aa.	17..336 1..320	239/320 (74%) 271/320 (84%)	e-137

PFam analysis indicates that the NOV30a protein contains the domains shown in the Table 30F.

Table 30F. Domain Analysis of NOV30a			
Pfam Domain	**NOV30a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
DUF40	58..299	43/277 (16%) 152/277 (55%)	0.22
7tm_1	60..309	54/276 (20%) 185/276 (67%)	9.4e-32

Example 31.

The NOV31 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 31A.

Table 31A. NOV31 Sequence Analysis		
NOV31a, CG53803-02 DNA Sequence	SEQ ID NO: 377	1049 bp
	ORF Start: ATG at 54	ORF Stop: TGA at 996
ATATTTTGCTTTGGCAGGAACAATTCTCTTCAACCCTTCCATTAAAAGGAATTATGATGATGGTTTTA		

AGGAATCTGAGCATGGAGCCCACCTTTGCCCTTTTAGGTTTCACAGATTACCCAAAGCTTCAGATTCC
TCTCTTCCTTGTGTTTCTGCTCATGTATGTTATCACAGTGGTAGGAAACCTTGGGATGATCATAATAA
TCAAGATTAACCCCAAATTTCACACTCCTATGTACTTTTTCCTTAGTCACCTCTCTTTTGTTGATTTT
TGTTACTCTTCCATTGTCACTCCCAAGCTGCTTGAGAACTTGGTAATGGCAGATAAAAGCATCTTCTA
CTTTAGCTGCATGATGCAGTACTTCCTGTCCTGCACTGCTGTGGTGACAGAGTCTTTCTTGCTGGCAG
TGATGGCCTATGACCGCTTTGTGGCCATCTGCAATCCTCTGCTTTATACAGTGGCCATGTCACAGAGG
CTCTGTGCCCTGCTGGTGGCTGGGTCATATCTCTGGGGCATGTTTGGCCCCTTGGTACTCCTTTGTTA
TGCTCTCCGGTTAAACTTCTCTGGACCTAATGTAATCAACCACTTCTTTTGTGAGTATACTGCTCTCA
TCTCTGTGTCTGGCTCTGATATACTCATCCCCCACCTGCTGCTTTTCAGCTTCGCCACCTTCAATGAG
ATGTGTACACTACTGATCATCCTCACTTCCTATGTTTTCATTTTTGTGACTGTACTAAAAATCCGTTC
TGTTAGTGGGCGCCACAAAGCCTTCTCCACCTGGGCCTCCCACCTGACTGCTATCACCATCTTCCATG
GGACCATCCTTTTCCTTTACTGTGTACCCAACTCCAAAAACTCTCGGCAAACAGTCAAAGTGGCCTCT
GTATTTTACACAGTTGTCAACCCCATGCTGAACCCTCCGATCTACAGCCTAAGGAATAAAGACGTGAA
GGATGCTTTCTGGAAGTTAATACATACACAAGTTCCATTTCACTGAACCAGTCTCAAAAGTTGTTTTC
AATCCAAATGAACAACCCANNNNNNNNNN

NOV31a, CG53803-02 Protein Sequence	SEQ ID NO: 378	314 aa	MW at 35790.4kD

MMMVLRNLSMEPTFALLGFTDYPKLQIPLFLVFLLMYVITVVGNLGMIIIIKINPKFHTPMYFFLSHL
SFVDFCYSSIVTPKLLENLVMADKSIFYFSCMMQYFLSCTAVVTESFLLAVMAYDRFVAICNPLLYTV
AMSQRLCALLVAGSYLWGMFGPLVLLCYALRLNFSGPNVINHFFCEYTALISVSGSDILIPHLLLFSF
ATFNEMCTLLIILTSYVFIFVTVLKIRSVSGRHKAFSTWASHLTAITIFHGTILFLYCVPNSKNSRQT
VKVASVFYTVVNPMLNPPIYSLRNKDVKDAFWKLIHTQVPFH

NOV31b, CG53803-01 DNA Sequence	SEQ ID NO: 379	1039 bp
	ORF Start: ATG at 54	ORF Stop: TGA at 996

ATATTTTGCTTTGGCAGGAACAATTCTCTTCAACCCTTCCATTAAAAGGAATTATGATGATGGTTTTA
AGGAATCTGAGCATGGAGCCCACCTTTGCCCTTTTAGGTTTCACAGATTACCCAAAGCTTCAGATTCC
TCTCTTCCTTGTGTTTCTGCTCATGTATGTTATCACAGTGGTAGGAAACCTTGGGATGATCATAATAA
TCAAGATTAACCCCAAATTTCACACTCCTATGTACTTTTTCCTTAGTCACCTCTCTTTTGTTGATTTT
TGTTACTCTTCCATTGTCACTCCCAAGCTGCTTGAGAACTTGGTAATGGCAGATAAAAGCATCTTCTA
CTTTAGCTGCATGATGCAGTACTTCCTGTCCTGCACTGCTGTGGTGACAGAGTCTTTCTTGCTGGCAG
TGATGGCCTATGACCGCTTTGTGGCCATCTGCAATCCTCTGCTTTATACAGTGGCCATGTCACAGAGG
CTCTGTGCCCTGCTGGTGGCTGGGTCATATCTCTGGGGCATGTTTGGCCCCTTGGTACTCCTTTGTTA
TGCTCTCCGGTTAAACTTCTCTGGACCTAATGTAATCAACCACTTCTTTTGTGAGTATACTGCTCTCA
TCTCTGTGTCTGGCTCTGATATACTCATCCCCCACCTGCTGCTTTTCAGCTTCGCCACCTTCAATGAG
ATGTGTACACTACTGATCATCCTCACTTCCTATGTTTTCATTTTTGTGACTGTACTAAAAATCCGTTC
TGTTAGTGGGCGCCACAAAGCCTTCTCCACCTGGGCCTCCCACCTGACTGCTATCACCATCTTCCATG
GGACCATCCTTTTCCTTTACTGTGTACCCAACTCCAAAAACTCTCGGCAAACAGTCAAAGTGGCCTCT
GTATTTTACACAGTTGTCAACCCCATGCTGAACCCTCCGATCTACAGCCTAAGGAATAAAGACGTGAA
GGATGCTTTCTGGAAGTTAATACATACACAAGTTCCATTTCACTGAACCAGTCTCAAAAGTTGTTTTC
AATCCAAATGAACAACCCA

NOV31b, CG53803-01 Protein Sequence	SEQ ID NO: 380	314 aa	MW at 35790.4kD

MMMVLRNLSMEPTFALLGFTDYPKLQIPLFLVFLLMYVITVVGNLGMIIIIKINPKFHTPMYFFLSHL
SFVDFCYSSIVTPKLLENLVMADKSIFYFSCMMQYFLSCTAVVTESFLLAVMAYDRFVAICNPLLYTV
AMSQRLCALLVAGSYLWGMFGPLVLLCYALRLNFSGPNVINHFFCEYTALISVSGSDILIPHLLLFSF
ATFNEMCTLLIILTSYVFIFVTVLKIRSVSGRHKAFSTWASHLTAITIFHGTILFLYCVPNSKNSRQT
VKVASVFYTVVNPMLNPPIYSLRNKDVKDAFWKLIHTQVPFH

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 31B.

Table 31B. Comparison of the NOV31 protein sequences.

NOV31a	MMMVLRNLSMEPTFALLGFTDYPKLQIPLFLVFLLMYVITVVGNLGMIIIIKINPKFHTP
NOV31b	MMMVLRNLSMEPTFALLGFTDYPKLQIPLFLVFLLMYVITVVGNLGMIIIIKINPKFHTP

```
NOV31a  MYFFLSHLSFVDFCYSSIVTPKLLENLVMADKSIFYFSCMMQYFLSCTAVVTESFLLAVM
NOV31b  MYFFLSHLSFVDFCYSSIVTPKLLENLVMADKSIFYFSCMMQYFLSCTAVVTESFLLAVM

NOV31a  AYDRFVAICNPLLYTVAMSQRLCALLVAGSYLWGMFGPLVLLCYALRLNFSGPNVINHFF
NOV31b  AYDRFVAICNPLLYTVAMSQRLCALLVAGSYLWGMFGPLVLLCYALRLNFSGPNVINHFF

NOV31a  CEYTALISVSGSDILIPHLLLFSFATFNEMCTLLIILTSYVFIFVTVLKIRSVSGRHKAF
NOV31b  CEYTALISVSGSDILIPHLLLFSFATFNEMCTLLIILTSYVFIFVTVLKIRSVSGRHKAF

NOV31a  STWASHLTAITIFHGTILFLYCVPNSKNSRQTVKVASVFYTVVNPMLNPPIYSLRNKDVK
NOV31b  STWASHLTAITIFHGTILFLYCVPNSKNSRQTVKVASVFYTVVNPMLNPPIYSLRNKDVK

NOV31a  DAFWKLIHTQVPFH
NOV31b  DAFWKLIHTQVPFH

NOV31a  (SEQ ID NO:  378)
NOV31b  (SEQ ID NO:  380)
```

Further analysis of the NOV31a protein yielded the following properties shown in Table 31C.

Table 31C. Protein Sequence Properties NOV31a

SignalP analysis:	Cleavage site between residues 44 and 45

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 1;  neg.chg 1
      H-region:  length 9;  peak value   9.41
      PSG score:   5.01

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -1.06
      possible cleavage site: between 43 and 44

>>> Seems to have a cleavable signal peptide (1 to 43)

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 44
      Tentative number of TMS(s) for the threshold  0.5:   5
      INTEGRAL    Likelihood = -2.23   Transmembrane  104 - 120
      INTEGRAL    Likelihood = -0.64   Transmembrane  145 - 161
      INTEGRAL    Likelihood = -0.80   Transmembrane  186 - 202
      INTEGRAL    Likelihood = -8.39   Transmembrane  212 - 228
      INTEGRAL    Likelihood = -2.71   Transmembrane  247 - 263
      PERIPHERAL  Likelihood =  2.07 (at 63)
      ALOM score:  -8.39  (number of TMSs: 5)

MTOP: Prediction of membrane topology (Hartmann et al.)
```

```
      Center position for calculation: 21
      Charge difference:  3.0   C( 1.0) - N(-2.0)
      C > N: C-terminal side will be inside

>>>Caution: Inconsistent mtop result with signal peptide
>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  7.94
      Hyd Moment(95):  8.95    G content:       1
      D/E content:     2       S/T content:     3
      Score: -5.12

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 16  LRN|LS

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   6.4%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:
      Leucine zipper pattern (PS00029):  *** found ***
         LLVAGSYLWGMFGPLVLLCYAL at 145
 none
```

```
checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       44.4 %: endoplasmic reticulum
       22.2 %: vacuolar
       22.2 %: mitochondrial
       11.1 %: Golgi

>> prediction for CG53803-02 is end (k=9)
```

A search of the NOV31a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 31D.

Table 31D. Geneseq Results for NOV31a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV31a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU85141	G-coupled olfactory receptor #2 - Homo sapiens, 314 aa. [WO200198526-A2, 27-DEC-2001]	1..314 1..314	314/314 (100%) 314/314 (100%)	0.0
AAG72434	Human OR-like polypeptide query sequence, SEQ ID NO: 2115 - Homo sapiens, 314 aa. [WO200127158-A2, 19-APR-2001]	1..314 1..314	314/314 (100%) 314/314 (100%)	0.0
AAG71725	Human olfactory receptor polypeptide, SEQ ID NO: 1406 - Homo sapiens, 314 aa. [WO200127158-A2, 19-APR-2001]	1..314 1..314	314/314 (100%) 314/314 (100%)	0.0
AAE07540	Human G-protein coupled receptor 2a (GPCR2a) variant - Homo sapiens, 314 aa. [WO200159113-A2, 16-AUG-2001]	1..314 1..314	314/314 (100%) 314/314 (100%)	0.0
AAU24517	Human olfactory receptor AOLFR2 - Homo sapiens, 314 aa. [WO200168805-A2, 20-SEP-2001]	1..314 1..314	314/314 (100%) 314/314 (100%)	0.0

In a BLAST search of public sequence databases, the NOV31a protein was found to have homology to the proteins shown in the BLASTP data in Table 31E.

Table 31E. Public BLASTP Results for NOV31a				
Protein Accession Number	**Protein/Organism/Length**	**NOV31a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
CAC69325	Sequence 3 from Patent WO0159113 - Homo sapiens (Human), 314 aa.	1..314 1..314	314/314 (100%) 314/314 (100%)	0.0
CAC69326	Sequence 5 from Patent WO0159113 - Homo sapiens (Human), 314 aa.	1..314 1..314	313/314 (99%) 313/314 (99%)	0.0
Q8NGL3	Seven transmembrane helix receptor - Homo sapiens (Human), 314 aa.	1..314 1..314	312/314 (99%) 313/314 (99%)	0.0
CAD37508	Sequence 39 from Patent WO0224726 - Homo sapiens (Human), 305 aa.	10..314 1..305	305/305 (100%) 305/305 (100%)	e-177
Q8VFR4	Olfactory receptor MOR174-8 - Mus musculus (Mouse), 316 aa.	7..313 7..313	243/307 (79%) 267/307 (86%)	e-139

PFam analysis indicates that the NOV31a protein contains the domains shown in the Table 31F.

Table 31F. Domain Analysis of NOV31a			
Pfam Domain	**NOV31a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	43..292	58/278 (21%) 174/278 (63%)	1.4e-24

Example 32.

The NOV32 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 32A.

Table 32A. NOV32 Sequence Analysis

NOV32a, CG53989-03 DNA Sequence	SEQ ID NO: 381	1092 bp
	ORF Start: ATG at 30	ORF Stop: TAG at 1047

CCGGGACAGTCACCTCCAGCAGCATCCTG**ATG**TTCTGGGGACACTGGTGCTGGGGGGTCAGTGGGGAA
GGGCTCCTGGGTCCTCATGACCCTCTCCCTTGGGTGAGCACAAAACACCATGGCACTTTGGGGTCTTG
GAAACAGACTAAAGGGGATGTCAAGTCCTCATTCCTTGGAGCTGCTGACGAGTCCAGAATGGGTCATG
TTTTCTTGCCCCGACCCCAGCACCTCAGGGCAGCGGAAGGTCCAGAGAGAGGTCGGGGACCGGGGCCG
CTCCTTGCATCCTGGGCTTGTGTCTGTTGCCCCCTGGCTGGTGACTTGCACTCTCCTGGAGCTGGTTC
TTGCAGCCGAGGCCGTCACGGGGCTGGGATGTCGCTGCTGCTTCTCTTCGTGGTGTTGACCATTTCTC
AGACCTCCCCCCGCCCCTGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTG
AGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAG
CCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGC
TCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGTGCCCTCGGGGAAGACCTGCTGGGTGACC
GGCTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGGTGAA
GGTCAGGAGCAACGTCCTCTGTAACCAGACCTGTCGCCGCCGCTTTCCTTCCAACCACACTGAGCGGT
TTGAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGAC
AACGGGGGCCCCCTCCTGTGCGGGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAA
ATTCTGCGGCCTTCGCGGCTATCCCGGCATGTACACTCGCGTGACGAGCTACGTGTCCTGGATCCGCC
AGTACGTCCCGCCGTTCCCCAGACGC**TAG**CTGGGGTGCAGTGGGGTCTGCATGATCCAGGAGGGCCCG
TCTT

NOV32a, CG53989-03 Protein Sequence	SEQ ID NO: 382	339 aa	MW at 37349.5kD

MFWGHWCWGVSGEGLLGPHDPLPWVSTKHHGTLGSWKQTKGDVKSSFLGAADESRMGHVFLPRPQHLR
AAEGPERGRGPGPLLASWACVCCPLAGDLHSPGAGSCSRGRHGAGMSLLLLFVVLTISQTSPRPCREE
LEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSA
SLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSNVLCNQTCRRRFPSNHTERFERLIKDDML
CAGDGNHGSWPGDNGGPLLCGRNCTWVQVEVVSWGKFCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR

NOV32b, CG53989-04 DNA Sequence	SEQ ID NO: 383	881 bp
	ORF Start: ATG at 31	ORF Stop: TAG at 832

GCTCTGGGAAGACCCTCGTCCGTCCCCCTC**ATG**AGCCCGGCACGGGGCGTGAGCTGGTGGGCATCACT
GGGGGCTGCGACGTCCCGGCCAGGAGGCACCCCTGGTAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAG
TGCAGGTGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCAC
CCCCAGTACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCC
GGTGCCGCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCGGGACGTGCCCTCGGGGA
AGACCTGCTGGGCCTGGGCTCCAGACGCGTCCTGGATGGCTTCCTGCCGCTGCCGAGACGGATGGCCA
CAGGCCAGGTGGCTCAGAGCAGCAGGAATGTACTATCTCACGGCTCTGCAGGCGGAACGTCCACACTC
CAGGCGTGGGCAGGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCA
ACGTCCTCTGTAACCAGACCTGTCGCCGCCGCTTTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTC
ATCAAGGACGACATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCC
CCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAATTCTGCGGCC
TTCGCGGCTATCCCGGCATGTACACTCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCG
CCGTTCCCCAGACGC**TAG**CTGGGGTGCAGTGGGGTCTGCATGATCCAGGAGGGCCCGTCTAAGCG

NOV32b, CG53989-04 Protein Sequence	SEQ ID NO: 384	267 aa	MW at 30126.0kD

MSPARGVSWWASLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQ
GGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWAWAPDASWMASCRCRDGWPQARWLRAAGM
YYLTALQAERPHSRRGQELLPWPLSLWEATVKVRSNVLCNQTCRRRFPSNHTERFERLIKDDMLCAGD
GNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKFCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR

NOV32c, 306076095 DNA Sequence	SEQ ID NO: 385	819 bp
	ORF Start: at 1	ORF Stop: end of sequence

CGCAAGCTTATGAGCCCGGCACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCGGCC
AGGAGGCACCCCTGGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGT
TCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGCCTG

TCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCTCAT
CCACCCGGTCTCGCTCCCGTCTGCCTCCCGGGACGTGCCCTCGGGGAAGACCTGCTGGGCCTGGGCTC
CAGACGCGTCCTGGATGGCTTCCTGCCGCTGCCGAGACGGATGGCCACAGGCCAGGTGGCTCAGAGCA
GCAGGAATGTACTATCTCACGGCTCTGCAGGCGGAACGTCCACACTCCAGGCGTGGGCAGGAACTACT
GCCCTGGCCCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGTAACCAGACCT
GTCGCCGCCGCTTTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGACATGCTGTGT
GCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCAGGCGGAATTG
CACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTATCCCGGCATGT
ACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAGACGCGTCGAC
GCG

NOV32c, 306076095 Protein Sequence	SEQ ID NO: 386	273 aa	MW at 30808.8kD

RKLMSPARGVSWWASLGAATSRPGGTPGREELEACAFRVQVGQLRFYEDDQRTKVVEIVRHPQYNESL
SAQGGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWAWAPDASWMASCRCRDGWPQARWLRA
AGMYYLTALQAERPHSRRGQELLPWPLSLWEATVKVRSNVLCNQTCRRRFPSNHTERFERLIKDDMLC
AGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRRVD
A

NOV32d, CG53989-01 DNA Sequence	SEQ ID NO: 387	858 bp
	ORF Start: ATG at 1	ORF Stop: TAG at 856

ATGCTGTGGCTACTGCTCCTGACCCTCCCCTGCCTGATGGGCTCTGTGCCCAGGAACCCAGGCGAGGG
CACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGG
TCAGCCTGAGGTTCTACAGCATGAAGAAGGGTCTGTGGGAGCCCATCTGTGGGGGCTCCCTCATCCAC
CCAGAGTGGGTGCTGACCGCCGCCCACTGCCTTTTGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCA
GGTGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCC
AGTACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTG
CCGCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGTGCCCTCGGGGAAGAC
CTGCTGGGTGACCGGCTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGCCCCTCAGCTTGTGGG
AGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGTAACCAGACCTGTCGCCGCCGCTTTCCTTCCAAC
CACACTGAGCGGTTTGAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGACGGGAACCACGGCTC
CTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGGTGGAGGTGG
TGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTATCCCGGCATGTACACCCGCGTGACGAGCTACGTG
TCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAGACGC**TAG**

NOV32d, CG53989-01 Protein Sequence	SEQ ID NO: 388	285 aa	MW at 31927.6kD

MLWLLLLTLPCLMGSVPRNPGEGTGRELVGITGGCDVSARRHPWQVSLRFYSMKKGLWEPICGGSLIH
PEWVLTAAHCLLEELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLKLEAPV
PLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSNVLCNQTCRRRFPSN
HTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTRVTSYV
SWIRQYVPPFPRR

NOV32e, CG53989-02 DNA Sequence	SEQ ID NO: 389	660 bp
	ORF Start: at 1	ORF Stop: end of sequence

TCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGGGAGGAGTTGGAGGCTTGCGCGTT
TAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCC
GTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAG
GCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCGGGACGTGCCCTC
GGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGCCCCTCA
GCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGTAACCAGACCTGTCGCCGCCGCTTT
CCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGACGGGAA
CCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGG
TGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTATCCCGGCATGTACACCCGCGTGACG
AGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAGACGC

5

NOV32e, CG53989-02 Protein Sequence	SEQ ID NO: 390	220 aa	MW at 24527.8kD

SLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLKLE
APVPLSELIHPVSLPSASRDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSNVLCNQTCRRRF
PSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTRVT
SYVSWIRQYVPPFPRR

NOV32f, CG53989-05 DNA Sequence	SEQ ID NO: 391	2847 bp
	ORF Start: ATG at 1	ORF Stop: TAG at 2845

ATGGTCAGCAAGGGGGGAGTTGCTGCAGAGCCAGAGCCACACTATTGTGAGGACAGTGAAAGAGGCCC
CAACACCCTCACAGGTCCGGGCAGCCTTCCTAGAGGAGGTGGCATTGAGGTGGGCATGGAGTTTCCGG
GATGCAGCGGTGAAGGGTGCGTGAAGCCCCATGAGGAGGCGGCCCGGGAGGGGGCGGGCAGAGGCAAG
AGGGCTGTGCCGGGACCCAAGCGACGGCAGCAGGGGTCAGCAGAGGGGCCTGCGGCGGGGTGGACGCT
GGAGCAGGAGACCAGGGGAGATGTCTTAGAGGATAAAAATGAGCGGGCAGATGAAGAGATACTCAGGC
TGGCACCAGGGAAAGGCAGGCTCCCAATAGACAGCAAACACCTGAAACCGGTGATCAGCAGCTTCCCG
GTAAGATCTCAGGAGCTGGGCGAGGGGGCTGGAGCAGGCACACTAAGAGGCAAAATGGCAGAGTTTAA
CTGGTCTATGGCCTTCAAGGGACCTGCGGCTGGTCATGAAGAGCGCCTCAACTCTGTGTCCAGCAGGG
CCAAGAAGGGCATTGGCTGGGATGTCGCTGCTGCTTCTCTTCGTGGTGTTGACCATTTCTCAGACCTC
CCCCCGCCCCTGCAGGTCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGAG
GCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGCC
TGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCTC
ATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGTGCCCTCGGGGAAGACCTGCTGGGTGACCGG
CTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGGTGAAGG
TCAGGAGCAACGTCCTCTGTAACCAGACCTGTCGCCGCCGCTTTCCTTCCAACCACACTGAGCGGTTT
GAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAA
CGGGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAAC
TCTGCGGCCTTCGCGGCTATCCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAG
CCATGCCCCTCAGCTCAGACCCCTGCTGTGGTCCGAAGATTTGTGCTCCCCCCAAATCCAGATGTTGA
AGCCCTAACTCCCAGTGTGATGGGATCAGGAGCGCCGCTGCCCCCGGCCCCCGACCTGCAAGAGGCCG
AGGTCCCCATCATGAGGACCCGAGCTTGCGAGAGGATGTATCACAAAGGCCCCACTGCCCACGGCCAG
GTCACCATCATCAAGGCTGCCATGCCGTGTGCAGGGAGGAAGGGGCAGGGTTCCTGCCAGGCCGCTCT
GAGGACGGAGGACCTCACCCCAACCACACCCAACACGGAGGTGTCTCCACGTGCAGACCCCAGGCTGA
GCCAGCCGGAGGACATCTGGCCAGAGTGGGCTTGGCCAGTTGTGGTGGGCACCACCATGCTGCTGCTG
CTGCTGTTCCTGGCTGTCTCCTCCCTGGGGAGCTGTAGCACTGGGAGTCCAGCTCCCGTCCCCGAGAA
TGACCTGGTGGGCATTGTGGGGGGCCACAACACCCCAGGGGAAGTGGTCGTGGCAGTGGGTGCTGACC
GCCGCTCACTGCATTTTCCGGAAGGACACCGACCCGTCCACCTACCGGATTCACACCAGGGATGTGTA
TCTGTACGGGGGCCGGGGGCTGCTGAATGTCAGCCAGATCGTCGTCCACCCAACTACTCTGTCTTCTT
CCTGGGGGCAGACATCGCCCTGCTGAAGCTGGCCACCAGTTCCCTGGAGTTCACTGACAGTGACAACT
GCTGGAACACAGGCTGGGGCATGGTCGGCTTGTTGGATATGCTGCCGCCTCCTTACCGCCCGCAGCAG
GTGAAGGTCCTCACACTGAGCAATGCAGACTGTGAGCGGCAGACCTACGATGCTTTTCCTGGTGCTGG
AGACAGAAAGTTCATCCAGGATGACATGATCTGTGCCGGCCGCACGGGCCGCCGCACCTGGAAGGGTG
ACTCAGGCGGCCCCCTGGTCTGCAAGAAGAAGGGTACCTGGCTCCAGGCGGGAGTAGTGAGCTGGGGA
TTTTACAGTGATCGGCCCAGCATTGGCGTCTACACACGCCCAGAGACCAGCTGGCAGGGTGCCAACCA
TGCAGACGCCCAGAGACCAGCTGGCAGGGTGCCAACCATGCAGAGGCCCAGAGACATGGGCCAGGGCC
AGGAGTGGGTCTGCAGGCCCTTCACCCACGTCACCTGCTACCCGACGGCCATCCCCAGGCCCTTCACC
CATGTCACCTGCTACCTGATGGCTGTCCCCAGCACCCTCACCCACGTCACCTGCTACCCGACGGCCGT
CCCCAGGCCCTTCACCCATGTCACCTGCTACCTGATGGCTGTCCCCAGCACCCTCACCCACATCACCT
GCTACATGATGGCCGTCCCCAGGCCCTTTACCCACATCACCTGCTACCCAATGGCTGTCCCCAGCACC
CTTACCCACGTCACCTGCCACCCGACGGCCATCCCCAGGCCCTTCACCCACATCACCTGCTACACGAT
GGCCATCCCCAGGCCTTCAACCACGCCACCTGCTACACGACGGCCATCCCCAGCACCCTCACCCACGT
CACCTGCTACACGATGGCCGTCCCCAGGCCCATCACCCATGTCACCTGCTACACGA**TAG**

NOV32f, CG53989-05 Protein Sequence	SEQ ID NO: 392	948 aa	MW at 102839.3kD

MVSKGGVAAEPEPHYCEDSERGPNTLTGPGSLPRGGGIEVGMEFPGCSGEGCVKPHEEAAREGAGRGK
RAVPGPKRRQQGSAEGPAAGWTLEQETRGDVLEDKNERADEEILRLAPGKGRLPIDSKHLKPVISSFP

VRSQELGEGAGAGTLRGKMAEFNWSMAFKGPAAGHEERLNSVSSRAKKGIGWDVAAASLRGVDHFSDL
PPPLQVREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLKLEAPVPLSEL
IHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSNVLCNQTCRRRFPSNHTERF
ERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTRVTSYVSWIRQ
PCPSAQTPAVVRRFVLPPNPDVEALTPSVMGSGAPLPPAPDLQEAEVPIMRTRACERMYHKGPTAHGQ
VTIIKAAMPCAGRKGQGSCQAALRTEDLTPTTPNTEVSPRADPRLSQPEDIWPEWAWPVVVGTTMLLL
LLFLAVSSLGSCSTGSPAPVPENDLVGIVGGHNTPGEVVVAVGADRRSLHFPEGHRPVHLPDSHQGCV
SVRGPGAAECQPDRRPPNYSVFFLGADIALLKLATSSLEFTDSDNCWNTGWGMVGLLDMLPPPYRPQQ
VKVLTLSNADCERQTYDAFPGAGDRKFIQDDMICAGRTGRRTWKGDSGGPLVCKKKGTWLQAGVVSWG
FYSDRPSIGVYTRPETSWQGANHADAQRPAGRVPTMQRPRDMGQGQEWVCRPFTHVTCYPTAIPRPFT
HVTCYLMAVPSTLTHVTCYPTAVPRPFTHVTCYLMAVPSTLTHITCYMMAVPRPFTHITCYPMAVPST
LTHVTCHPTAIPRPFTHITCYTMAIPRPSTTPPATRRPSPAPSPTSPATRWPSPGPSPMSPATR

NOV32g, CG53989-06 DNA Sequence	SEQ ID NO: 393	660 bp
	ORF Start: at 1	ORF Stop: end of sequence

TCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGGGAGGAGTTGGAGGCTTGCGCGTT
TAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCC
GTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAG
GCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCGGGACGTGCCCTC
GGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGCCCCTCA
GCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGTAACCAGACCTGTCGCCGCCGCTTT
CCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGACGGGAA
CCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGG
TGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTATCCCGGCATGTACACCCGCGTGACG
AGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAGACGC

NOV32g, CG53989-06 Protein Sequence	SEQ ID NO: 394	220 aa	MW at 24527.8kD

SLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLKLE
APVPLSELIHPVSLPSASRDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSNVLCNQTCRRRF
PSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTRVT
SYVSWIRQYVPPFPRR

NOV32h, CG53989-07 DNA Sequence	SEQ ID NO: 395	672 bp
	ORF Start: at 7	ORF Stop: at 667

AGATCTTCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGGGAGGAGTTGGAGGCTTG
CGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGA
TCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAG
CTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGT
GCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGC
CCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGTAACCAGACCTGTCGCCGC
CGCTTTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGA
CGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGG
TCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTATCCCGGCATGTACACCCGC
GTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAGACGCCTCGAG

NOV32h, CG53989-07 Protein Sequence	SEQ ID NO: 396	220 aa	MW at 24484.8kD

SLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLKLE
APVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSNVLCNQTCRRRF
PSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTRVT
SYVSWIRQYVPPFPRR

NOV32i, CG53989-08 DNA Sequence	SEQ ID NO: 397	1171 bp
	ORF Start: at 187	ORF Stop: TGA at 805

ATGCTGTGGCTACTGCTCCTGACCCTCCCCTGCCTGATGGGCTCTGTGCCCAGGAACCCAGGCGAGTC
CGCCCCACCCAATGCCCCTGCTGCCCAGGACCCCCTCCTTGCCCTGCCCCGGGCTCAGAGTGCCAGCC
CTGGGGTGGGTGGGGACCATCTGATTGCCGGTCTCTCCTGGTGCCCCTGAGCTCTGGGAAGACCCTCG

TCCGTCCCCCTCATGAGCCCGGCACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCG
GCCAGGAGGCACCCCTGGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGA
GGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGC
CTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCT
CATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGTGCCCTCGGGGAAGACCTGCTGGGTGACCG
GCTGGGGTGTCATTGGACGTGGAGGTCAGGAGCAGGACCACTTGGGTGGGATGTGGAGAGATGACCCG
GAATGTCGGTGCAGGCCTGGGCTCCAGACGCGTCCTGGATGGCTTCCTGCCGCTGCCGAGACGGATGG
CCACAGGCCAGGTGGCTCAGAGCAGCAGGAATGTACTATCTCACGGCTCTGCAGGCGGAACGTCCACA
CTCCAGGCGTGGGCAGGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGA
GCAACGTCCTCTGTAACCAGACCTGTCGCCGCCGCTTTCCTTCCAACCACACTGAGCGGTTTGAGCGG
CTCATCAAGGACGACATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGG
CCCCCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCG
GCCTTCGCGGCTATCCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTC
CCGCCGTTCCCCAGACGCTAGCTGGGGTGCAGTGGGGTCTGCATGATCCAGGAGGGCCCGTCTTCCTT
GTGGACACGCCTGCT

NOV32i, CG53989-08 Protein Sequence	SEQ ID NO: 398	206 aa	MW at 22191.8kD

ALGRPSSVPLMSPARGVSWWASLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRH
PQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGGQEQDHLGG
MWRDDPECRCRPGLQTRPGWLPAAAETDGHRPGGSEQQECTISRLCRRNVHTPGVGRNYCPGPSACGR
RR

NOV32j, CG53989-09 DNA Sequence	SEQ ID NO: 399	843 bp
	ORF Start: ATG at 26	ORF Stop: TAG at 836

GAGGTGGAGGTTGCAGTAAGCCAAGATGGCGCCACTGCACTCTAGCCTGTTTCTGCTGAGCGGGACCA
TGAGCCCAAAAGTGGGCATCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGGTCAGC
CTGAGGTTCTACAGCATGAAGAAGGGTCTGTGGGAGCCCATCTGTGGGGGCTCCCTCATCCACCCAGA
GTGGGTGCTGACCGCCGCCCACTGCGTCGAGCTTGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGG
TGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAG
TACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCC
GCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGTGCCCTCGGGGAAGACCT
GCTGGGTGACCGGCTGGGGTGTCATTTGGGGACACGTTTTCCTGCTCCCGCCACCCCACCTCAGGGCA
GCGGAAGGTCCAATCATGAGGACCCGAGCTTGCGAGAGGATGTATCACAAAGGCCCCACTGCCCACGT
CACCATCATCAAGGCTGCCATGCCGTGTGCAGGGGCTGAGCGCCATCTCTCCCCACAGGGCGACAACG
GGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTC
TGCGGCCTTCGCTATCCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGT
CCCGCCGTTCCCCAGACGCTAGCTGGG

NOV32j, CG53989-09 Protein Sequence	SEQ ID NO: 400	270 aa	MW at 29993.6kD

MAPLHSSLFLLSGTMSPKVGITGGCDVSARRHPWQVSLRFYSMKKGLWEPICGGSLIHPEWVLTAAHC
VELEELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPV
SLPSASLDVPSGKTCWVTGWGVIWGHVFLLPPPHLRAAEGPIMRTRACERMYHKGPTAHVTIIKAAMP
CAGAERHLSPQGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRYPGMYTRVTSYVSWIRQYVPPFPRR

NOV32k, CG53989-10 DNA Sequence	SEQ ID NO: 401	964 bp
	ORF Start: ATG at 67	ORF Stop: TGA at 655

GACCATCTGATTGCCGGTCTCTCCTGGTGCCCCTGAGCTCTGGGAAGACCCTCGTCCGTCCCCCTCAT
GAGCCCGGCACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCC
CTGGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGAC
GACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGG
CGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCT
CGCTCCCGTCTGCCTCCCGGGACGTGCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATT
GGACGTGGAGGTCAGGAGCAGGACCACTTGGGTGGGATGTGGAGAGATGACCCGGAATGTCGGTGCAG
GCCTGGGCTCCAGACGCGTCCTGGATGGCTTCCTGCCGCTGCCGAGACGGATGGCCACAGGCCAGGTG
GCTCAGAGCAGCAGGAATGTACTATCTCACGGCTCTGCAGGCGGAACGTCCACACTCCAGGCGTGGGC
AGGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGT

AACCAGACCTGTCGCCGCCGCATTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGA
CATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCA
GGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTAT
CCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAG
ACGCTAGCTGGG

NOV32k, CG53989-10 Protein Sequence	SEQ ID NO: 402	196 aa	MW at 21256.7kD

MSPARGVSWWASLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQ
GGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWVTGWGVIGRGGQEQDHLGGMWRDDPECRC
RPGLQTRPGWLPAAAETDGHRPGGSEQQECTISRLCRRNVHTPGVGRNYCPGPSACGRRR

NOV32l, CG53989-11 DNA Sequence	SEQ ID NO: 403	948 bp
	ORF Start: ATG at 61	ORF Stop: TAG at 931

TGACCCTCCCCTGCCTGATGGGCTCTGTGCCCAGGAACCCAGGCGAGTCCGCCCCACCCA**ATG**CCCCT
GCTGCCCAGCCGGTCTCTCCTGGTGCCCCTGAGCTCTGGGAAGACCCTCGTCCGTCCCCCTCATGAGC
CCGGCACGGGGCGTGAGCTGGTGGGCATCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGG
CAGGTCAGCCTGAGGTTCTACAGCATGAAGAAGGGTCTGTGGGAGCCCATCTGTGGGGGCTCCCTCAT
CCACCCAGAGTGGGTGCTGACCGCCGCCCACTGCCTTGGGCCTGAGGAGTTGGAGGCTTGCGCGTTTA
GAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCCGT
CACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAGGC
CCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGTGCCCTCGG
GGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGCCCCTCAGC
TTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGTAACCAGACCTGTCGCCGCCGCTTTCC
TTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGACGAGCGCC
ATCTCTCCCCACAGGGCGACAACGGGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGGTG
GAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTATCCCGGCATGTACACCCGCGTGACGAG
CTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAGACGC**TAG**CTGGGGTGCAGTGGG

NOV32l, CG53989-11 Protein Sequence	SEQ ID NO: 404	290 aa	MW at 32449.2kD

MPLLPSRSLLVPLSSGKTLVRPPHEPGTGRELVGITGGCDVSARRHPWQVSLRFYSMKKGLWEPICGG
SLIHPEWVLTAAHCLGPEELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLK
LEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSNVLCNQTCRR
RFPSNHTERFERLIKDDMLCAGDERHLSPQGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTR
VTSYVSWIRQYVPPFPRR

NOV32m, CG53989-12 DNA Sequence	SEQ ID NO: 405	843 bp
	ORF Start: ATG at 11	ORF Stop: TAG at 836

TGAGAGATAA**ATG**GGCTCCCAGAGATGCCAGGGAGGAGGCCCCGGCACGGGGCGTGAGCTGGTGGGCA
TCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGGTCAGCCTGAGGTTCTACAGCATG
AAGAAGGGTCTGTGGGAGCCCATCTGTGGGGGCTCCCTCATCCACCCAGAGTGGGTGCTGACCGCCGC
CCACTGCCTTGGCAGGGAGGAGTTGGAGGCTTGCGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCT
ATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCT
GCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCA
CCCGGTCTCGCTCCCGTCTGCCTCCCGGCCTGGGCTCCAGACGCGTCCTGGATGGCTTCCTGCCGCTG
CCGAGACGGATGGGCAGGAACTACTGCCCTGGCCCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGG
AGCAACGTCCTCTGTAACCAGACCTGTCGCCGCCGCTTTCCTTCCAACCACACTGAGCGGTTTGAGCG
GCTCATCAAGGACGACATGCTGTGTGCCGGGGACGGGAACCACGGCTCCTGGCCAGGCGACAACGGGG
GCCCCCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGC
GGCCTTCGCGGCTATCCCGGCATGTACACCCGCGTGACGAGCTACGTGTCCTGGATCCGCCAGTACGT
CCCGCCGTTCCCCAGACGC**TAG**CTGGG

NOV32m, CG53989-12 Protein Sequence	SEQ ID NO: 406	275 aa	MW at 30683.8kD

MGSQRCQGGGPGTGRELVGITGGCDVSARRHPWQVSLRFYSMKKGLWEPICGGSLIHPEWVLTAAHCL
GREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVS
LPSASRPGLQTRPGWLPAAAETDGQELLPWPLSLWEATVKVRSNVLCNQTCRRRFPSNHTERFERLIK

DDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPF PRR

NOV32n, CG53989-13 DNA Sequence	SEQ ID NO: 407	870 bp
	ORF Start: ATG at 43	ORF Stop: TAA at 868

ATCTGGCCAGAGTGGGCTTGGCCAGTTGTGGTGGGCACCACC**ATG**CTGCTGCTGCTGCTGTTCCTGGC TGTCTCCTCCCTGGGGAGCTGTAGCACTGGGAGTCCAGCTCCCGTCCCCGAGAATGACCTGGTGGGCA TTGTGGGGGGCCACAACACCCAGGGGAAGTGGTCGTGGCAGGTCAGCCTGAGGATCTATAGCTACCAC TGGGCCTCCTGGGTGCCCATCTGCGGGGGCTCCCTCATCCACCCCCAGTGGGTGCTGACCGCCGCTCA CTGCATTTTCCGGAAGGACACCGACCCGTCCACCTACCGGATTCACACCAGGGATGTGTATCTGTACG GGGGCCGGGGGCTGCTGAATGTCAGCCAGATCGTCGTCCACCCCAACTACTCTGTCTTCTTCCTGGGG GCAGACATCGCCCTGCTGAAGCTGGCCACCAGTGTGAGAACAACAAACACTCTCGCGGCAGTCGCCCT GCCGTCATTGTCCCTGGAGTTCACTGACAGTGACAACTGCTGGAACACAGGCTGGGGCATGGTCGGCT TGTTGGATATGCTGCCGCCTCCTTACCGCCCGCAGCAGGTGAAGGTCCTCACACTGAGCAATGCAGAC TGTGAGCGGCAGACCTACGATGCTTTTCCTGGTGCTGGAGACAGAAAGTTCATCCAGGATGACATGAT CTGTGCCGGCCGCACGGGCCGCCGCACCTGGAAGGGTGACTCAGGCGGCCCCCTGGTCTGCAAGAAGA AGGGTACCTGGCTCCAGGCGGGAGTAGTGAGCTGGGGATTTTACAGTGATCGGCCCAGCATTGGCGTC TACACGTGGGTCCAGACCTATGTGCCCTGGATCCTGCAGCAAATGCACCTC**TAA**

NOV32n, CG53989-13 Protein Sequence	SEQ ID NO: 408	275 aa	MW at 30467.7kD

MLLLLLFLAVSSLGSCSTGSPAPVPENDLVGIVGGHNTQGKWSWQVSLRIYSYHWASWVPICGGSLIH PQWVLTAAHCIFRKDTDPSTYRIHTRDVYLYGGRGLLNVSQIVVHPNYSVFFLGADIALLKLATSVRT TNTLAAVALPSLSLEFTDSDNCWNTGWGMVGLLDMLPPPYRPQQVKVLTLSNADCERQTYDAFPGAGD RKFIQDDMICAGRTGRRTWKGDSGGPLVCKKKGTWLQAGVVSWGFYSDRPSIGVYTWVQTYVPWILQQ MHL

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 32B.

Table 32B. Comparison of the NOV32 protein sequences.

```
NOV32a  ------------------------------------------------------------
NOV32b  ------------------------------------------------------------
NOV32c  ------------------------------------------------------------
NOV32d  ------------------------------------------------------------
NOV32e  ------------------------------------------------------------
NOV32f  MVSKGGVAAEPEPHYCEDSERGPNTLTGPGSLPRGGGIEVGMEFPGCSGEGCVKPHEEAA
NOV32g  ------------------------------------------------------------
NOV32h  ------------------------------------------------------------
NOV32i  ------------------------------------------------------------
NOV32j  ------------------------------------------------------------
NOV32k  ------------------------------------------------------------
NOV32l  ------------------------------------------------------------
NOV32m  ------------------------------------------------------------
NOV32n  ------------------------------------------------------------

NOV32a  ----------MFWGHWCWGVSGEGLLGPHDPLPWVSTKHHGTLGSWKQTKGDVKS
NOV32b  ------------------------------------------------------------
NOV32c  ------------------------------------------------------------
NOV32d  ----------------------------------------------------MLWLLLLTLP
NOV32e  ------------------------------------------------------------
NOV32f  REGAGRGKRAVPGPKRRQQGSAEGPAAGWTLEQETRGDVLEDKNERADEEILRLAPGKGR
NOV32g  ------------------------------------------------------------
NOV32h  ------------------------------------------------------------
```

```
NOV32i  ------------------------------------------------------------
NOV32j  ----------------------------------------------------------MAPLH
NOV32k  ------------------------------------------------------------
NOV32l  ---------------------------------------------------------MPLLPSR
NOV32m  ----------------------------------------------------------MGSQR
NOV32n  ----------------------------------------------------------MLLL

NOV32a  SFLGAADESRMGHVFLPRPQHLRAAEGPERGRGPGPLLASWACVCCPLAGDLHSPGAGSC
NOV32b  ------------------------MSPARG------VSWWA-----------SLGAATS
NOV32c  ----------------------RKLMSPARG------VSWWA-----------SLGAATS
NOV32d  CLMGSVPRNPG-------EGTGRELVGITGGCDVSARRHPWQVS---LR--FYSMKKGLW
NOV32e  ------------------------------------------------------SLGAATS
NOV32f  LPIDSKHLKPVISSFPVRSQELGEGAGAGTLRGKMAEFNWSMAFKGPAAGHEERLNSVSS
NOV32g  ----------------------------------------------------SLGAATS
NOV32h  ----------------------------------------------------SLGAATS
NOV32i  -------------ALGRPS-SVPLMSPARG------VSWWA-----------SLGAATS
NOV32j  SSLFLLS-----------GTMSPKVGITGGCDVSARRHPWQVS---LR--FYSMKKGLW
NOV32k  ------------------------MSPARG------VSWWA-----------SLGAATS
NOV32l  SLLVPLSSGKTLVRPPHEPGTGRELVGITGGCDVSARRHPWQVS---LR--FYSMKKGLW
NOV32m  CQGG--G----------PGTGRELVGITGGCDVSARRHPWQVS---LR--FYSMKKGLW
NOV32n  LLFLAVSSLGSCSTGSPAPVPENDLVGIVGGHNTQG-KWSWQVSLR-----IYSYHWASW

NOV32a  SRGR-HGAGMSLLLLFVVLTISQTSPRPC-REELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32b  RPG---G----------------TPG---REELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32c  RPG---G----------------TPG---REELEACAFRVQVGQLRFYEDDQRTKVVEI
NOV32d  EPIC-GG-------SLIHPEWVLTAAHCL-LEELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32e  RPG-------------------GTPG---REELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32f  RAKKGIGWDVAAASLRGVDHFSDLPPPLQVREELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32g  RPG-------------------GTPG---REELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32h  RPG-------------------GTPG---REELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32i  RPG---G----------------TPG---REELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32j  EPIC-GG-------SLIHPEWVLTAAHCVELEELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32k  RPG---G----------------TPG---REELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32l  EPIC-GG-------SLIHPEWVLTAAHCLGPEELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32m  EPIC-GG-------SLIHPEWVLTAAHCLGREELEACAFRVQVGQLRLYEDDQRTKVVEI
NOV32n  VPIC-GG-------SLIHPQWVLTAAHCIFRKDTDPSTYRIHTRDVYLYGGRGLLNVSQI

NOV32a  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTG----W
NOV32b  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWAWAPDASW
NOV32c  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWAWAPDASW
NOV32d  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTG----W
NOV32e  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWVTG----W
NOV32f  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTG----W
NOV32g  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWVTG----W
NOV32h  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTG----W
NOV32i  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTG----W
NOV32j  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTG----W
NOV32k  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWVTG----W
NOV32l  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTG----W
NOV32m  VRHPQYNESLSAQGGADIALLKLEAPVPLSELIHPVSLPSASRPGLQTRPGWLPA----A
NOV32n  VVHPNYSVFF---LGADIALLKLATSVRTTNTLAAVALPSLSLEFTDSDNCWNTG----W

NOV32a  GVIG--R-G--------ELLPWP----------------------LSLWEATVKVRSNV
NOV32b  MASCRCRDGWPQARWLRAAGMYYLTALQAERPHSR-R-GQELLPWPLSLWEATVKVRSNV
```

```
NOV32c  MASCRCRDGWPQARWLRAAGMYYLTALQAERPHSR-R-GQELLPWPLSLWEATVKVRSNV
NOV32d  GVIG----R-------GELLPWP----------------------LSLWEATVKVRSNV
NOV32e  GVIG--R-G--------ELLPWP----------------------LSLWEATVKVRSNV
NOV32f  GVIG---RG--------ELLPWP----------------------LSLWEATVKVRSNV
NOV32g  GVIG--R-G--------ELLPWP----------------------LSLWEATVKVRSNV
NOV32h  GVIG--R-G--------ELLPWP----------------------LSLWEATVKVRSNV
NOV32i  GVIG--RGGQEQD---HLGGMWR------DDPECRCRPGLQTRPGWLPAAAETDGHRPGG
NOV32j  GVIW----G-------HVFLLPP----------------------PHLRAAEGPIMR--
NOV32k  GVIG--RGGQEQD---HLGGMWR------DDPECRCRPGLQTRPGWLPAAAETDGHRPGG
NOV32l  GVIG----R-------GELLPWP----------------------LSLWEATVKVRSNV
NOV32m  AETD----G-------QELLPWP----------------------LSLWEATVKVRSNV
NOV32n  GMVG----L---------LDMLP----------------------PPYRPQQVKVLT--

NOV32a  LCNQTCRRRFPSNHTERFER-LIKDDMLCAGDGNHGSWPGDNGGPLLCGRNCTWVQVEVV
NOV32b  LCNQTCRRRFPSNHTERFER-LIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVV
NOV32c  LCNQTCRRRFPSNHTERFER-LIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVV
NOV32d  LCNQTCRRRFPSNHTERFER-LIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVV
NOV32e  LCNQTCRRRFPSNHTERFER-LIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVV
NOV32f  LCNQTCRRRFPSNHTERFER-LIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVV
NOV32g  LCNQTCRRRFPSNHTERFER-LIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVV
NOV32h  LCNQTCRRRFPSNHTERFER-LIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVV
NOV32i  SEQQECT----------ISR-LCRRNVHTPGVG-RNYCPG----PSACGRRR--------
NOV32j  --TRACERMYHKGPTAHVT--IIKAAMPCAGAERHLSPQGDNGGPLLCRRNCTWVQVEVV
NOV32k  SEQQECT----------ISR-LCRRNVHTPGVG-RNYCPG----PSACGRRR--------
NOV32l  LCNQTCRRRFPSNHTERFER-LIKDDMLCAGDERHLSPQGDNGGPLLCRRNCTWVQVEVV
NOV32m  LCNQTCRRRFPSNHTERFER-LIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVV
NOV32n  LSNADCERQTYDAFPGAGDRKFIQDDMICAGRTGRRTWKGDSGGPLVCKKKGTWLQAGVV

NOV32a  SWGKFCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR------------------------
NOV32b  SWGKFCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR------------------------
NOV32c  SWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRRVDA--------------------
NOV32d  SWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR------------------------
NOV32e  SWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR------------------------
NOV32f  SWGKLCGLRGYPGMYTRVTSYVSWIRQPCPSAQTPAVVRRFVLPPNPDVEALTPSVMGSG
NOV32g  SWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR------------------------
NOV32h  SWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR------------------------
NOV32i  ------------------------------------------------------------
NOV32j  SWGKLCGLR-YPGMYTRVTSYVSWIRQYVPPFPRR------------------------
NOV32k  ------------------------------------------------------------
NOV32l  SWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR------------------------
NOV32m  SWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR------------------------
NOV32n  SWGFYSDRP-SIGVYTWVQTYVPWILQQMHL----------------------------

NOV32a  ------------------------------------------------------------
NOV32b  ------------------------------------------------------------
NOV32c  ------------------------------------------------------------
NOV32d  ------------------------------------------------------------
NOV32e  ------------------------------------------------------------
NOV32f  APLPPAPDLQEAEVPIMRTRACERMYHKGPTAHGQVTIIKAAMPCAGRKGQGSCQAALRT
NOV32g  ------------------------------------------------------------
NOV32h  ------------------------------------------------------------
NOV32i  ------------------------------------------------------------
NOV32j  ------------------------------------------------------------
NOV32k  ------------------------------------------------------------
```

```
NOV32l  ------------------------------------------------------------
NOV32m  ------------------------------------------------------------
NOV32n  ------------------------------------------------------------

NOV32a  ------------------------------------------------------------
NOV32b  ------------------------------------------------------------
NOV32c  ------------------------------------------------------------
NOV32d  ------------------------------------------------------------
NOV32e  ------------------------------------------------------------
NOV32f  EDLTPTTPNTEVSPRADPRLSQPEDIWPEWAWPVVVGTTMLLLLLFLAVSSLGSCSTGSP
NOV32g  ------------------------------------------------------------
NOV32h  ------------------------------------------------------------
NOV32i  ------------------------------------------------------------
NOV32j  ------------------------------------------------------------
NOV32k  ------------------------------------------------------------
NOV32l  ------------------------------------------------------------
NOV32m  ------------------------------------------------------------
NOV32n  ------------------------------------------------------------

NOV32a  ------------------------------------------------------------
NOV32b  ------------------------------------------------------------
NOV32c  ------------------------------------------------------------
NOV32d  ------------------------------------------------------------
NOV32e  ------------------------------------------------------------
NOV32f  APVPENDLVGIVGGHNTPGEVVVAVGADRRSLHFPEGHRPVHLPDSHQGCVSVRGPGAAE
NOV32g  ------------------------------------------------------------
NOV32h  ------------------------------------------------------------
NOV32i  ------------------------------------------------------------
NOV32j  ------------------------------------------------------------
NOV32k  ------------------------------------------------------------
NOV32l  ------------------------------------------------------------
NOV32m  ------------------------------------------------------------
NOV32n  ------------------------------------------------------------

NOV32a  ------------------------------------------------------------
NOV32b  ------------------------------------------------------------
NOV32c  ------------------------------------------------------------
NOV32d  ------------------------------------------------------------
NOV32e  ------------------------------------------------------------
NOV32f  CQPDRRPPNYSVFFLGADIALLKLATSSLEFTDSDNCWNTGWGMVGLLDMLPPPYRPQQV
NOV32g  ------------------------------------------------------------
NOV32h  ------------------------------------------------------------
NOV32i  ------------------------------------------------------------
NOV32j  ------------------------------------------------------------
NOV32k  ------------------------------------------------------------
NOV32l  ------------------------------------------------------------
NOV32m  ------------------------------------------------------------
NOV32n  ------------------------------------------------------------

NOV32a  ------------------------------------------------------------
NOV32b  ------------------------------------------------------------
NOV32c  ------------------------------------------------------------
NOV32d  ------------------------------------------------------------
NOV32e  ------------------------------------------------------------
```

```
NOV32f  KVLTLSNADCERQTYDAFPGAGDRKFIQDDMICAGRTGRRTWKGDSGGPLVCKKKGTWLQ
NOV32g  ------------------------------------------------------------
NOV32h  ------------------------------------------------------------
NOV32i  ------------------------------------------------------------
NOV32j  ------------------------------------------------------------
NOV32k  ------------------------------------------------------------
NOV32l  ------------------------------------------------------------
NOV32m  ------------------------------------------------------------
NOV32n  ------------------------------------------------------------

NOV32a  ------------------------------------------------------------
NOV32b  ------------------------------------------------------------
NOV32c  ------------------------------------------------------------
NOV32d  ------------------------------------------------------------
NOV32e  ------------------------------------------------------------
NOV32f  AGVVSWGFYSDRPSIGVYTRPETSWQGANHADAQRPAGRVPTMQRPRDMGQGQEWVCRPF
NOV32g  ------------------------------------------------------------
NOV32h  ------------------------------------------------------------
NOV32i  ------------------------------------------------------------
NOV32j  ------------------------------------------------------------
NOV32k  ------------------------------------------------------------
NOV32l  ------------------------------------------------------------
NOV32m  ------------------------------------------------------------
NOV32n  ------------------------------------------------------------

NOV32a  ------------------------------------------------------------
NOV32b  ------------------------------------------------------------
NOV32c  ------------------------------------------------------------
NOV32d  ------------------------------------------------------------
NOV32e  ------------------------------------------------------------
NOV32f  THVTCYPTAIPRPFTHVTCYLMAVPSTLTHVTCYPTAVPRPFTHVTCYLMAVPSTLTHIT
NOV32g  ------------------------------------------------------------
NOV32h  ------------------------------------------------------------
NOV32i  ------------------------------------------------------------
NOV32j  ------------------------------------------------------------
NOV32k  ------------------------------------------------------------
NOV32l  ------------------------------------------------------------
NOV32m  ------------------------------------------------------------
NOV32n  ------------------------------------------------------------

NOV32a  ------------------------------------------------------------
NOV32b  ------------------------------------------------------------
NOV32c  ------------------------------------------------------------
NOV32d  ------------------------------------------------------------
NOV32e  ------------------------------------------------------------
NOV32f  CYMMAVPRPFTHITCYPMAVPSTLTHVTCHPTAIPRPFTHITCYTMAIPRPSTTPPATRR
NOV32g  ------------------------------------------------------------
NOV32h  ------------------------------------------------------------
NOV32i  ------------------------------------------------------------
NOV32j  ------------------------------------------------------------
NOV32k  ------------------------------------------------------------
NOV32l  ------------------------------------------------------------
NOV32m  ------------------------------------------------------------
NOV32n  ------------------------------------------------------------
```

```
NOV32a  ----------------------------
NOV32b  ----------------------------
NOV32c  ----------------------------
NOV32d  ----------------------------
NOV32e  ----------------------------
NOV32f  PSPAPSPTSPATRWPSPGPSPMSPATR
NOV32g  ----------------------------
NOV32h  ----------------------------
NOV32i  ----------------------------
NOV32j  ----------------------------
NOV32k  ----------------------------
NOV32l  ----------------------------
NOV32m  ----------------------------
NOV32n  ----------------------------

NOV32a  (SEQ ID NO:  382)
NOV32b  (SEQ ID NO:  384)
NOV32c  (SEQ ID NO:  386)
NOV32d  (SEQ ID NO:  388)
NOV32e  (SEQ ID NO:  390)
NOV32f  (SEQ ID NO:  392)
NOV32g  (SEQ ID NO:  394)
NOV32h  (SEQ ID NO:  396)
NOV32i  (SEQ ID NO:  398)
NOV32j  (SEQ ID NO:  400)
NOV32k  (SEQ ID NO:  402)
NOV32l  (SEQ ID NO:  404)
NOV32m  (SEQ ID NO:  406)
NOV32n  (SEQ ID NO:  408)
```

Further analysis of the NOV32a protein yielded the following properties shown in Table 32C.

Table 32C. Protein Sequence Properties NOV32a

SignalP analysis:	Cleavage site between residues 13 and 14

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 12;  peak value   6.67
      PSG score:   2.27

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -6.63
      possible cleavage site: between 34 and 35

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
```

```
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -2.76   Transmembrane  110 - 126
      PERIPHERAL  Likelihood =  1.80 (at 81)
      ALOM score:  -2.76  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 117
      Charge difference: -2.0   C( 0.0) - N( 2.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 2   (cytoplasmic tail 1 to 110)

MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  4.93
      Hyd Moment(95):   1.50    G content:       5
      D/E content:      2       S/T content:     1
      Score: -9.83

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  10.6%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: too long tail

Dileucine motif in the tail: found
      LL at 15
```

```
      LL at 82

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 89

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

---------------------------

Final Results (k = 9/23):

       34.8 %: cytoplasmic
       30.4 %: mitochondrial
       13.0 %: Golgi
        8.7 %: endoplasmic reticulum
        4.3 %: extracellular, including cell wall
        4.3 %: nuclear
        4.3 %: vesicles of secretory system

>> prediction for CG53989-03 is cyt (k=23)
```

A search of the NOV32a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 32D.

Table 32D. Geneseq Results for NOV32a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV32a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAE14347	Human protease PRTS-12 protein - Homo sapiens, 262 aa. [WO200183775-A2, 08-NOV-2001]	130..339 53..262	205/210 (97%) 205/210 (97%)	e-123
AAE08591	Human NOV12 protein - Homo sapiens, 220 aa. [WO200161009-A2, 23-AUG-2001]	126..339 7..220	205/214 (95%) 206/214 (95%)	e-122
AAE08590	Human NOV11 protein - Homo sapiens, 285 aa. [WO200161009-A2, 23-AUG-2001]	135..339 81..285	203/205 (99%) 203/205 (99%)	e-122
AAU82736	Amino acid sequence of novel human protease #35 - Homo sapiens, 948 aa. [WO200200860-A2, 03-JAN-2002]	9..334 46..411	237/368 (64%) 252/368 (68%)	e-118
AAE08587	Human NOV8 protein - Homo sapiens, 290 aa. [WO200161009-A2, 23-AUG-2001]	135..339 86..290	198/205 (96%) 198/205 (96%)	e-116

In a BLAST search of public sequence databases, the NOV32a protein was found to have homology to the proteins shown in the BLASTP data in Table 32E.

Table 32E. Public BLASTP Results for NOV32a				
Protein Accession Number	**Protein/Organism/Length**	**NOV32a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAP21675	Mast cell protease-11 - Mus musculus (Mouse), 318 aa.	106..336 60..284	129/231 (55%) 155/231 (66%)	2e-60
AAA30855	Mastin precursor - Canis sp, 251 aa (fragment).	98..337 22..249	120/240 (50%) 142/240 (59%)	7e-51
P19236	Mastocytoma protease precursor (EC 3.4.21.-) - Canis familiaris (Dog), 269 aa.	98..337 40..267	120/240 (50%) 142/240 (59%)	7e-51
Q8SQ44	Tryptase precursor - Sus scrofa (Pig), 277 aa.	143..337 90..275	106/195 (54%) 128/195 (65%)	1e-49
Q9XSM1	Tryptase (EC 3.4.21.59) - Ovis aries (Sheep), 273 aa.	109..337 54..273	95/229 (41%) 129/229 (55%)	6e-43

PFam analysis indicates that the NOV32a protein contains the domains shown in the Table 32F.

Table 32F. Domain Analysis of NOV32a			
Pfam Domain	**NOV32a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
trypsin	125..329	80/272 (29%) 156/272 (57%)	5.6e-32

Example 33.

The NOV33 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 33A.

Table 33A. NOV33 Sequence Analysis		
NOV33a, CG54203-02 DNA Sequence	SEQ ID NO: 409	1028 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 975

TGCAGAGGGGATATCACATGGAAAAAGCCAATGAGACCTCCCCTGTGATGGGGTTCGTTCTCCTGGGG
CTCTCTGCCCACCCAGAGCTGGAAAAGACATTCTTCGTGCTCATCCTGCTGATGTACCTCGTGATCCT
GCTGGGCAATGGGGTCCTCATCCTGGTGACCATCCTTGACTCCCGCCTGCACACGCCCATGTACTTCT
TCCTAGGGAACCTCTCCTTCCTGGACATCTGCTTCACTACCTCCTCAGTCCCACTGGTCCTGGACAGC
TTTTTGACTCCCCAGGAAACCATCTCCTTCTCAGCCTGTGCTGTGCAGATGGCACTCTCCTTTGCCAT

GGCAGGAACAGAGTGCTTGCTCCTGAGCATGATGGCATTTGATCGCTATGTGGCCATCTGCAACCCCC
TTAGGTACTCCGTGATCATGAGCAAGGCTGCCTACATGCCCATGGCTGCCAGCTCCTGGGCTATTGGT
GGTGCTGCTTCCGTGGTACACACATCCTTGGCAATTCAGCTGCCCTTCTGTGGAGACAATGTCATCAA
CCACTTCACCTGTGAGATTCTGGCTGTTCTAAAGTTGGCCTGTGCTGACATTTCCATCAATGTGATCA
GCATGGAGGTGACGAATGTGATCTTCCTAGGAGTCCCGGTTCTGTTCATCTCTTTCTCCTATGTCTTC
ATCATCACCACCATCCTGAGGATCCCCTCAGCTGAGGGGAGGAAAAAGGTCTTCTCCACCTGCTCTGC
CCACCTCACCGTGGTGATCGTCTTCTACGGGACCTTATTCTTCATGTATGGGAAGCCTAAGTCTAAGG
ACTCCATGGGAGCAGACAAAGAGGATCTTTCAGACAAACTCATCCCCCTTTTCTATGGGGTGGTGACC
CCGATGCTCAACCCCATCATCTATAGCCTGAGGAACAAGGATGTGAAGGCTGCTGTGAGGAGACTGCT
GAGACCAAAAGGCTTCACTCAGTGATGGTGGAAGGGTCCTCTGTGATTGTCACCCACATGGAAGTAAG
GAATCACA

NOV33a, CG54203-02 Protein Sequence	SEQ ID NO: 410	319 aa	MW at 35072.3kD

MEKANETSPVMGFVLLGLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMYFFLGNLS
FLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAFDRYVAICNPLRYSVI
MSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLACADISINVISMEVTN
VIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFSTCSAHLTVVIVFYGTLFFMYGKPKSKDSMGAD
KEDLSDKLIPLFYGVVTPMLNPIIYSLRNKDVKAAVRRLLRPKGFTQ

NOV33b, CG54203-01 DNA Sequence	SEQ ID NO: 411	1031 bp
	ORF Start: ATG at 22	ORF Stop: TGA at 979

TGATGGCAGAGGGGATATCACATGGAAAAAGCCAATGAGACCTCCCCTGTGATGGGGTTCGTTCTCCT
GAGGCTCTCTGCCCACCCAGAGCTGGAAAAGACATTCTTCGTGCTCATCCTGCTGATGTACCTCGTGA
TCCTGCTGGGCAATGGGGTCCTCATCCTGGTGACCATCCTTGACTCCCGCCTGCACACGCCCATGTAC
TTCTTCCTAGGGAACCTCTCCTTCCTGGACATCTGCTTCACTACCTCCTCAGTCCCACTGGTCCTGGA
CAGCTTTTTGACTCCCCAGGAAACCATCTCCTTCTCAGCCTGTGCTGTGCAGATGGCACTCTCCTTTG
CCATGGCAGGAACAGAGTGCTTGCTCCTGAGCATGATGGCATTTGATCGCTATGTGGCCATCTGCAAC
CCCCTTAGGTACTCCGTGATCATGAGCAAGGCTGCCTACATGCCCATGGCTGCCAGCTCCTGGGCTAT
TGGTGGTGCTGCTTCCGTGGTACACACATCCTTGGCAATTCAGCTGCCCTTCTGTGGAGACAATGTCA
TCAACCACTTCACCTGTGAGATTCTGGCTGTTCTAAAGTTGGCCTGTGCTGACATTTCCATCAATGTG
ATCAGCATGGAGGTGACGAATGTGATCTTCCTAGGAGTCCCGGTTCTGTTCATCTCTTTCTCCTATGT
CTTCATCATCACCACCATCCTGAGGATCCCCTCAGCTGAGGGGAGGAAAAAGGTCTTCTCCACCTGCT
CTGCCCACCTCACCGTGGTGATCGTCTTCTACGGGACCTTATTCTTCATGTATGGGAAGCCTAAGTCT
AAGGACTCCATGGGAGCAGACAAAGAGGATCTTTCAGACAAACTCATCCCCCTTTTCTATGGGGTGGT
GACCCCGATGCTCAACCCCATCATCTATAGCCTGAGGAACAAGGATGTGAAGGCTGCTGTGAGGAGAC
TGCTGAGACCAAAAGGCTTCACTCAGTGATGGTGGAAGGGTCCTCTGTGATTGTCACCCACATGGAAG
TAAGGAATCAC

NOV33b, CG54203-01 Protein Sequence	SEQ ID NO: 412	319 aa	MW at 35171.4kD

MEKANETSPVMGFVLLRLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMYFFLGNLS
FLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAFDRYVAICNPLRYSVI
MSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLACADISINVISMEVTN
VIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFSTCSAHLTVVIVFYGTLFFMYGKPKSKDSMGAD
KEDLSDKLIPLFYGVVTPMLNPIIYSLRNKDVKAAVRRLLRPKGFTQ

NOV33c, SNP13382442 of CG54203-02, DNA Sequence	SEQ ID NO: 413	1028 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 975
	SNP Pos: 66	SNP Change: G to A

TGCAGAGGGGATATCACATGGAAAAAGCCAATGAGACCTCCCCTGTGATGGGGTTCGTTCTCCTGAGG
CTCTCTGCCCACCCAGAGCTGGAAAAGACATTCTTCGTGCTCATCCTGCTGATGTACCTCGTGATCCT
GCTGGGCAATGGGGTCCTCATCCTGGTGACCATCCTTGACTCCCGCCTGCACACGCCCATGTACTTCT
TCCTAGGGAACCTCTCCTTCCTGGACATCTGCTTCACTACCTCCTCAGTCCCACTGGTCCTGGACAGC
TTTTTGACTCCCCAGGAAACCATCTCCTTCTCAGCCTGTGCTGTGCAGATGGCACTCTCCTTTGCCAT
GGCAGGAACAGAGTGCTTGCTCCTGAGCATGATGGCATTTGATCGCTATGTGGCCATCTGCAACCCCC
TTAGGTACTCCGTGATCATGAGCAAGGCTGCCTACATGCCCATGGCTGCCAGCTCCTGGGCTATTGGT
GGTGCTGCTTCCGTGGTACACACATCCTTGGCAATTCAGCTGCCCTTCTGTGGAGACAATGTCATCAA

CCACTTCACCTGTGAGATTCTGGCTGTTCTAAAGTTGGCCTGTGCTGACATTTCCATCAATGTGATCA
GCATGGAGGTGACGAATGTGATCTTCCTAGGAGTCCCGGTTCTGTTCATCTCTTTCTCCTATGTCTTC
ATCATCACCACCATCCTGAGGATCCCCTCAGCTGAGGGGAGGAAAAAGGTCTTCTCCACCTGCTCTGC
CCACCTCACCGTGGTGATCGTCTTCTACGGGACCTTATTCTTCATGTATGGGAAGCCTAAGTCTAAGG
ACTCCATGGGAGCAGACAAAGAGGATCTTTCAGACAAACTCATCCCCCTTTTCTATGGGGTGGTGACC
CCGATGCTCAACCCCATCATCTATAGCCTGAGGAACAAGGATGTGAAGGCTGCTGTGAGGAGACTGCT
GAGACCAAAAGGCTTCACTCAG**TGA**TGGTGGAAGGGTCCTCTGTGATTGTCACCCACATGGAAGTAAG
GAATCACA

NOV33c, SNP13382442 of CG54203-02, Protein Sequence	SEQ ID NO: 414	319 aa	MW at 35171.4kD
	SNP Pos: 17		SNP Change: Gly to Arg

MEKANETSPVMGFVLLRLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMYFFLGNLS
FLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAFDRYVAICNPLRYSVI
MSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLACADISINVISMEVTN
VIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFSTCSAHLTVVIVFYGTLFFMYGKPKSKDSMGAD
KEDLSDKLIPLFYGVVTPMLNPIIYSLRNKDVKAAVRRLLRPKGFTQ

NOV33d, SNP13373880 of CG54203-02, DNA Sequence	SEQ ID NO: 415	1028 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 975
	SNP Pos: 324	SNP Change: G to A

TGCAGAGGGGATATCACATGGAAAAAGCCAATGAGACCTCCCCTGTGATGGGGTTCGTTCTCCTGGGG
CTCTCTGCCCACCCAGAGCTGGAAAAGACATTCTTCGTGCTCATCCTGCTGATGTACCTCGTGATCCT
GCTGGGCAATGGGGTCCTCATCCTGGTGACCATCCTTGACTCCCGCCTGCACACGCCCATGTACTTCT
TCCTAGGGAACCTCTCCTTCCTGGACATCTGCTTCACTACCTCCTCAGTCCCACTGGTCCTGGACAGC
TTTTTGACTCCCCAGGAAACCATCTCCTTCTCAGCCTGTGCTGTGCAGATGACACTCTCCTTTGCCAT
GGCAGGAACAGAGTGCTTGCTCCTGAGCATGATGGCATTTGATCGCTATGTGGCCATCTGCAACCCCC
TTAGGTACTCCGTGATCATGAGCAAGGCTGCCTACATGCCCATGGCTGCCAGCTCCTGGGCTATTGGT
GGTGCTGCTTCCGTGGTACACACATCCTTGGCAATTCAGCTGCCCTTCTGTGGAGACAATGTCATCAA
CCACTTCACCTGTGAGATTCTGGCTGTTCTAAAGTTGGCCTGTGCTGACATTTCCATCAATGTGATCA
GCATGGAGGTGACGAATGTGATCTTCCTAGGAGTCCCGGTTCTGTTCATCTCTTTCTCCTATGTCTTC
ATCATCACCACCATCCTGAGGATCCCCTCAGCTGAGGGGAGGAAAAAGGTCTTCTCCACCTGCTCTGC
CCACCTCACCGTGGTGATCGTCTTCTACGGGACCTTATTCTTCATGTATGGGAAGCCTAAGTCTAAGG
ACTCCATGGGAGCAGACAAAGAGGATCTTTCAGACAAACTCATCCCCCTTTTCTATGGGGTGGTGACC
CCGATGCTCAACCCCATCATCTATAGCCTGAGGAACAAGGATGTGAAGGCTGCTGTGAGGAGACTGCT
GAGACCAAAAGGCTTCACTCAG**TGA**TGGTGGAAGGGTCCTCTGTGATTGTCACCCACATGGAAGTAAG
GAATCACA

NOV33d, SNP13373880 of CG54203-02, Protein Sequence	SEQ ID NO: 416	319 aa	MW at 35102.3kD
	SNP Pos: 103		SNP Change: Ala to Thr

MEKANETSPVMGFVLLGLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMYFFLGNLS
FLDICFTTSSVPLVLDSFLTPQETISFSACAVQMTLSFAMAGTECLLLSMMAFDRYVAICNPLRYSVI
MSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLACADISINVISMEVTN
VIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFSTCSAHLTVVIVFYGTLFFMYGKPKSKDSMGAD
KEDLSDKLIPLFYGVVTPMLNPIIYSLRNKDVKAAVRRLLRPKGFTQ

NOV33e, SNP13373881 of CG54203-02, DNA Sequence	SEQ ID NO: 417	1028 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 975
	SNP Pos: 429	SNP Change: A to G

TGCAGAGGGGATATCACATGGAAAAAGCCAATGAGACCTCCCCTGTGATGGGGTTCGTTCTCCTGGGG
CTCTCTGCCCACCCAGAGCTGGAAAAGACATTCTTCGTGCTCATCCTGCTGATGTACCTCGTGATCCT
GCTGGGCAATGGGGTCCTCATCCTGGTGACCATCCTTGACTCCCGCCTGCACACGCCCATGTACTTCT
TCCTAGGGAACCTCTCCTTCCTGGACATCTGCTTCACTACCTCCTCAGTCCCACTGGTCCTGGACAGC
TTTTTGACTCCCCAGGAAACCATCTCCTTCTCAGCCTGTGCTGTGCAGATGGCACTCTCCTTTGCCAT
GGCAGGAACAGAGTGCTTGCTCCTGAGCATGATGGCATTTGATCGCTATGTGGCCATCTGCAACCCCC
TTAGGTACTCCGTGATCATGGGCAAGGCTGCCTACATGCCCATGGCTGCCAGCTCCTGGGCTATTGGT
GGTGCTGCTTCCGTGGTACACACATCCTTGGCAATTCAGCTGCCCTTCTGTGGAGACAATGTCATCAA
CCACTTCACCTGTGAGATTCTGGCTGTTCTAAAGTTGGCCTGTGCTGACATTTCCATCAATGTGATCA

GCATGGAGGTGACGAATGTGATCTTCCTAGGAGTCCCGGTTCTGTTCATCTCTTTCTCCTATGTCTTC
ATCATCACCACCATCCTGAGGATCCCCTCAGCTGAGGGGAGGAAAAAGGTCTTCTCCACCTGCTCTGC
CCACCTCACCGTGGTGATCGTCTTCTACGGGACCTTATTCTTCATGTATGGGAAGCCTAAGTCTAAGG
ACTCCATGGGAGCAGACAAAGAGGATCTTTCAGACAAACTCATCCCCCTTTTCTATGGGGTGGTGACC
CCGATGCTCAACCCCATCATCTATAGCCTGAGGAACAAGGATGTGAAGGCTGCTGTGAGGAGACTGCT
GAGACCAAAAGGCTTCACTCAG**TGA**TGGTGGAAGGGTCCTCTGTGATTGTCACCCACATGGAAGTAAG
GAATCACA

NOV33e, SNP13373881 of CG54203-02, Protein Sequence	SEQ ID NO: 418	319 aa	MW at 35042.3kD
	SNP Pos: 138		SNP Change: Ser to Gly

MEKANETSPVMGFVLLGLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMYFFLGNLS
FLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAFDRYVAICNPLRYSVI
MGKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLACADISINVISMEVTN
VIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFSTCSAHLTVVIVFYGTLFFMYGKPKSKDSMGAD
KEDLSDKLIPLFYGVVTPMLNPIIYSLRNKDVKAAVRRLLRPKGFTQ

NOV33f, SNP13382443 of CG54203-02, DNA Sequence	SEQ ID NO: 419	1028 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 975
	SNP Pos: 444	SNP Change: A to G

TGCAGAGGGGATATCAC**ATG**GAAAAAGCCAATGAGACCTCCCCTGTGATGGGGTTCGTTCTCCTGGGG
CTCTCTGCCCACCCAGAGCTGGAAAAGACATTCTTCGTGCTCATCCTGCTGATGTACCTCGTGATCCT
GCTGGGCAATGGGGTCCTCATCCTGGTGACCATCCTTGACTCCCGCCTGCACACGCCCATGTACTTCT
TCCTAGGGAACCTCTCCTTCCTGGACATCTGCTTCACTACCTCCTCAGTCCCACTGGTCCTGGACAGC
TTTTTGACTCCCCAGGAAACCATCTCCTTCTCAGCCTGTGCTGTGCAGATGGCACTCTCCTTTGCCAT
GGCAGGAACAGAGTGCTTGCTCCTGAGCATGATGGCATTTGATCGCTATGTGGCCATCTGCAACCCCC
TTAGGTACTCCGTGATCATGAGCAAGGCTGCCTAC**G**TGCCCATGGCTGCCAGCTCCTGGGCTATTGGT
GGTGCTGCTTCCGTGGTACACACATCCTTGGCAATTCAGCTGCCCTTCTGTGGAGACAATGTCATCAA
CCACTTCACCTGTGAGATTCTGGCTGTTCTAAAGTTGGCCTGTGCTGACATTTCCATCAATGTGATCA
GCATGGAGGTGACGAATGTGATCTTCCTAGGAGTCCCGGTTCTGTTCATCTCTTTCTCCTATGTCTTC
ATCATCACCACCATCCTGAGGATCCCCTCAGCTGAGGGGAGGAAAAAGGTCTTCTCCACCTGCTCTGC
CCACCTCACCGTGGTGATCGTCTTCTACGGGACCTTATTCTTCATGTATGGGAAGCCTAAGTCTAAGG
ACTCCATGGGAGCAGACAAAGAGGATCTTTCAGACAAACTCATCCCCCTTTTCTATGGGGTGGTGACC
CCGATGCTCAACCCCATCATCTATAGCCTGAGGAACAAGGATGTGAAGGCTGCTGTGAGGAGACTGCT
GAGACCAAAAGGCTTCACTCAG**TGA**TGGTGGAAGGGTCCTCTGTGATTGTCACCCACATGGAAGTAAG
GAATCACA

NOV33f, SNP13382443 of CG54203-02, Protein Sequence	SEQ ID NO: 420	319 aa	MW at 35040.2kD
	SNP Pos: 143		SNP Change: Met to Val

MEKANETSPVMGFVLLGLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMYFFLGNLS
FLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAFDRYVAICNPLRYSVI
MSKAAYVPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLACADISINVISMEVTN
VIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFSTCSAHLTVVIVFYGTLFFMYGKPKSKDSMGAD
KEDLSDKLIPLFYGVVTPMLNPIIYSLRNKDVKAAVRRLLRPKGFTQ

NOV33g, SNP13373844 of CG54203-02, DNA Sequence	SEQ ID NO: 421	1028 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 975
	SNP Pos: 719	SNP Change: G to A

TGCAGAGGGGATATCAC**ATG**GAAAAAGCCAATGAGACCTCCCCTGTGATGGGGTTCGTTCTCCTGGGG
CTCTCTGCCCACCCAGAGCTGGAAAAGACATTCTTCGTGCTCATCCTGCTGATGTACCTCGTGATCCT
GCTGGGCAATGGGGTCCTCATCCTGGTGACCATCCTTGACTCCCGCCTGCACACGCCCATGTACTTCT
TCCTAGGGAACCTCTCCTTCCTGGACATCTGCTTCACTACCTCCTCAGTCCCACTGGTCCTGGACAGC
TTTTTGACTCCCCAGGAAACCATCTCCTTCTCAGCCTGTGCTGTGCAGATGGCACTCTCCTTTGCCAT
GGCAGGAACAGAGTGCTTGCTCCTGAGCATGATGGCATTTGATCGCTATGTGGCCATCTGCAACCCCC
TTAGGTACTCCGTGATCATGAGCAAGGCTGCCTACATGCCCATGGCTGCCAGCTCCTGGGCTATTGGT
GGTGCTGCTTCCGTGGTACACACATCCTTGGCAATTCAGCTGCCCTTCTGTGGAGACAATGTCATCAA
CCACTTCACCTGTGAGATTCTGGCTGTTCTAAAGTTGGCCTGTGCTGACATTTCCATCAATGTGATCA
GCATGGAGGTGACGAATGTGATCTTCCTAGGAGTCCCGGTTCTGTTCATCTCTTTCTCCTATGTCTTC

ATCATCACCACCATCCTGAGGATCCCCTCAGCTGAGGGAAGGAAAAAGGTCTTCTCCACCTGCTCTGC
CCACCTCACCGTGGTGATCGTCTTCTACGGGACCTTATTCTTCATGTATGGGAAGCCTAAGTCTAAGG
ACTCCATGGGAGCAGACAAAGAGGATCTTTCAGACAAACTCATCCCCCTTTTCTATGGGGTGGTGACC
CCGATGCTCAACCCCATCATCTATAGCCTGAGGAACAAGGATGTGAAGGCTGCTGTGAGGAGACTGCT
GAGACCAAAAGGCTTCACTCAGTGATGGTGGAAGGGTCCTCTGTGATTGTCACCCACATGGAAGTAAG
GAATCACA

NOV33g, SNP13373844 of CG54203-02, Protein Sequence	SEQ ID NO: 422	319 aa	MW at 35072.3kD
	SNP Pos: 234		SNP Change: Gly to Gly

MEKANETSPVMGFVLLGLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMYFFLGNLS
FLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAFDRYVAICNPLRYSVI
MSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLACADISINVISMEVTN
VIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFSTCSAHLTVVIVFYGTLFFMYGKPKSKDSMGAD
KEDLSDKLIPLFYGVVTPMLNPIIYSLRNKDVKAAVRRLLRPKGFTQ

NOV33h, SNP13373883 of CG54203-02, DNA Sequence	SEQ ID NO: 423	1028 bp
	ORF Start: ATG at 18	ORF Stop: TGA at 975
	SNP Pos: 758	SNP Change: C to T

TGCAGAGGGGATATCACATGGAAAAAGCCAATGAGACCTCCCCTGTGATGGGGTTCGTTCTCCTGGGG
CTCTCTGCCCACCCAGAGCTGGAAAAGACATTCTTCGTGCTCATCCTGCTGATGTACCTCGTGATCCT
GCTGGGCAATGGGGTCCTCATCCTGGTGACCATCCTTGACTCCCGCCTGCACACGCCCATGTACTTCT
TCCTAGGGAACCTCTCCTTCCTGGACATCTGCTTCACTACCTCCTCAGTCCCACTGGTCCTGGACAGC
TTTTTGACTCCCCAGGAAACCATCTCCTTCTCAGCCTGTGCTGTGCAGATGGCACTCTCCTTTGCCAT
GGCAGGAACAGAGTGCTTGCTCCTGAGCATGATGGCATTTGATCGCTATGTGGCCATCTGCAACCCCC
TTAGGTACTCCGTGATCATGAGCAAGGCTGCCTACATGCCCATGGCTGCCAGCTCCTGGGCTATTGGT
GGTGCTGCTTCCGTGGTACACACATCCTTGGCAATTCAGCTGCCCTTCTGTGGAGACAATGTCATCAA
CCACTTCACCTGTGAGATTCTGGCTGTTCTAAAGTTGGCCTGTGCTGACATTTCCATCAATGTGATCA
GCATGGAGGTGACGAATGTGATCTTCCTAGGAGTCCCGGTTCTGTTCATCTCTTTCTCCTATGTCTTC
ATCATCACCACCATCCTGAGGATCCCCTCAGCTGAGGGGAGGAAAAAGGTCTTCTCCACCTGCTCTGC
CCACCTCACTGTGGTGATCGTCTTCTACGGGACCTTATTCTTCATGTATGGGAAGCCTAAGTCTAAGG
ACTCCATGGGAGCAGACAAAGAGGATCTTTCAGACAAACTCATCCCCCTTTTCTATGGGGTGGTGACC
CCGATGCTCAACCCCATCATCTATAGCCTGAGGAACAAGGATGTGAAGGCTGCTGTGAGGAGACTGCT
GAGACCAAAAGGCTTCACTCAGTGATGGTGGAAGGGTCCTCTGTGATTGTCACCCACATGGAAGTAAG
GAATCACA

NOV33h, SNP13373883 of CG54203-02, Protein Sequence	SEQ ID NO: 424	319 aa	MW at 35072.3kD
	SNP Pos: 247		SNP Change: Thr to Thr

MEKANETSPVMGFVLLGLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPMYFFLGNLS
FLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMAFDRYVAICNPLRYSVI
MSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTCEILAVLKLACADISINVISMEVTN
VIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFSTCSAHLTVVIVFYGTLFFMYGKPKSKDSMGAD
KEDLSDKLIPLFYGVVTPMLNPIIYSLRNKDVKAAVRRLLRPKGFTQ

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 33B.

Table 33B. Comparison of the NOV33 protein sequences.

```
NOV33a  MEKANETSPVMGFVLLGLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPM
NOV33b  MEKANETSPVMGFVLLRLSAHPELEKTFFVLILLMYLVILLGNGVLILVTILDSRLHTPM

NOV33a  YFFLGNLSFLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMA
NOV33b  YFFLGNLSFLDICFTTSSVPLVLDSFLTPQETISFSACAVQMALSFAMAGTECLLLSMMA

NOV33a  FDRYVAICNPLRYSVIMSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTC
```

```
NOV33b  FDRYVAICNPLRYSVIMSKAAYMPMAASSWAIGGAASVVHTSLAIQLPFCGDNVINHFTC

NOV33a  EILAVLKLACADISINVISMEVTNVIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFS
NOV33b  EILAVLKLACADISINVISMEVTNVIFLGVPVLFISFSYVFIITTILRIPSAEGRKKVFS

NOV33a  TCSAHLTVVIVFYGTLFFMYGKPKSKDSMGADKEDLSDKLIPLFYGVVTPMLNPIIYSLR
NOV33b  TCSAHLTVVIVFYGTLFFMYGKPKSKDSMGADKEDLSDKLIPLFYGVVTPMLNPIIYSLR

NOV33a  NKDVKAAVRRLLRPKGFTQ
NOV33b  NKDVKAAVRRLLRPKGFTQ

NOV33a  (SEQ ID NO:  410)
NOV33b  (SEQ ID NO:  412)
```

Further analysis of the NOV33a protein yielded the following properties shown in Table 33C.

Table 33C. Protein Sequence Properties NOV33a

SignalP analysis:	Cleavage site between residues 45 and 46

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 6;  pos.chg 1;  neg.chg 2
      H-region:  length 16;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.59
      possible cleavage site: between 44 and 45

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   7
      INTEGRAL    Likelihood = -9.50   Transmembrane   31 -  47
      INTEGRAL    Likelihood = -0.22   Transmembrane   67 -  83
      INTEGRAL    Likelihood = -0.32   Transmembrane   93 - 109
      INTEGRAL    Likelihood = -3.77   Transmembrane  182 - 198
      INTEGRAL    Likelihood = -6.95   Transmembrane  206 - 222
      INTEGRAL    Likelihood = -1.81   Transmembrane  242 - 258
      INTEGRAL    Likelihood = -0.22   Transmembrane  280 - 296
      PERIPHERAL  Likelihood =  2.38 (at 151)
      ALOM score:  -9.50  (number of TMSs: 7)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 38
      Charge difference:  1.0   C( 0.5) - N(-0.5)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b
```

```
MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  6.12
      Hyd Moment(95): 11.03     G content:       0
      D/E content:      2       S/T content:     0
      Score: -6.47

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.2%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: KGFT

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
```

```
      total: 0 residues

-------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       11.1 %: mitochondrial
       11.1 %: vacuolar
       11.1 %: vesicles of secretory system
       11.1 %: Golgi

>> prediction for CG54203-02 is end (k=9)
```

A search of the NOV33a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 33D.

5

Table 33D. Geneseq Results for NOV33a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV33a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU85170	G-coupled olfactory receptor #31 - Homo sapiens, 319 aa. [WO200198526-A2, 27-DEC-2001]	1..319 1..319	318/319 (99%) 318/319 (99%)	e-179
AAU97924	Novel odourant receptor NOV3 protein - Unidentified, 319 aa. [WO200236632-A2, 10-MAY-2002]	1..319 1..319	318/319 (99%) 318/319 (99%)	e-179
AAG71989	Human olfactory receptor polypeptide, SEQ ID NO: 1670 - Homo sapiens, 319 aa. [WO200127158-A2, 19-APR-2001]	1..319 1..319	318/319 (99%) 318/319 (99%)	e-179
AAU07096	Human odorant receptor (OR) polypeptide #13 - Homo sapiens, 319 aa. [WO200157215-A2, 09-AUG-2001]	1..319 1..319	318/319 (99%) 318/319 (99%)	e-179
AAU07086	Human odorant receptor (OR) polypeptide #3 - Homo sapiens, 319 aa. [WO200157215-A2, 09-AUG-2001]	1..319 1..319	318/319 (99%) 318/319 (99%)	e-179

In a BLAST search of public sequence databases, the NOV33a protein was found to have homology to the proteins shown in the BLASTP data in Table 33E.

Table 33E. Public BLASTP Results for NOV33a

Protein Accession Number	Protein/Organism/Length	NOV33a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NQN1	Olfactory receptor 2S2 - Homo sapiens (Human), 319 aa.	1..319 1..319	318/319 (99%) 318/319 (99%)	e-179
CAC88321	Sequence 3 from Patent WO0164879 - Homo sapiens (Human), 318 aa.	1..317 1..317	263/317 (82%) 289/317 (90%)	e-151
Q9QZ22	Olfactory receptor GA_x5J8B7W5BNN-979337-980296 - Mus musculus (Mouse), 319 aa.	1..319 1..319	263/319 (82%) 292/319 (91%)	e-150
CAC88320	Sequence 1 from Patent WO0164879 - Homo sapiens (Human), 319 aa.	1..317 1..318	263/318 (82%) 288/318 (89%)	e-149
Q9QZ19	Olfactory receptor (Olfactory receptor MOR262-5) - Mus musculus (Mouse), 319 aa.	1..317 1..317	257/317 (81%) 283/317 (89%)	e-147

PFam analysis indicates that the NOV33a protein contains the domains shown in the Table 33F.

Table 33F. Domain Analysis of NOV33a

Pfam Domain	NOV33a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1	42..297	57/280 (20%) 179/280 (64%)	3.7e-35

Example 34.

The NOV34 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 34A.

Table 34A. NOV34 Sequence Analysis			
NOV34a, CG54212-02 DNA Sequence	SEQ ID NO: 425		953 bp
	ORF Start: ATG at 11		ORF Stop: TGA at 947

ACTCTGCAGC**ATG**GAGCCGCTCAACAGAACAGAGGTGTCCGAGTTCTTTCTGAAAGGATTTTCTGGCT
ACCCAGCCCTGGAGCATCTGCTCTTCCCTCTGTGCTCAGCCATGTACCTGGTGACCCTCCTGGGGAAC
ACAGCCATCATGGCGGTGAGCGTGCTAGATATCCACCTGCACACGCCTGTGTACTTCTTCCTGGGCAA
CCTCTCTACCCTGGACATCTGCTACACGCCCACCTTTGTGCCTCTGATGCTGGTCCACCTCCTGTCAT
CCCGGAAGACCATCTCCTTTGCTGTCTGTGCCATCCAGATGTGTCTGAGCCTGTCCACGGGCTCCACG
GAGTGCCTGCTACTGGCCATCACGGCCTATGACCGCTACCTGGCCATCTGCCAGCCACTCAGGTACCG
CGTGCTCATGAGCCACCGGCTCTGCGTGCTGCTGATGGGAGCTGCCTGGGTCCTCTGCCTCCTCAAGT
CGGTGACTGAGATGGTCATCTCCATGAGGCTGCCCTTCTGTGGCCACCACGTGGTCAGTCACTTCACC
TGCAAGATCCTGGCAGTGCTGAAGCTGGCATGCGGCAACACGTCGGTCAGCGAAGACTTCCTGCTGGC
GGGCTCCATCCTGCTGCTGCCTGTACCCCTGGCATTCATCTGCCTGTCCTACTTGCTCATCCTGGCCA
CCATCCTGAGGGTGCCCTCGGCCGCCAGGTGCTGCAAAGCCTTCTCCACCTGCTTGGCACACCTGGCT
GTAGTGCTGCTTTTCTACGGCACCATCATCTTCATGTACTTGAAGCCCAAGAGTAAGGAAGCCCACAT
CTCTGATGAGGTCTTCACAGTCCTCTATGCCATGGTCACGACCATGCTGAACCCCACCATCTACAGCC
TGAGGAACAAGGAGGTGAAGGAGGCCGCCAGGAAGGTGTGGGGCAGGAGTCGGGCCTCCAGG**TGA**GGG
A

NOV34a, CG54212-02 Protein Sequence	SEQ ID NO: 426	312 aa	MW at 34707.2kD

MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVYFFLGNLST
LDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYRVLM
SHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLLAGSI
LLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLLFYGTIIFMYLKPKSKEAHISDE
VFTVLYAMVTTMLNPTIYSLRNKEVKEAARKVWGRSRASR

NOV34b, CG54212-01 DNA Sequence	SEQ ID NO: 427	1050 bp
	ORF Start: ATG at 59	ORF Stop: TGA at 995

CCCTGTACCCTCTCTCCTTCCATCCCAGCTGTGGACCATCTCTTCAGAACTCTGCAGC**ATG**GAGCCGC
TCAACAGAACAGAGGTGTCCGAGTTCTTTCTGAAAGGATTTTCTGGCTACCCAGCCCTGGAGCATCTG
CTCTTCCCTCTGTGCTCAGCCATGTACCTGGTGACCCTCCTGGGGAACACAGCCATCATGGCGGTGAG
CGTGCTAGATATCCACCTGCACACGCCCGTGTACTTCTTCCTGGGCAACCTCTCTACCCTGGACATCT
GCTACACGCCCACCTTTGTGCCTCTGATGCTGGTCCACCTCCTGTCATCCCGGAAGACCATCTCCTTT
GCTGTCTGTGCCATCCAGATGTGTCTGAGCCTGTCCACGGGCTCCACGGAGTGCCTGCTACTGGCCAT
CACGGCCTATGACCGCTACCTGGCCATCTGCCAGCCACTCAGGTACCACGTGCTCATGAGCCACCGGC
TCTGCGTGCTGCTGATGGGAGCTGCCTGGGTCCTCTGCCTCCTCAAGTCGGTGACTGAGATGGTCATC
TCCATGAGGCTGCCCTTCTGTGGCCACCACGTGGTCAGTCACTTCACCTGCAAGATCCTGGCAGTGCT
GAAGCTGGCATGCGGCAACACGTCGGTCAGCGAAGACTTCCTGCTGGCGGGCTCCATCCTGCTGCTGC
CTGTACCCCTGGCATTCATCTGCCTGTCCTACTTGCTCATCCTGGCCACCATCCTGAGGGTGCCCTCG
GCCGCCAGGTGCTGCAAAGCCTTCTCCACCTGCTTGGCACACCTGGCTGTAGTGCTGCTTTTCTACGG
CACCATCATCTTCATGTACTTGAAGCCCAAGAGTAAGGAAGCCCACATCTCTGATGAGGTCTTCACAG
TCCTCTATGCCATGGTCACGACCATGCTGAACCCCACCATCTACAGCCTGAGGAACAAGGAGGTGAAG
GAGGCCGCCAGGAAGGTGTGGGGCAGGAGTCGGGCCTCCAGG**TGA**GGGAGGGCGGGGCTCTGTACAGA
CGCAGGTCTCAGGTTAGTAGCTGAGGCCAT

NOV34b, CG54212-01 Protein Sequence	SEQ ID NO: 428	312 aa	MW at 34688.2kD

MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVYFFLGNLST
LDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLM
SHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLLAGSI
LLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLLFYGTIIFMYLKPKSKEAHISDE
VFTVLYAMVTTMLNPTIYSLRNKEVKEAARKVWGRSRASR

NOV34c, CG54212-03 DNA Sequence	SEQ ID NO: 429	917 bp
	ORF Start: at 3	ORF Stop: TGA at 861

TGCTCTTCCCTCTGTGCTCAGCCATGTACCTGGTGACCCTCCTGGGGAACACAGCCATCATGGCGGTG

AGCGTGCTAGATATCCACCTGCACACGCCCGTGTACTTCTTCCTGGGCAACCTCTCTACCCTGGACAT
CTGCTACACGCCCACCTTTGTGCCTCTGATGCTGGTCCACCTCCTGTCATCCCGGAAGACCATCTCCT
TTGCTGTCTGTGCCATCCAGATGTGTCTGAGCCTGTCCACGGGCTCCACGGAGTGCCTGCTACTGGCC
ATCACGGCCTATGACCGCTACCTGGCCATCTGCCAGCCACTCAGGTACCACGTGCTCATGAGCCACCG
GCTCTGCGTGCTGCTGATGGGAGCTGCCTGGGTCCTCTGCCTCCTCAAGTCGGTGACTGAGATGGTCA
TCTCCATGAGGCTGCCCTTCTGTGGCCACCACGTGGTCAGTCACTTCACCTGCAAGATCCTGGCAGTG
CTGAAGCTGGCATGCGGCAACACGTCGGTCAGCGAAGACTTCCTGCTGGCGGGCTCCATCCTGCTGCT
GCCTGTACCCCTGGCATTCATCTGCCTGTCCTACTTGCTCATCCTGGCCACCATCCTGAGGGTGCCCT
CGGCCGCCAGGTGCTGCAAAGCCTTCTCCACCTGCTTGGCACACCTGGCTGTAGTGCTGCTTTTCTAC
GGCACCATCATCTTCATGTACTTGAAGCCCAAGAGTAAGGAAGCCCACATCTCTGATGAGGTCTTCAC
AGTCCTCTATGCCATGGTCACGACCATGCTGAACCCCACCATCTACAGCCTGAGGAACAAGGAGGTGA
AGGAGGCCGCCAGGAAGGTGTGGGGCAGGAGTCGGGCCTCCAGGTGAGGGAGGGCGGGGCTCTGTACA
GACGCAGGTCTCAGGTTAGTAGCTGAGGCCATC

NOV34c, CG54212-03 Protein Sequence	SEQ ID NO: 430	286 aa	MW at 31693.8kD

LFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVYFFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISF
AVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLLKSVTEMVI
SMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPS
AARCCKAFSTCLAHLAVVLLFYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVK
EAARKVWGRSRASR

NOV34d, CG54212-04 DNA Sequence	SEQ ID NO: 431	1025 bp
	ORF Start: ATG at 33	ORF Stop: TGA at 969

AGCTGTGGACCATCTCTTCAGAACTCTGCAGCATGGAGCCGCTCAACAGAACAGAGGTGTCCGAGTTC
TTTCTGAAAGGATTTTCTGGCTACCCAGCCCTGGAGCATCTGCTCTTCCCTCTGTGCTCAGCCATGTA
CCTGGTGACCCTCCTGGGGAACACAGCCATCATGGCGGTGAGCGTGCTAGATATCCACCTGCACACGC
CCGTGTACTTCTTCCTGGGCAACCTCTCTACCCTGGACATCTGCTACACGCCCACCTTTGTGCCTCTG
ATGCTGGTCCACCTCCTGTCATCCCGGAAGACCATCTCCTTTGCTGTCTGTGCCATCCAGATGTGTCT
GAGCCTGTCCACGGGCTCCACGGAGTGCCTGCTACTGGCCATCACGGCCTATGACCGCTACCTGGCCA
TCTGCCAGCCACTCAGGTACCACGTGCTCATGAGCCACCGGCTCTGCGTGCTGCTGATGGGAGCTGCC
TGGGTCCTCTGCCTCCTCAAGTCGGTGACTGAGATGGTCATCTCCATGAGGCTGCCCTTCTGTGGCCA
CCACGTGGTCAGTCACTTCACCTGCAAGATCCTGGCAGTGCTGAAGCTGGCATGCGGCAACACGTCGG
TCAGCGAAGACTTCCTGCTGGCGGGCTCCATCCTGCTGCTGCCTGTACCCCTGGCATTCATCTGCCTG
TCCTACTTGCTCATCCTGGCCACCATCCTGAGGGTGCCCTCGGCCGCCAGGTGCTGCAAAGCCTTCTC
CACCTGCTTGGCACACCTGGCTGTAGTGCTGCTTTTCTACGGCACCATCATCTTCATGTACTTGAAGC
CCAAGAGTAAGGAAGCCCACATCTCTGATGAGGTCTTCACAGTCCTCTATGCCATGGTCACGACCATG
CTGAACCCCACCATCTACAGCCTGAGGAACAAGGAGGTGAAGGAGGCCGCCAGGAAGGTGTGGGGCAG
GAGTCGGGCCTCCAGGTGAGGGAGGGCGGGGCTCTGTACAGACGCAGGTCTCAGGTTAGTAGCTGAGG
CCATC

NOV34d, CG54212-04 Protein Sequence	SEQ ID NO: 432	312 aa	MW at 34688.2kD

MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVYFFLGNLST
LDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLM
SHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLLAGSI
LLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLLFYGTIIFMYLKPKSKEAHISDE
VFTVLYAMVTTMLNPTIYSLRNKEVKEAARKVWGRSRASR

NOV34e, SNP13373981 of CG54212-02, DNA Sequence	SEQ ID NO: 433	953 bp
	ORF Start: ATG at 11	ORF Stop: TGA at 947
	SNP Pos: 184	SNP Change: T to C

ACTCTGCAGCATGGAGCCGCTCAACAGAACAGAGGTGTCCGAGTTCTTTCTGAAAGGATTTTCTGGCT
ACCCAGCCCTGGAGCATCTGCTCTTCCCTCTGTGCTCAGCCATGTACCTGGTGACCCTCCTGGGGAAC
ACAGCCATCATGGCGGTGAGCGTGCTAGATATCCACCTGCACACGCCCGTGTACTTCTTCCTGGGCAA
CCTCTCTACCCTGGACATCTGCTACACGCCCACCTTTGTGCCTCTGATGCTGGTCCACCTCCTGTCAT
CCCGGAAGACCATCTCCTTTGCTGTCTGTGCCATCCAGATGTGTCTGAGCCTGTCCACGGGCTCCACG
GAGTGCCTGCTACTGGCCATCACGGCCTATGACCGCTACCTGGCCATCTGCCAGCCACTCAGGTACCG

CGTGCTCATGAGCCACCGGCTCTGCGTGCTGCTGATGGGAGCTGCCTGGGTCCTCTGCCTCCTCAAGT
CGGTGACTGAGATGGTCATCTCCATGAGGCTGCCCTTCTGTGGCCACCACGTGGTCAGTCACTTCACC
TGCAAGATCCTGGCAGTGCTGAAGCTGGCATGCGGCAACACGTCGGTCAGCGAAGACTTCCTGCTGGC
GGGCTCCATCCTGCTGCTGCCTGTACCCCTGGCATTCATCTGCCTGTCCTACTTGCTCATCCTGGCCA
CCATCCTGAGGGTGCCCTCGGCCGCCAGGTGCTGCAAAGCCTTCTCCACCTGCTTGGCACACCTGGCT
GTAGTGCTGCTTTTCTACGGCACCATCATCTTCATGTACTTGAAGCCCAAGAGTAAGGAAGCCCACAT
CTCTGATGAGGTCTTCACAGTCCTCTATGCCATGGTCACGACCATGCTGAACCCCACCATCTACAGCC
TGAGGAACAAGGAGGTGAAGGAGGCCGCCAGGAAGGTGTGGGGCAGGAGTCGGGCCTCCAGGTGAGGG
A

NOV34e, SNP13373981 of CG54212-02, Protein Sequence	SEQ ID NO: 434	312 aa	MW at 34707.2kD
	SNP Pos: 58		SNP Change: Pro to Pro

MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVYFFLGNLST
LDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYRVLM
SHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLLAGSI
LLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLLFYGTIIFMYLKPKSKEAHISDE
VFTVLYAMVTTMLNPTIYSLRNKEVKEAARKVWGRSRASR

NOV34f, SNP13373982 of CG54212-02, DNA Sequence	SEQ ID NO: 435	953 bp
	ORF Start: ATG at 11	ORF Stop: TGA at 947
	SNP Pos: 188	SNP Change: T to C

ACTCTGCAGCATGGAGCCGCTCAACAGAACAGAGGTGTCCGAGTTCTTTCTGAAAGGATTTTCTGGCT
ACCCAGCCCTGGAGCATCTGCTCTTCCCTCTGTGCTCAGCCATGTACCTGGTGACCCTCCTGGGGAAC
ACAGCCATCATGGCGGTGAGCGTGCTAGATATCCACCTGCACACGCCTGTGCACTTCTTCCTGGGCAA
CCTCTCTACCCTGGACATCTGCTACACGCCCACCTTTGTGCCTCTGATGCTGGTCCACCTCCTGTCAT
CCCGGAAGACCATCTCCTTTGCTGTCTGTGCCATCCAGATGTGTCTGAGCCTGTCCACGGGCTCCACG
GAGTGCCTGCTACTGGCCATCACGGCCTATGACCGCTACCTGGCCATCTGCCAGCCACTCAGGTACCG
CGTGCTCATGAGCCACCGGCTCTGCGTGCTGCTGATGGGAGCTGCCTGGGTCCTCTGCCTCCTCAAGT
CGGTGACTGAGATGGTCATCTCCATGAGGCTGCCCTTCTGTGGCCACCACGTGGTCAGTCACTTCACC
TGCAAGATCCTGGCAGTGCTGAAGCTGGCATGCGGCAACACGTCGGTCAGCGAAGACTTCCTGCTGGC
GGGCTCCATCCTGCTGCTGCCTGTACCCCTGGCATTCATCTGCCTGTCCTACTTGCTCATCCTGGCCA
CCATCCTGAGGGTGCCCTCGGCCGCCAGGTGCTGCAAAGCCTTCTCCACCTGCTTGGCACACCTGGCT
GTAGTGCTGCTTTTCTACGGCACCATCATCTTCATGTACTTGAAGCCCAAGAGTAAGGAAGCCCACAT
CTCTGATGAGGTCTTCACAGTCCTCTATGCCATGGTCACGACCATGCTGAACCCCACCATCTACAGCC
TGAGGAACAAGGAGGTGAAGGAGGCCGCCAGGAAGGTGTGGGGCAGGAGTCGGGCCTCCAGGTGAGGG
A

NOV34f, SNP13373982 of CG54212-02, Protein Sequence	SEQ ID NO: 436	312 aa	MW at 34681.2kD
	SNP Pos: 60		SNP Change: Tyr to His

MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVHFFLGNLST
LDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYRVLM
SHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLLAGSI
LLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLLFYGTIIFMYLKPKSKEAHISDE
VFTVLYAMVTTMLNPTIYSLRNKEVKEAARKVWGRSRASR

NOV34g, SNP13373984 of CG54212-02, DNA Sequence	SEQ ID NO: 437	953 bp
	ORF Start: ATG at 11	ORF Stop: TGA at 947
	SNP Pos: 408	SNP Change: G to A

ACTCTGCAGCATGGAGCCGCTCAACAGAACAGAGGTGTCCGAGTTCTTTCTGAAAGGATTTTCTGGCT
ACCCAGCCCTGGAGCATCTGCTCTTCCCTCTGTGCTCAGCCATGTACCTGGTGACCCTCCTGGGGAAC
ACAGCCATCATGGCGGTGAGCGTGCTAGATATCCACCTGCACACGCCTGTGTACTTCTTCCTGGGCAA
CCTCTCTACCCTGGACATCTGCTACACGCCCACCTTTGTGCCTCTGATGCTGGTCCACCTCCTGTCAT
CCCGGAAGACCATCTCCTTTGCTGTCTGTGCCATCCAGATGTGTCTGAGCCTGTCCACGGGCTCCACG
GAGTGCCTGCTACTGGCCATCACGGCCTATGACCGCTACCTGGCCATCTGCCAGCCACTCAGGTACCA
CGTGCTCATGAGCCACCGGCTCTGCGTGCTGCTGATGGGAGCTGCCTGGGTCCTCTGCCTCCTCAAGT
CGGTGACTGAGATGGTCATCTCCATGAGGCTGCCCTTCTGTGGCCACCACGTGGTCAGTCACTTCACC
TGCAAGATCCTGGCAGTGCTGAAGCTGGCATGCGGCAACACGTCGGTCAGCGAAGACTTCCTGCTGGC

GGGCTCCATCCTGCTGCTGCCTGTACCCCTGGCATTCATCTGCCTGTCCTACTTGCTCATCCTGGCCA
CCATCCTGAGGGTGCCCTCGGCCGCCAGGTGCTGCAAAGCCTTCTCCACCTGCTTGGCACACCTGGCT
GTAGTGCTGCTTTTCTACGGCACCATCATCTTCATGTACTTGAAGCCCAAGAGTAAGGAAGCCCACAT
CTCTGATGAGGTCTTCACAGTCCTCTATGCCATGGTCACGACCATGCTGAACCCCACCATCTACAGCC
TGAGGAACAAGGAGGTGAAGGAGGCCGCCAGGAAGGTGTGGGGCAGGAGTCGGGCCTCCAGGTGAGGG
A

NOV34g, SNP13373984 of CG54212-02, Protein Sequence	SEQ ID NO: 438	312 aa	MW at 34688.2kD
	SNP Pos: 133		SNP Change: Arg to His

MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVYFFLGNLST
LDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYHVLM
SHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLLAGSI
LLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLLFYGTIIFMYLKPKSKEAHISDE
VFTVLYAMVTTMLNPTIYSLRNKEVKEAARKVWGRSRASR

NOV34h, SNP13373985 of CG54212-02, DNA Sequence	SEQ ID NO: 439	953 bp
	ORF Start: ATG at 11	ORF Stop: TGA at 947
	SNP Pos: 501	SNP Change: T to C

ACTCTGCAGCATGGAGCCGCTCAACAGAACAGAGGTGTCCGAGTTCTTTCTGAAAGGATTTTCTGGCT
ACCCAGCCCTGGAGCATCTGCTCTTCCCTCTGTGCTCAGCCATGTACCTGGTGACCCTCCTGGGGAAC
ACAGCCATCATGGCGGTGAGCGTGCTAGATATCCACCTGCACACGCCTGTGTACTTCTTCCTGGGCAA
CCTCTCTACCCTGGACATCTGCTACACGCCCACCTTTGTGCCTCTGATGCTGGTCCACCTCCTGTCAT
CCCGGAAGACCATCTCCTTTGCTGTCTGTGCCATCCAGATGTGTCTGAGCCTGTCCACGGGCTCCACG
GAGTGCCTGCTACTGGCCATCACGGCCTATGACCGCTACCTGGCCATCTGCCAGCCACTCAGGTACCG
CGTGCTCATGAGCCACCGGCTCTGCGTGCTGCTGATGGGAGCTGCCTGGGTCCTCTGCCTCCTCAAGT
CGGTGACTGAGATGGTCATCTCCACGAGGCTGCCCTTCTGTGGCCACCACGTGGTCAGTCACTTCACC
TGCAAGATCCTGGCAGTGCTGAAGCTGGCATGCGGCAACACGTCGGTCAGCGAAGACTTCCTGCTGGC
GGGCTCCATCCTGCTGCTGCCTGTACCCCTGGCATTCATCTGCCTGTCCTACTTGCTCATCCTGGCCA
CCATCCTGAGGGTGCCCTCGGCCGCCAGGTGCTGCAAAGCCTTCTCCACCTGCTTGGCACACCTGGCT
GTAGTGCTGCTTTTCTACGGCACCATCATCTTCATGTACTTGAAGCCCAAGAGTAAGGAAGCCCACAT
CTCTGATGAGGTCTTCACAGTCCTCTATGCCATGGTCACGACCATGCTGAACCCCACCATCTACAGCC
TGAGGAACAAGGAGGTGAAGGAGGCCGCCAGGAAGGTGTGGGGCAGGAGTCGGGCCTCCAGGTGAGGG
A

NOV34h, SNP13373985 of CG54212-02, Protein Sequence	SEQ ID NO: 440	312 aa	MW at 34677.1kD
	SNP Pos: 164		SNP Change: Met to Thr

MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVYFFLGNLST
LDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAYDRYLAICQPLRYRVLM
SHRLCVLLMGAAWVLCLLKSVTEMVISTRLPFCGHHVVSHFTCKILAVLKLACGNTSVSEDFLLAGSI
LLLPVPLAFICLSYLLILATILRVPSAARCCKAFSTCLAHLAVVLLFYGTIIFMYLKPKSKEAHISDE
VFTVLYAMVTTMLNPTIYSLRNKEVKEAARKVWGRSRASR

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 34B.

Table 34B. Comparison of the NOV34 protein sequences.

```
NOV34a  MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVY
NOV34b  MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVY
NOV34c  -------------------------LFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVY
NOV34d  MEPLNRTEVSEFFLKGFSGYPALEHLLFPLCSAMYLVTLLGNTAIMAVSVLDIHLHTPVY

NOV34a  FFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAY
NOV34b  FFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAY
NOV34c  FFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAY
```

```
NOV34d  FFLGNLSTLDICYTPTFVPLMLVHLLSSRKTISFAVCAIQMCLSLSTGSTECLLLAITAY

NOV34a  DRYLAICQPLRYRVLMSHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCK
NOV34b  DRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCK
NOV34c  DRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCK
NOV34d  DRYLAICQPLRYHVLMSHRLCVLLMGAAWVLCLLKSVTEMVISMRLPFCGHHVVSHFTCK

NOV34a  ILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFST
NOV34b  ILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFST
NOV34c  ILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFST
NOV34d  ILAVLKLACGNTSVSEDFLLAGSILLLPVPLAFICLSYLLILATILRVPSAARCCKAFST

NOV34a  CLAHLAVVLLFYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEA
NOV34b  CLAHLAVVLLFYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEA
NOV34c  CLAHLAVVLLFYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEA
NOV34d  CLAHLAVVLLFYGTIIFMYLKPKSKEAHISDEVFTVLYAMVTTMLNPTIYSLRNKEVKEA

NOV34a  ARKVWGRSRASR
NOV34b  ARKVWGRSRASR
NOV34c  ARKVWGRSRASR
NOV34d  ARKVWGRSRASR

NOV34a  (SEQ ID NO:  426)
NOV34b  (SEQ ID NO:  428)
NOV34c  (SEQ ID NO:  430)
NOV34d  (SEQ ID NO:  432)
```

Further analysis of the NOV34a protein yielded the following properties shown in Table 34C.

Table 34C. Protein Sequence Properties NOV34a

SignalP analysis:	Cleavage site between residues 48 and 49

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 1;  neg.chg 3
      H-region:  length 3;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -2.43
      possible cleavage site: between 33 and 34

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood = -1.86   Transmembrane   35 -  51
      INTEGRAL    Likelihood = -0.11   Transmembrane   92 - 108
      INTEGRAL    Likelihood = -4.88   Transmembrane  141 - 157
      INTEGRAL    Likelihood = -0.27   Transmembrane  173 - 189
```

```
      INTEGRAL    Likelihood = -7.80   Transmembrane  204 - 220
      INTEGRAL    Likelihood = -6.26   Transmembrane  241 - 257
      PERIPHERAL  Likelihood =  0.79 (at 71)
      ALOM score:  -7.80  (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 42
      Charge difference:  0.5   C( 0.0) - N(-0.5)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:        1       Hyd Moment(75):  6.33
      Hyd Moment(95):  9.16     G content:       0
      D/E content:     2        S/T content:     1
      Score: -5.44

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   8.3%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none
```

```
checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       11.1 %: Golgi
       11.1 %: vacuolar
       11.1 %: vesicles of secretory system
       11.1 %: mitochondrial

>> prediction for CG54212-02 is end (k=9)
```

A search of the NOV34a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 34D.

Table 34D. Geneseq Results for NOV34a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV34a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU85168	G-coupled olfactory receptor #29 - Homo sapiens, 312 aa. [WO200198526-A2, 27-DEC-2001]	1..312 1..312	311/312 (99%) 311/312 (99%)	e-178
AAU95519	Human olfactory and pheromone G protein-coupled receptor #6 - Homo sapiens, 312 aa. [WO200224726-A2, 28-MAR-2002]	1..312 1..312	311/312 (99%) 311/312 (99%)	e-178
AAU97925	Novel odourant receptor NOV4 protein - Unidentified, 312 aa. [WO200236632-A2, 10-MAY-2002]	1..312 1..312	311/312 (99%) 311/312 (99%)	e-178
AAU97923	Novel odourant receptor NOV2 protein - Unidentified, 312 aa. [WO200236632-A2, 10-MAY-2002]	1..312 1..312	311/312 (99%) 311/312 (99%)	e-178
AAU97922	Novel odourant receptor NOV1 protein - Unidentified, 312 aa. [WO200236632-A2, 10-MAY-2002]	1..312 1..312	311/312 (99%) 311/312 (99%)	e-178

In a BLAST search of public sequence databases, the NOV34a protein was found to have homology to the proteins shown in the BLASTP data in Table 34E.

Table 34E. Public BLASTP Results for NOV34a				
Protein Accession Number	**Protein/Organism/Length**	**NOV34a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8NGT2	Seven transmembrane helix receptor - Homo sapiens (Human), 312 aa.	1..312 1..312	311/312 (99%) 311/312 (99%)	e-178
CAD57893	Sequence 9 from Patent WO0236632 - Homo sapiens (Human), 312 aa.	1..312 1..312	310/312 (99%) 311/312 (99%)	e-177
Q9QZ18	Olfactory receptor (Olfactory receptor MOR262-4) - Mus musculus (Mouse), 312 aa.	1..312 1..312	256/312 (82%) 275/312 (88%)	e-146
Q96R40	Olfactory receptor - Homo sapiens (Human), 216 aa (fragment).	68..283 1..216	215/216 (99%) 215/216 (99%)	e-120
CAD42439	Sequence 121 from Patent WO0212343 - Homo sapiens (Human), 311 aa.	1..307 1..305	167/307 (54%) 225/307 (72%)	5e-94

PFam analysis indicates that the NOV34a protein contains the domains shown in the Table 34F.

Table 34F. Domain Analysis of NOV34a			
Pfam Domain	**NOV34a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	41..290	63/276 (23%) 182/276 (66%)	2.8e-34

Example 35.

The NOV35 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 35A.

Table 35A. NOV35 Sequence Analysis

NOV35a, CG54236-01 DNA Sequence	SEQ ID NO: 441	1260 bp
	ORF Start: ATG at 105	ORF Stop: TAA at 1143

TGCTCCCTGTTTCATTAAAACCTAGAGAGATGTAATCAGTAAGCAAGAAGGAAAAAGGGAAATTCACA
AAGTAACTTTTTGTGTCTGTTTCTTTTTAACCCAGC**ATG**GAGAGAAAATTTATGTCCTTGCAACCATC
CATCTCCGTATCAGAAATGGAACCAAATGGCACCTTCAGCAATAACAACAGCAGGAACTGCACAATTG
AAAACTTCAAGAGAGAATTTTTCCCAATTGTATATCTGATAATATTTTTCTGGGGAGTCTTGGGAAAT
GGGTTGTCCATATATGTTTTCCTGCAGCCTTATAAGAAGTCCACATCTGTGAACGTTTTCATGCTAAA
TCTGGCCATTTCAGATCTCCTGTTCATAAGCACGCTTCCCTTCAGGGCTGACTATTATCTTAGAGGCT
CCAATTGGATATTTGGAGACCTGGCCTGCAGGATTATGTCTTATTCCTTGTATGTCAACATGTACAGC
AGTATTTATTTCCTGACCGTGCTGAGTGTTGTGCGTTTCCTGGCAATGGTTCACCCCTTTCGGCTTCT
GCATGTCACCAGCATCAGGAGTGCCTGGATCCTCTGTGGGATCATATGGATCCTTATCATGGCTTCCT
CAATAATGCTCCTGGACAGTGGCTCTGAGCAGAACGGCAGTGTCACATCATGCTTAGAGCTGAATCTC
TATAAAATTGCTAAGCTGCAGACCATGAACTATATTGCCTTGGTGGTGGGCTGCCTGCTGCCATTTTT
CACACTCAGCATCTGTTATCTGCTGATCATTCGGGTTCTGTTAAAAGTGGAGGTCCCAGAATCGGGGC
TGCGGGTTTCTCACAGGAAGGCACTGACCACCATCATCATCACCTTGATCATCTTCTTCTTGTGTTTC
CTGCCCTATCACACACTGAGGACCGTCCACTTGACGACATGGAAAGTGGGTTTATGCAAAGACAGACT
GCATAAAGCTTTGGTTATCACACTGGCCTTGGCAGCAGCCAATGCCTGCTTCAATCCTCTGCTCTATT
ACTTTGCTGGGGAGAATTTTAAGGACAGACTAAAGTCTGCACTCAGAAAAGGCCATCCACAGAAGGCA
AAGACAAAGTGTGTTTTCCCTGTTAGTGTGTGGTTGAGAAAGGAAACAAGAGTA**TAA**GGAGCTCTTAG
ATGAGACCTGTTCTTGTATCCTTGTGTCCATCTTCATTCACTCATAGTCTCCAAATGACTTTGTATTT
ACATCACTCCCAACAAATGTTGATTCTTAATATTTA

NOV35a, CG54236-01 Protein Sequence	SEQ ID NO: 442	346 aa	MW at 39634.7kD

MERKFMSLQPSISVSEMEPNGTFSNNNSRNCTIENFKREFFPIVYLIIFFWGVLGNGLSIYVFLQPYK
KSTSVNVFMLNLAISDLLFISTLPFRADYYLRGSNWIFGDLACRIMSYSLYVNMYSSIYFLTVLSVVR
FLAMVHPFRLLHVTSIRSAWILCGIIWILIMASSIMLLDSGSEQNGSVTSCLELNLYKIAKLQTMNYI
ALVVGCLLPFFTLSICYLLIIRVLLKVEVPESGLRVSHRKALTTIIITLIIFFLCFLPYHTLRTVHLT
TWKVGLCKDRLHKALVITLALAAANACFNPLLYYFAGENFKDRLKSALRKGHPQKAKTKCVFPVSVWL
RKETRV

NOV35b, CG54236-02 DNA Sequence	SEQ ID NO: 443		1193 bp
	ORF Start: ATG at 105		ORF Stop: TAA at 1143

TGCTCCCTGTTTCATTAAAACCTAGAGAGATGTAATCAGTAAGCAAGAAGGAAAAAGGGAAATTCACA
AAGTAACTTTTTGTGTCTGTTTCTTTTTAACCCAGC**ATG**GAGAGAAAATTTATGTCCTTGCAACCATC
CATCTCCGTATCAGAAATGGAACCAAATGGCACCTTCAGCAATAACAACAGCAGGAACTGCACAATTG
AAAACTTCAAGAGAGAATTTTTCCCAATTGTATATCTGATAATATTTTTCTGGGGAGTCTTGGGAAAT
GGGTTGTCCATATATGTTTTCCTGCAGCCTTATAAGAAGTCCACATCTGTGAACGTTTTCATGCTAAA
TCTGGCCATTTCAGATCTCCTGTTCATAAGCACGCTTCCCTTCAGGGCTGACTATTATCTTAGAGGCT
CCAATTGGATATTTGGAGACCTGGCCTGCAGGATTATGTCTTATTCCTTGTATGTCAACATGTACAGC
AGTATTTATTTCCTGACCGTGCTGAGTGTTGTGCGTTTCCTGGCAATGGTTCACCCCTTTCGGCTTCT
GCATGTCACCAGCATCAGGAGTGCCTGGATCCTCTGTGGGATCATATGGATCCTTATCATGGCTTCCT
CAATAATGCTCCTGGACAGTGGCTCTGAGCAGAACGGCAGTGTCACATCATGCTTAGAGCTGAATCTC
TATAAAATTGCTAAGCTGCAGACCATGAACTATATTGCCTTGGTGGTGGGCTGCCTGCTGCCATTTTT
CACACTCAGCATCTGTTATCTGCTGATCATTCGGGTTCTGTTAAAAGTGGAGGTCCCAGAATCGGGGC
TGCGGGTTTCTCACAGGAAGGCACTGACCACCATCATCATCACCTTGATCATCTTCTTCTTGTGTTTC
CTGCCCTATCACACACTGAGGACCGTCCACTTGACGACATGGAAAGTGGGTTTATGCAAAGACAGACT
GCATAAAGCTTTGGTTATCACACTGGCCTTGGCAGCAGCCAATGCCTGCTTCAATCCTCTGCTCTATT
ACTTTGCTGGGGAGAATTTTAAGGACAGACTAAAGTCTGCACTCAGAAAAGGCCATCCACAGAAGGCA
AAGACAAAGTGTGTTTTCCCTGTTAGTGTGTGGTTGAGAAAGGAAACAAGAGTA**TAA**GGAGCTCTTAG
ATGAGACCTGTTCTTGTATCCTTGTGTCCATCTTCAT

NOV35b, CG54236-02 Protein Sequence	SEQ ID NO: 444	346 aa	MW at 39634.7kD

MERKFMSLQPSISVSEMEPNGTFSNNNSRNCTIENFKREFFPIVYLIIFFWGVLGNGLSIYVFLQPYK

KSTSVNVFMLNLAISDLLFISTLPFRADYYLRGSNWIFGDLACRIMSYSLYVNMYSSIYFLTVLSVVR
FLAMVHPFRLLHVTSIRSAWILCGIIWILIMASSIMLLDSGSEQNGSVTSCLELNLYKIAKLQTMNYI
ALVVGCLLPFFTLSICYLLIIRVLLKVEVPESGLRVSHRKALTTIIITLIIFFLCFLPYHTLRTVHLT
TWKVGLCKDRLHKALVITLALAAANACFNPLLYYFAGENFKDRLKSALRKGHPQKAKTKCVFPVSVWL
RKETRV

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 35B.

Table 35B. Comparison of the NOV35 protein sequences.

```
NOV35a  MERKFMSLQPSISVSEMEPNGTFSNNNSRNCTIENFKREFFPIVYLIIFFWGVLGNGLSI
NOV35b  MERKFMSLQPSISVSEMEPNGTFSNNNSRNCTIENFKREFFPIVYLIIFFWGVLGNGLSI

NOV35a  YVFLQPYKKSTSVNVFMLNLAISDLLFISTLPFRADYYLRGSNWIFGDLACRIMSYSLYV
NOV35b  YVFLQPYKKSTSVNVFMLNLAISDLLFISTLPFRADYYLRGSNWIFGDLACRIMSYSLYV

NOV35a  NMYSSIYFLTVLSVVRFLAMVHPFRLLHVTSIRSAWILCGIIWILIMASSIMLLDSGSEQ
NOV35b  NMYSSIYFLTVLSVVRFLAMVHPFRLLHVTSIRSAWILCGIIWILIMASSIMLLDSGSEQ

NOV35a  NGSVTSCLELNLYKIAKLQTMNYIALVVGCLLPFFTLSICYLLIIRVLLKVEVPESGLRV
NOV35b  NGSVTSCLELNLYKIAKLQTMNYIALVVGCLLPFFTLSICYLLIIRVLLKVEVPESGLRV

NOV35a  SHRKALTTIIITLIIFFLCFLPYHTLRTVHLTTWKVGLCKDRLHKALVITLALAAANACF
NOV35b  SHRKALTTIIITLIIFFLCFLPYHTLRTVHLTTWKVGLCKDRLHKALVITLALAAANACF

NOV35a  NPLLYYFAGENFKDRLKSALRKGHPQKAKTKCVFPVSVWLRKETRV
NOV35b  NPLLYYFAGENFKDRLKSALRKGHPQKAKTKCVFPVSVWLRKETRV

NOV35a  (SEQ ID NO:  442)
NOV35b  (SEQ ID NO:  444)
```

Further analysis of the NOV35a protein yielded the following properties shown in Table 35C.

Table 35C. Protein Sequence Properties NOV35a

SignalP analysis:	Cleavage site between residues 60 and 61

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 4;  pos.chg 2;  neg.chg 1
      H-region:  length 11;  peak value   5.18
      PSG score:   0.78

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -0.37
      possible cleavage site: between 55 and 56

>>> Seems to have a cleavable signal peptide (1 to 55)

ALOM: Klein et al's method for TM region allocation
```

```
      Init position for calculation: 56
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood = -2.81   Transmembrane   75 -  91
      INTEGRAL    Likelihood = -3.72   Transmembrane  125 - 141
      INTEGRAL    Likelihood = -8.86   Transmembrane  157 - 173
      INTEGRAL    Likelihood = -7.75   Transmembrane  204 - 220
      INTEGRAL    Likelihood =-11.36   Transmembrane  245 - 261
      INTEGRAL    Likelihood = -1.06   Transmembrane  287 - 303
      PERIPHERAL  Likelihood =  2.38 (at 222)
      ALOM score: -11.36  (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 27
      Charge difference:  2.0   C( 1.0) - N(-1.0)
      C > N: C-terminal side will be inside

>>>Caution: Inconsistent mtop result with signal peptide
>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75): 11.72
      Hyd Moment(95):  9.21    G content:      0
      D/E content:     2       S/T content:    4
      Score: -3.79

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 13  ERK|FM

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  10.4%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: ERKF


      KKXX-like motif in the C-terminus: KETR

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none
```

```
NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       11.1 %: Golgi
       11.1 %: vacuolar
       11.1 %: cytoplasmic

>> prediction for CG54236-01 is end (k=9)
```

A search of the NOV35a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 35D.

Table 35D. Genescq Results for NOV35a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV35a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABU11923	Human G-protein coupled receptor HGPRBMY11v1 - Homo sapiens, 346 aa. [WO200286123-A2, 31-OCT-2002]	1..346 1..346	346/346 (100%) 346/346 (100%)	0.0
ABP81707	Human cysteinyl leukotriene CYSLT2 receptor protein SEQ ID NO:589 - Homo sapiens, 346 aa. [WO200261087-A2, 08-AUG-2002]	1..346 1..346	346/346 (100%) 346/346 (100%)	0.0
ABB05229	Human LTD4-like G protein-coupled receptor protein SEQ ID NO:2 - Homo sapiens, 346 aa. [WO200194580-A1, 13-DEC-2001]	1..346 1..346	346/346 (100%) 346/346 (100%)	0.0
AAG77965	Human G-protein coupled receptor PFI-017* - Homo sapiens, 346 aa. [US2001039037-A1, 08-NOV-2001]	1..346 1..346	346/346 (100%) 346/346 (100%)	0.0
AAE17231	Human CysLT2 GPCR (G-protein coupled receptor) - Homo sapiens, 346 aa. [WO200192302-A2, 06-DEC-2001]	1..346 1..346	346/346 (100%) 346/346 (100%)	0.0

In a BLAST search of public sequence databases, the NOV35a protein was found to have homology to the proteins shown in the BLASTP data in Table 35E.

Table 35E. Public BLASTP Results for NOV35a				
Protein Accession Number	**Protein/Organism/Length**	**NOV35a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9NS75	Cysteinyl leukotriene receptor 2 (CysLTR2) (PSEC0146) (HG57) (HPN321) (hGPCR21) - Homo sapiens (Human), 346 aa.	1..346 1..346	346/346 (100%) 346/346 (100%)	0.0
CAC69290	Sequence 1 from Patent WO0159118 - Homo sapiens (Human), 346 aa.	1..346 1..346	344/346 (99%) 345/346 (99%)	0.0
Q95N03	Cysteinyl leukotriene receptor 2 (CysLTR2) - Sus scrofa (Pig), 345 aa.	1..346 1..345	275/347 (79%) 300/347 (86%)	e-158
Q8R528	Cysteinyl leukotriene 2 receptor - Mus musculus (Mouse), 309 aa.	17..324 1..308	226/308 (73%) 256/308 (82%)	e-132
Q920A1	Cysteinyl leukotriene receptor 2 (CysLTR2) - Mus musculus (Mouse), 309 aa.	17..324 1..308	224/308 (72%) 255/308 (82%)	e-131

PFam analysis indicates that the NOV35a protein contains the domains shown in the Table 35F.

Table 35F. Domain Analysis of NOV35a			
Pfam Domain	**NOV35a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	55..305	89/277 (32%) 185/277 (67%)	6.1e-54
TAS2R	33..324	64/314 (20%) 173/314 (55%)	0.019

Example 36.

The NOV36 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 36A.

Table 36A. NOV36 Sequence Analysis		
NOV36a, CG54479-06 DNA Sequence	SEQ ID NO: 445	2066 bp
	ORF Start: ATG at 22	ORF Stop: TAG at 2032

ACAGGTTTCACAACTTCCCGG**ATG**GGGGCTGTGGTGGGTCACAGTGCAGCCTCCAGCCAGAAGGATGGG
GTGGCTCCCACTCCTGCTGCTTCTGACTCAATGCTTAGGGGTCCCTGGGCAGCGCTCGCCATTGAATG
ACTTCCAAGTGCTCCGGGGCACAGAGCTACAGCACCTGCTACATGCGGTGGTGCCCGGGCCTTGGCAG
GAGGATGTGGCAGATGCTGAAGAGTGTGCTGGTCGCTGTGGGCCCTTAATGGACTGCCGGGCCTTCCA
CTACAACGTGAGCAGCCATGGTTGCCAACTGCTGCCATGGACTCAACACTCGCCCCACACGAGGCTGC
GGCGTTCTGGGCGCTGTGACCTCTTCCAGAAGAAAGACTACGTACGGACCTGCATCATGAACAATGGG
GTTGGGTACCGGGGCACCATGGCCACGACCGTGGGTGGCCTGCCCTGCCAGGCTTGGAGCCACAAGTT
CCCGAATGATCACAAGTACACGCCCACTCTCCGGAATGGCCTGGAAGAGAACTTCTGCCGTAACCCTG
ATGGCGACCCCGGAGGTCCTTGGTGCTACACAACAGACCCTGCTGTGCGCTTCCAGAGCTGCGGCATC
AAATCCTGCCGGGAGGCCGCGTGTGTCTGGTGCAATGGCGAGGAATACCGCGGCGCGGTAGACCGCAC
GGAGTCAGGGCGCGAGTGCCAGCGCTGGGATCTTCAGCACCCGCACCAGCACCCCTTCGAGCCGGGCA
AGTTCCTCGACCAAGGTCTGGACGACAACTATTGCCGGAATCCTGACGGCTCCGAGCGGCCATGGTGC
TACACTACGGATCCGCAGATCGAGCGAGAGTTCTGTGACCTCCCCCGCTGCGGGTCCGAGGCACAGCC
CCGCCAAGAGGCCACAACTGTCAGCTGCTTCCGCGGGAAGGGTGAGGGCTACCGGGGCACAGCCAATA
CCACCACTGCGGGCGTACCTTGCCAGCGTTGGGACGCGCAAATCCCTCATCAGCACCGATTTACGCCA
GAAAAATACGCGTGCAAAGACCTTCGGGAGAACTTCTGCCGGAACCCCGACGGCTCAGAGGCGCCCTG
GTGCTTCACACTGCGGCCCGGCATGCGCGCGGCCTTTTGCTACCAGATCCGGCGTTGTACAGACGACG
TGCGGCCCCAGGACTGCTACCACGGCGCAGGGGAGCAGTACCGCGGCACGGTCAGCAAGACCCGCAAG
GGTGTCCAGTGCCAGCGCTGGTCCGCTGAGACGCCGCACAAGCCGCAGTTCACGTTTACCTCCGAACC
GCATGCACAACTGGAGGAGAACTTCTGCCGGAACCCAGATGGGGATAGCCATGGGCCCTGGTGCTACA
CGATGGACCCAAGGACCCCATTCGACTACTGTGCCCTGCGACGCTGCGCTGATGACCAGCCGCCATCA
ATCCTGGACCCCCCAGACCAGGTGCAGTTTGAGAAGTGTGGCAAGAGGGTGGATCGGCTGGATCAGCG
GCGTTCCAAGCTGCGCGTGGTTGGGGGCCATCCGGGCAACTCACCCTGGACAGTCAGCTTGCGGAATC
GGCAGGGCCAGCATTTCTGCGGGGGGTCTCTAGTGAAGGAGCAGTGGATACTGACTGCCCGGCAGTGC
TTCTCCTCCTGCCATATGCCTCTCACGGGCTATGAGGTATGGTTGGGCACCCTGTTCCAGAACCCACA
GCATGGAGAGCCAAGCCTACAGCGGGTCCCAGTAGCCAAGATGGTGTGTGGGCCCTCAGGCTCCCAGC
TTGTCCTGCTCAAGCTGGAGAGATCTGTGACCCTGAACCAGCGTGTGGCCCTGATCTGCCTGCCCCCT
GAATGGTATGTGGTGCCTCCAGGGACCAAGTGTGAGGGTGACTACGGGGGCCCACTTGCCTGCTTTAC
CCACAACTGCTGGGTCCTGGAAGGAATTATAATCCCCAACCGAGTATGCGCAAGGTCCCGCTGGCCAG
CTGTCTTCACGCGTGTCTCTGTGTTTGTGGACTGGATTCACAAGGTCATGAGACTGGGT**TAG**GCCCAG
CCTTGATGCCATATGCCTTGGGGAGG

NOV36a, CG54479-06 Protein Sequence	SEQ ID NO: 446	670 aa	MW at 76160.6kD

MGLWWVTVQPPARRMGWLPLLLLLTQCLGVPGQRSPLNDFQVLRGTELQHLLHAVVPGPWQEDVADAE
ECAGRCGPLMDCRAFHYNVSSHGCQLLPWTQHSPHTRLRRSGRCDLFQKKDYVRTCIMNNGVGYRGTM
ATTVGGLPCQAWSHKFPNDHKYTPTLRNGLEENFCRNPDGDPGGPWCYTTDPAVRFQSCGIKSCREAA
CVWCNGEEYRGAVDRTESGRECQRWDLQHPHQHPFEPGKFLDQGLDDNYCRNPDGSERPWCYTTDPQI
EREFCDLPRCGSEAQPRQEATTVSCFRGKGEGYRGTANTTTAGVPCQRWDAQIPHQHRFTPEKYACKD
LRENFCRNPDGSEAPWCFTLRPGMRAAFCYQIRRCTDDVRPQDCYHGAGEQYRGTVSKTRKGVQCQRW
SAETPHKPQFTFTSEPHAQLEENFCRNPDGDSHGPWCYTMDPRTPFDYCALRRCADDQPPSILDPPDQ
VQFEKCGKRVDRLDQRRSKLRVVGGHPGNSPWTVSLRNRQGQHFCGGSLVKEQWILTARQCFSSCHMP
LTGYEVWLGTLFQNPQHGEPSLQRVPVAKMVCGPSGSQLVLLKLERSVTLNQRVALICLPPEWYVVPP
GTKCEGDYGGPLACFTHNCWVLEGIIIPNRVCARSRWPAVFTRVSVFVDWIHKVMRLG

NOV36b, CG54479-05 DNA Sequence	SEQ ID NO: 447	1698 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1696

ATGACTTCTAGGTGCTCCGGGGCACAGAGCTACCTACAAGCGGTGGTGCCCGGGCCTTGGCAGGAGGA
TGTGGCAGATGCTGAAGAGTGTGCTGGTCGCTGTGGGCCCTTAATGGACTGCGCGTTCCACTACAATG
TGAGCAGCCATGGTTGCCAACTGCTGCCATGGACTCAACACTCACCCCACACGAGGCTGCGGCATTCT
GGGCGCTGTGACCTCTTCCAGGAGAAAGACTACATACGGACCTGCATCATGAACAATGGGGTTGGGTA
CCGGGGCACCATGGCCACGACCGTGGGTGGCCTGTCCTGCCAGGCTTGGAGCCACAAGTTCCCGAACG
ATCACCAGTACATGCCCACGCTCCGGAATGGCCTGGAAGAGAACTTCTGCCGTAACCCTGATGGCGAC

CCCGGAGGTCCTTGGTGCCACACAACAGACCCTGCCGTGCGCTTCCAGAGCTGCGGCATCAAATCCTG
CCGGGTGGCCGCGTGTGTCTGGTGCAATGGCGAGGAATACCGCGGCGCGGTAGACCGCACCGAGTCAG
GGCGCGAGTGCCAGCGCTGGGATCTTCAGCACCCGCACCAGCACCCCTTCGAGCCGGGCAAGTTCCTC
GACCAAGGTCTGGACGACAACTATTGCCGGAATCCTGACGGCTCCGAGCGGCCATGGTGCTACACTAC
GGATCCGCAGATCGAGCGAGAATTCTGTGACCTCCCCCGCTGCGGTTCCGAGGCACAGCCCCGCCAAG
AGGCCACAAGTGTCAGCTGCTTCCGCGGGAAGGGTGAGGGCTACCGGGGCACAGCCAATACCACCACC
GCGGGCGTACCTTGCCAGCGTTGGGACGCGCAAATCCCGCATCAGCACCGATTTACGCCAGAAAAATA
CGCGTGCAAGGACCTTCGGGAGAACTTCTGCCGGAACCCCGACGGCTCAGAGGCGCCCTGGTGCTTCA
CACTGCGGCCCGGCATGCGCGTGGGCTTTTGCTACCAGATCCGGCGTTGTACAGACGACGTGCGGCCC
CAGGACTGCTACCACGGCGCGGGGGAGCAGTACCGCGGCACGGTCAGCAAGACCCGCAAGGGTGTCCA
GTGCCAGCGCGGGTCCGCTGAGACGCCGCACAAGCCGCAGTTCACGTTTACCTCCGAACCGCATGCAC
AACTGGAGGAGAACTTCTGCCAGGACCCAGATGGGGATAGCCATGGGCCCTGGTGCTACACGATGGAC
CCAAGGACCCCATTCGACTACTGTGCCCTGCGACGCTGCGCTGATGACCAGCCGCCATCAATCCTGGA
CCCCCCCGACCAGGTGCAGTTTGAGAAGTGTGGCAAGAGGGTGGATCGGCTGGATCAGCGTTGTTCCA
AGCTGCGCGTGGCTGGGGGCCATCCGGGCAACTCACCCTGGACAGTCAGCTTGCGGAATAGGCAGGGC
CAGCATTTCTGCGGGGGGTCTCTAGTGAAGGAGCAGTGGATACTGACTGCCCGGCAGTGCTTCTCCTC
CAGCCATATGCCTCTCACGGGCTATGAGGTATGGTTGGGCACCCTGTTCCAGAACCCACAACATGGAG
AGCCAGGCCTACAGCGGGTCCCAGTAGCCAAGATGCTGTGTGGGCCCTCAGGCTCTCAGCTTGTCCTG
CTCAAGCTGGAGAGGTCTGTGACCCTGAACCAGCGTGTGGCCCTGATCTGCCTGCCGCCTG**AATGA**

NOV36b, CG54479-05 Protein Sequence	SEQ ID NO: 448	565 aa	MW at 63751.8kD

MTSRCSGAQSYLQAVVPGPWQEDVADAEECAGRCGPLMDCAFHYNVSSHGCQLLPWTQHSPHTRLRHS
GRCDLFQEKDYIRTCIMNNGVGYRGTMATTVGGLSCQAWSHKFPNDHQYMPTLRNGLEENFCRNPDGD
PGGPWCHTTDPAVRFQSCGIKSCRVAACVWCNGEEYRGAVDRTESGRECQRWDLQHPHQHPFEPGKFL
DQGLDDNYCRNPDGSERPWCYTTDPQIEREFCDLPRCGSEAQPRQEATSVSCFRGKGEGYRGTANTTT
AGVPCQRWDAQIPHQHRFTPEKYACKDLRENFCRNPDGSEAPWCFTLRPGMRVGFCYQIRRCTDDVRP
QDCYHGAGEQYRGTVSKTRKGVQCQRGSAETPHKPQFTFTSEPHAQLEENFCQDPDGDSHGPWCYTMD
PRTPFDYCALRRCADDQPPSILDPPDQVQFEKCGKRVDRLDQRCSKLRVAGGHPGNSPWTVSLRNRQG
QHFCGGSLVKEQWILTARQCFSSSHMPLTGYEVWLGTLFQNPQHGEPGLQRVPVAKMLCGPSGSQLVL
LKLERSVTLNQRVALICLPPE

NOV36c, CG54479-01 DNA Sequence	SEQ ID NO: 449	2200 bp
	ORF Start: ATG at 21	ORF Stop: TAG at 2157

TGCAGCCTCCAGCCAGAAGG**ATG**GGGTGGCTCCCACTCCTGCTGCTTCTGACTCAATGCTTAGGGGTC
CCTGGGCAGCGCTCGCCATTGAATGACTTCGAGGTGCTCCGGGGCACAGAGCTACAGCGGCTGCTACA
AGCGGTGGTGCCCGGGCCTTGGCAGGAGGATGTGGCAGATGCTGAAGAGTGTGCTGGTCGCTGTGGGC
CCTTAATGGACTGCCGGGCGTTCCACTACAATGTGAGCAGCCATGGTTGCCAACTGCTGCCATGGACT
CAACACTCACCCCACACGAGGCTGCGGCATTCTGGGCGCTGTGACCTCTTCCAGGAGAAAGACTACAT
ACGGACCTGCATCATGAACAATGGGGTTGGGTACCGGGGCACCATGGCCACGACCGTGGGTGGCCTGT
CCTGCCAGGCTTGGAGCCACAAGTTCCCGAACGATCACAGGTACATGCCCACGCTCCGGAATGGCCTG
GAAGAGAACTTCTGCCGTAACCCTGATGGCGACCCCGGAGGTCCTTGGTGCCACACAACAGACCCTGC
CGTGCGCTTCCAGAGCTGCGGCATCAAATCCTGCCGGTCTGCCGCGTGTGTCTGGTGCAATGGCGAGG
AATACCGCGGCGCGGTAGACCGCACCGAGTCAGGGCGCGAGTGCCAGCGCTGGGATCTTCAGCACCCG
CACCAGCACCCCTTCGAGCCGGGCAAGTACCCCGACCAAGGTCTGGACGACAACTATTGCCGGAATCC
TGACGGCTCCGAGCGGCCATGGTGCTACACTACGGATCCGCAGATCGAGCGAGAATTCTGTGACCTCC
CCCGCTGCGGTTCCGAGGCACAGCCCCGCCAAGAGGCCACAAGTGTCAGCTGCTTCCGCGGGAAGGGT
GAGGGCTACCGGGGCACAGCCAATACCACCACCGCGGGCGTACCTTGCCAGCGTTGGGACGCGCAAAT
CCCGCATCAGCACCGATTTACGCCAGAAAAATACGCGTGCAAGGACCTTCGGGAGAACTTCTGCCGGA
ACCCCGACGGCTCAGAGGCGCCCTGGTGCTTCACACTGCGGCCCGGCATGCGCGTGGGCTTTTGCTAC
CAGATCCGGCGTTGTACAGACGACGTGCGGCCCCAGGGTTGCTACCACGGCGCGGGGGAGCAGTACCG
CGGCACGGTCAGCAAGACCCGCAAGGGTGTCCAGTGCCAGCGCGCGTCCGCTGAGACGCCGCACAAGC
CGCAGTTTACCTTTACCTCCGAACCGCATGCACAACTGGAGGAGAACTTCTGCCGCGACCCAGATGGG
GATAGCTATGGGCCCTGGTGCTACACGATGGACCCAAGGACCCCATTCGACTACTGTGCCCTGCGACG
CTGCGCTGATGACCAGCCGCCATCAATCCTGGACCCCCCCGACCAGGTGCAGTTTGAGAAGTGTGGCA
AGAGGGTGGATCGGCTGGATCAGCGTTGTTCCAAGCTGCGCGTGGCTGGGGGCCATCCGGGCAACTCA
CCCTGGACAGTCAGCTTGCGGAATAGGCAGGGCCAGCATTTCTGCGGGGGGTCTCTAGTGAAGGAGCA
GTGGATACTGACTGCCCGGCAGTGCTTCTCCTCCAGCCATATGCCTCTCACGGGCTATGAGGTATGGT
TGGGCACCCTGTTCCAGAACCCACAACATGGAGAGCCAGGCCTACAGCGGGTCCCAGTAGCCAAGATG

CTGTGTGGGCCCTCAGGCTCTCAGCTTGTCCTGCTCAAGCTGGAGAGATCTGTGACCCTGAACCAGCG
TGTGGCCCTGATCTGCCTGCCGCCTGAATGGTATGTGGTGCCTCCAGGGACCAAGTGTGAGATTGCAG
GCCGGGGTGAGACCAAAGGTACGGGTAATGACACAGTCCTAAATGTGGCCTTGCTGAATGTCATCTCC
AACCAGGAGTGTAACATCAAGCACCGAGGACATGTGCGGGAGAGCGAGATGTGCACTGAGGGACTGTT
GGCCCCTGTGGGGGCCTGTGAGGGGGGTGACTACGGGGGCCCACTTGCCTGCTTTACCCACAACTGCT
GGGTCCTGGAAGGAATTAGAATCCCCAACCGAGTATGCGCAAGGTCGCGCTGGCCAGCCGTCTTCACA
CGTGTCTCTGTGTTTGTGGACTGGATTCACAAGGTCATGAGACTGGGTTAGGCCCAGCCTTGACGCCA
TATGCTTTGGGGAGGACAAAACTT

NOV36c, CG54479-01 Protein Sequence	SEQ ID NO: 450	712 aa	MW at 80097.8kD

MGWLPLLLLLTQCLGVPGQRSPLNDFEVLRGTELQRLLQAVVPGPWQEDVADAEECAGRCGPLMDCRA
FHYNVSSHGCQLLPWTQHSPHTRLRHSGRCDLFQEKDYIRTCIMNNGVGYRGTMATTVGGLSCQAWSH
KFPNDHRYMPTLRNGLEENFCRNPDGDPGGPWCHTTDPAVRFQSCGIKSCRSAACVWCNGEEYRGAVD
RTESGRECQRWDLQHPHQHPFEPGKYPDQGLDDNYCRNPDGSERPWCYTTDPQIEREFCDLPRCGSEA
QPRQEATSVSCFRGKGEGYRGTANTTTAGVPCQRWDAQIPHQHRFTPEKYACKDLRENFCRNPDGSEA
PWCFTLRPGMRVGFCYQIRRCTDDVRPQGCYHGAGEQYRGTVSKTRKGVQCQRASAETPHKPQFTFTS
EPHAQLEENFCRDPDGDSYGPWCYTMDPRTPFDYCALRRCADDQPPSILDPPDQVQFEKCGKRVDRLD
QRCSKLRVAGGHPGNSPWTVSLRNRQGQHFCGGSLVKEQWILTARQCFSSSHMPLTGYEVWLGTLFQN
PQHGEPGLQRVPVAKMLCGPSGSQLVLLKLERSVTLNQRVALICLPPEWYVVPPGTKCEIAGRGETKG
TGNDTVLNVALLNVISNQECNIKHRGHVRESEMCTEGLLAPVGACEGGDYGGPLACFTHNCWVLEGIR
IPNRVCARSRWPAVFTRVSVFVDWIHKVMRLG

NOV36d, CG54479-02 DNA Sequence	SEQ ID NO: 451	1710 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1705

ATGACTTCCAGGTGCTCCGGGGCACAGAGCTACCTGCTACATGCGGTGGTGCCTGGGCCTTGGCAGGA
GGATGTGGCAGATGCTGAAGAGTGTGCTGGTCGCTGTGGGCCCTTAACGGACTGCTGGGCCTTCCACT
ACAATGTGAGCAGCCATGGTTGCCAACTGCTGCCATGGACTCAACACTCGCCCCACTCAAGGCTGTGG
CATTCTGGGCGCTGTGACCTCTTCCAGAAGAAAGACTACATACGGACCTGCATCATGAACAATGGGGT
TGGGTACCGGGGCACCATGGCCACGACCGTGGGTGGCCTGTCCTGCCAGGCTTGGAGCCACAAGTTCC
CGAATGATCACAAGTACATGCCCACGCTCCGGAATGGCCTGGAAGAGAACTTCTGCCATAACCCTGAT
GGCGACCCCGGAGGTCCTTGGTGCCACACAACAGACCCTGCCGTGCGCTTCCAGAGCTGCGGCATCAA
ATCCTGCCGGGTGGCCGCGTGTGTCTGGTGCAATGGCGAGGAATACCGCGGCGCGGTAGACCGCACCG
AGTCAGGGCGCGAGTGCCAGCGCTGGGATCTTCAGCACCCGCACCAGCACCCCTTCGAGCCGGGCAAG
TACCTCGACCAAGGTCTGGACGACAACTATTGCCGGAATCCTGACGGCTCCGAGCGGCCATGGTGCTA
CACTACGGATCCGCAGATCGAGCGAGAATTCTGTGACCTCCCCCGCTGCGGTTCCGAGGCACAGCCCC
GCCAAGAGGCCACAAGTGTCAGCTGCTTCCGCGGGAAGGGTGAGGGCTACCGGGGCACAGCCAATACC
ACCACCGCGGGCGTACCTTGCCAGCGTTGGGACGCGCAAATCCCGCATCAGCACCGATTTACGCCAGA
AAAATACGCGTGCAAGGACCTTCGGGAGAACTTCTGCCGGAACCCCGACGGCTCAGAGGCGCCCTGGT
GCTTCACACTGCGGCCCGGCATGCGCGTGGGCTTTTGCTACCAGATCCGGCGTTGTACAGACGACGTG
CGGCCCCAGGACTGCTACCACGGCGCGGGGGAGCAGTACCGCGGCACGGTCAGCAAGACCCGCAAGGG
TGTCCAGTGCCAGCGCGCGTCCGCTGAGACGCCGCACAAGCCGCAGTTCACGTTTACCTCCGAACCGC
ATGCACAACTGGAGGAGAACTTCTGCCAGGACCCAGATGGGGATAGCCATGGGCCCTGGTGCTACACG
ATGGACCCAAGGACCCCATTCGACTACTGTGCCCTGCGACGCTGCGCTGATGACCAGCCGCCATCAAT
CCTGGACCCCCCCACAGACCAGGTGCAGTTTGAGAAGTGTGGCAAGAGGGTGGATCGGCTGGATCAGC
GTCGTTCCAAGCTGCGCGTGGCTGGGGGCCATCCGGGCAACTCACCCTGGACAGTCAGCTTGGGGAAT
CGGAGGCAGGGCCAGCATTTCTGCGGGGGGTCTCTAGTGAAGGAGCAGTGGATACTGACTGCCCGGCA
GTGCTTCTCCTCCCATATGCCTCTCACGGGCTATGAGGTATGGTTGGGCACCCTGTTCCAGAACCCAC
AACATGGAGAGCCAGGCCTACAGCGGGTCCCAGTAGCCAAGATGCTGTGTGGGCCCTCAGGCTCCCAG
CTTGTCCTGCTCAAGCTGGAGAGATCTGTGACCCTGAACCAGCGTGTGGCCCTGATCTGCCTGCCGCC
TGAATGATAT

NOV36d, CG54479-02 Protein Sequence	SEQ ID NO: 452	568 aa	MW at 64180.3kD

MTSRCSGAQSYLLHAVVPGPWQEDVADAEECAGRCGPLTDCWAFHYNVSSHGCQLLPWTQHSPHSRLW
HSGRCDLFQKKDYIRTCIMNNGVGYRGTMATTVGGLSCQAWSHKFPNDHKYMPTLRNGLEENFCHNPD
GDPGGPWCHTTDPAVRFQSCGIKSCRVAACVWCNGEEYRGAVDRTESGRECQRWDLQHPHQHPFEPGK
YLDQGLDDNYCRNPDGSERPWCYTTDPQIEREFCDLPRCGSEAQPRQEATSVSCFRGKGEGYRGTANT
TTAGVPCQRWDAQIPHQHRFTPEKYACKDLRENFCRNPDGSEAPWCFTLRPGMRVGFCYQIRRCTDDV

RPQDCYHGAGEQYRGTVSKTRKGVQCQRASAETPHKPQFTFTSEPHAQLEENFCQDPDGDSHGPWCYT
MDPRTPFDYCALRRCADDQPPSILDPPTDQVQFEKCGKRVDRLDQRRSKLRVAGGHPGNSPWTVSLGN
RRQGQHFCGGSLVKEQWILTARQCFSSHMPLTGYEVWLGTLFQNPQHGEPGLQRVPVAKMLCGPSGSQ
LVLLKLERSVTLNQRVALICLPPE

NOV36e, CG54479-03 DNA Sequence	SEQ ID NO: 453	1011 bp
	ORF Start: at 7	ORF Stop: at 1006

AAGCTTTGCATCATGAACAATGGGGTTGGGTACCGGGGCACCATGGCCACGACCGTGGGTGGCCTGCC
CTGCCAGGCTTGGAGCCACAAGTTCCCAAATGATCACAAGTACACGCCCACTCTCCGGAATGGCCTGG
AAGAGAACTTCTGCCGTAACCCTGATGGCGACCCCGGAGGTCCTTGGTGCTACACAACAGACCCTGCT
GTGCGCTTCCAGAGCTGCGGCATCGAATCCTGCCGGGAGGCCGCGTGTGTCTGGTGCAATGGCGAGGA
ATACCGCGGCGCGGTAGACCGCACGGAGTCAGGGCGCGAGTGCCAGCGCTGGGATCTTCAGCACCCGC
ACCAGCACCCCTTCGAGCCGGGCAAGTTCCTCGACCAAGGTCTGGACGACAACTATTGCCGGAATCCT
GACGGCTCCGAGCGGCCATGGTGCTACACTACGGATCCGCAGATCGAGCGAGAGTTCTGTGACCTCCC
CCGCTGCGGGTCCGAGGCACAGCCCCGCCAAGAGGCCACAACTGTCAGCTGCTTCCGCGGGAAGGGTG
AGGGCTACCGGGGCACAGCCAATACCACCACTGCGGGCGTACCTTGCCAGCGTTGGGACGCGCAAATC
CCTCATCAGCACCGATTTACGCCAGAAAAATACGCGTGCAAAGACCTTCGGGAGAACTTCTGCCGGAA
CCCCGACGGCTCAGAGGCGCCCTGGTGCTTCACACTGCGGCCCGGCATGCGCGCGGCCTTTTGCTACC
AGATCCGGCGTTGTACAGACGACGTGCGGCCCCAGGGGGAGCAGTACCGCGGCACGGTCAGCAAGACC
CGCAAGGGTGTCCAGTGCCAGCGCTGGTCCGCTGAGACGCCGCACAAGCCGCAGTTCACGTTTACCTC
CGAACCGCATGCACAACTGGAGGAGAACTTCTGCCGGAACCCAGATGGGGATAGCCATGGGCCCTGGT
GCTACACGATGGACCCAAGGACCCCATTCGACTACTGTGCCCTGCGACGCTGCCTCGAG

NOV36e, CG54479-03 Protein Sequence	SEQ ID NO: 454	333 aa	MW at 38129.9kD

CIMNNGVGYRGTMATTVGGLPCQAWSHKFPNDHKYTPTLRNGLEENFCRNPDGDPGGPWCYTTDPAVR
FQSCGIESCREAACVWCNGEEYRGAVDRTESGRECQRWDLQHPHQHPFEPGKFLDQGLDDNYCRNPDG
SERPWCYTTDPQIEREFCDLPRCGSEAQPRQEATTVSCFRGKGEGYRGTANTTTAGVPCQRWDAQIPH
QHRFTPEKYACKDLRENFCRNPDGSEAPWCFTLRPGMRAAFCYQIRRCTDDVRPQGEQYRGTVSKTRK
GVQCQRWSAETPHKPQFTFTSEPHAQLEENFCRNPDGDSHGPWCYTMDPRTPFDYCALRRC

NOV36f, CG54479-04 DNA Sequence	SEQ ID NO: 455	1881 bp
	ORF Start: ATG at 76	ORF Stop: TGA at 1777

ACACATTACTGACATGTATGCCCACCTGACCTGCACCCACTCATGCCCACTCTGCAGGGCAGCGCTCG
CCATTGAATGACTTCCAGGTGCTCCGGGGCACAGAGCTACCTGCTACATGCGGTGGTGCCTGGGCCTT
GGCAGGAGGATGTGGCAGATGCTGAAGAGTGTGCTGGTCGCTGTGGGCCCTTAACGGACTGCTGGGCC
TTCCACTACAATGTGAGCAGCCATGGTTGCCAACTGCTGCCATGGACTCAACACTCGCCCCACTCAAG
GCTGTGGCATTCTGGGCGCTGTGACCTCTTCCAGAAGAAAGACTACATACGGACCTGCATCATGAACA
ATGGGGTTGGGTACCGGGGCACCATGGCCACGACCGTGGGTGGCCTGTCCTGCCAGGCTTGGAGCCAC
AAGTTCCCGAATGATCACAAGTACATGCCCACGCTCCGGAATGGCCTGGAAGAGAACTTCTGCCATAA
CCCTGATGGCGACCCCGGAGGTCCTTGGTGCCACACAACAGACCCTGCCGTGCGCTTCCAGAGCTGCG
GCATCAAATCCTGCCGGGTGGCCGCGTGTGTCTGGTGCAATGGCGAGGAATACCGCGGCGCGGTAGAC
CGCACCGAGTCAGGGCGCGAGTGCCAGCGCTGGGATCTTCAGCACCCGCACCAGCACCCCTTCGAGCC
GGGCAGGTTCCTCGACCAAGGTCTGGACGACAACTATTGCCGGAATCCTGACGGCTCCGAGCGGCCAT
GGTGCTACACTACGGATCCGCAGATCGAGCGAGAATTCTGTGACCTCCCCCGCTGCGGTTCCGAGGCA
CAGCCCCGCCAAGAGGCCACAAGTGTCAGCTGCTTCCGCGGGAAGGGTGAGGGCTACCGGGGCACAGC
CAATACCACCACCGCGGGCGTACCTTGCCAGCGTTGGGACGCGCAAATCCCGCATCAGCACCGATTTA
CGCCAGAAAAATACGCGTGCAAGGACCTTCGGGAGAACTTCTGCCGGAACCTCGACGGCTCAGAGGCG
CCCTGGTGCTTCACACTGCGGCCCGGCATGCGCGTGGGCTTTTGCTACCAGATCCGGCGTTGTACAGA
CGACGTGCGGCCCCAGGACTGCTACCACGGCGCGGGGGAGCAGTACCGCGGCACGGTCAGCAAGACCC
GCAAGGGTGTCCAGTGCCAGCGCGCGTCCGCTGAGACGCCGCACAAGCCGCAGTTCACGTTTACCTCC
GAACCGCATGCACAACTGGAGGAGAACTTCTGCCAGACCCCAGATGGGGATAGCCATGGGCCCTGGTG
CTACACGATGGACCCAAGGACCCCATTCGACTACTGTGCCCTGCGACGCTGCGCTGATGACCAGCCGC
CATCAATCCTGGACCCCCCCGACCAGGTGCAGTTTGAGAAGTGTGGCAAGAGGGTGGATCGGCTGGAT
CAGCGTCGTTCCAAGCTGCGCGTGGCTGGGGGCCATCCGGGCAACTCACCCTGGACAGTCAGCTTGGG
GAATCGGCAGGGCCAGCATTTCTGCGGGGGGTCTCTAGTGAAGGAGCAGTGGATACTGACTGCCCGGC
AGTGCTTCTCCTCCCAGCATATGCCTCTCACGGGCTATGAGGTATGGTTGGGCACCCTGTTCCAGAAC
CCACAACATGGAGAGCCAGGCCTACAGCGGGTCCCAGTAGCCAAGATGCTGTGTGGGCCCTCAGGCTC

CCAGCTTGTCCTGCTCAAGCTGGAGAGGTCTGTGACCCTGAACCAGCGTGTGGCCCTGATCTGCCTGC
CGCCTGAATGATATGTGGTGCCTCCAGGGACCAAGTGTGAGATTGCAGGCCGGGGTGAGACCAAAGGT
AAGAGCATAGTGCACAGGACTGCTGGTGGCCAGGAGGCCCAGCCC

NOV36f, CG54479-04 Protein Sequence	SEQ ID NO: 456	567 aa	MW at 64065.2kD

MTSRCSGAQSYLLHAVVPGPWQEDVADAEECAGRCGPLTDCWAFHYNVSSHGCQLLPWTQHSPHSRLW
HSGRCDLFQKKDYIRTCIMNNGVGYRGTMATTVGGLSCQAWSHKFPNDHKYMPTLRNGLEENFCHNPD
GDPGGPWCHTTDPAVRFQSCGIKSCRVAACVWCNGEEYRGAVDRTESGRECQRWDLQHPHQHPFEPGR
FLDQGLDDNYCRNPDGSERPWCYTTDPQIEREFCDLPRCGSEAQPRQEATSVSCFRGKGEGYRGTANT
TTAGVPCQRWDAQIPHQHRFTPEKYACKDLRENFCRNLDGSEAPWCFTLRPGMRVGFCYQIRRCTDDV
RPQDCYHGAGEQYRGTVSKTRKGVQCQRASAETPHKPQFTFTSEPHAQLEENFCQTPDGDSHGPWCYT
MDPRTPFDYCALRRCADDQPPSILDPPDQVQFEKCGKRVDRLDQRRSKLRVAGGHPGNSPWTVSLGNR
QGQHFCGGSLVKEQWILTARQCFSSQHMPLTGYEVWLGTLFQNPQHGEPGLQRVPVAKMLCGPSGSQL
VLLKLERSVTLNQRVALICLPPE

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 36B.

Table 36B. Comparison of the NOV36 protein sequences.

```
NOV36a  MGLWWVTVQPPARRMGWLPLLLLLTQCLGVPGQRSPLNDFQVLRGTELQHLLHAVVPGPW
NOV36b  ----------------------MTSRCSGAQSY-----------------LQAVVPGPW
NOV36c  --------------MGWLPLLLLLTQCLGVPGQRSPLNDFEVLRGTELQRLLQAVVPGPW
NOV36d  ----------------------MTSRCSGAQSYL----------------LHAVVPGPW
NOV36e  ------------------------------------------------------------
NOV36f  ----------------------MTSRCSGAQSYL----------------LHAVVPGPW

NOV36a  QEDVADAEECAGRCGPLMDCRAFHYNVSSHGCQLLPWTQHSPHTRLRRSGRCDLFQKKDY
NOV36b  QEDVADAEECAGRCGPLMDC-AFHYNVSSHGCQLLPWTQHSPHTRLRHSGRCDLFQEKDY
NOV36c  QEDVADAEECAGRCGPLMDCRAFHYNVSSHGCQLLPWTQHSPHTRLRHSGRCDLFQEKDY
NOV36d  QEDVADAEECAGRCGPLTDCWAFHYNVSSHGCQLLPWTQHSPHSRLWHSGRCDLFQKKDY
NOV36e  ------------------------------------------------------------
NOV36f  QEDVADAEECAGRCGPLTDCWAFHYNVSSHGCQLLPWTQHSPHSRLWHSGRCDLFQKKDY

NOV36a  VRTCIMNNGVGYRGTMATTVGGLPCQAWSHKFPNDHKYTPTLRNGLEENFCRNPDGDPGG
NOV36b  IRTCIMNNGVGYRGTMATTVGGLSCQAWSHKFPNDHQYMPTLRNGLEENFCRNPDGDPGG
NOV36c  IRTCIMNNGVGYRGTMATTVGGLSCQAWSHKFPNDHRYMPTLRNGLEENFCRNPDGDPGG
NOV36d  IRTCIMNNGVGYRGTMATTVGGLSCQAWSHKFPNDHKYMPTLRNGLEENFCHNPDGDPGG
NOV36e  ---CIMNNGVGYRGTMATTVGGLPCQAWSHKFPNDHKYTPTLRNGLEENFCRNPDGDPGG
NOV36f  IRTCIMNNGVGYRGTMATTVGGLSCQAWSHKFPNDHKYMPTLRNGLEENFCHNPDGDPGG

NOV36a  PWCYTTDPAVRFQSCGIKSCREAACVWCNGEEYRGAVDRTESGRECQRWDLQHPHQHPFE
NOV36b  PWCHTTDPAVRFQSCGIKSCRVAACVWCNGEEYRGAVDRTESGRECQRWDLQHPHQHPFE
NOV36c  PWCHTTDPAVRFQSCGIKSCRSAACVWCNGEEYRGAVDRTESGRECQRWDLQHPHQHPFE
NOV36d  PWCHTTDPAVRFQSCGIKSCRVAACVWCNGEEYRGAVDRTESGRECQRWDLQHPHQHPFE
NOV36e  PWCYTTDPAVRFQSCGIESCREAACVWCNGEEYRGAVDRTESGRECQRWDLQHPHQHPFE
NOV36f  PWCHTTDPAVRFQSCGIKSCRVAACVWCNGEEYRGAVDRTESGRECQRWDLQHPHQHPFE

NOV36a  PGKFLDQGLDDNYCRNPDGSERPWCYTTDPQIEREFCDLPRCGSEAQPRQEATTVSCFRG
NOV36b  PGKFLDQGLDDNYCRNPDGSERPWCYTTDPQIEREFCDLPRCGSEAQPRQEATSVSCFRG
NOV36c  PGKYPDQGLDDNYCRNPDGSERPWCYTTDPQIEREFCDLPRCGSEAQPRQEATSVSCFRG
NOV36d  PGKYLDQGLDDNYCRNPDGSERPWCYTTDPQIEREFCDLPRCGSEAQPRQEATSVSCFRG
NOV36e  PGKFLDQGLDDNYCRNPDGSERPWCYTTDPQIEREFCDLPRCGSEAQPRQEATTVSCFRG
```

```
NOV36f  PGRFLDQGLDDNYCRNPDGSERPWCYTTDPQIEREFCDLPRCGSEAQPRQEATSVSCFRG

NOV36a  KGEGYRGTANTTTAGVPCQRWDAQIPHQHRFTPEKYACKDLRENFCRNPDGSEAPWCFTL
NOV36b  KGEGYRGTANTTTAGVPCQRWDAQIPHQHRFTPEKYACKDLRENFCRNPDGSEAPWCFTL
NOV36c  KGEGYRGTANTTTAGVPCQRWDAQIPHQHRFTPEKYACKDLRENFCRNPDGSEAPWCFTL
NOV36d  KGEGYRGTANTTTAGVPCQRWDAQIPHQHRFTPEKYACKDLRENFCRNPDGSEAPWCFTL
NOV36e  KGEGYRGTANTTTAGVPCQRWDAQIPHQHRFTPEKYACKDLRENFCRNPDGSEAPWCFTL
NOV36f  KGEGYRGTANTTTAGVPCQRWDAQIPHQHRFTPEKYACKDLRENFCRNLDGSEAPWCFTL

NOV36a  RPGMRAAFCYQIRRCTDDVRPQDCYHGAGEQYRGTVSKTRKGVQCQRWSAETPHKPQFTF
NOV36b  RPGMRVGFCYQIRRCTDDVRPQDCYHGAGEQYRGTVSKTRKGVQCQRGSAETPHKPQFTF
NOV36c  RPGMRVGFCYQIRRCTDDVRPQGCYHGAGEQYRGTVSKTRKGVQCQRASAETPHKPQFTF
NOV36d  RPGMRVGFCYQIRRCTDDVRPQDCYHGAGEQYRGTVSKTRKGVQCQRASAETPHKPQFTF
NOV36e  RPGMRAAFCYQIRRCTDDVRPQ-----G-EQYRGTVSKTRKGVQCQRWSAETPHKPQFTF
NOV36f  RPGMRVGFCYQIRRCTDDVRPQDCYHGAGEQYRGTVSKTRKGVQCQRASAETPHKPQFTF

NOV36a  TSEPHAQLEENFCRNPDGDSHGPWCYTMDPRTPFDYCALRRCADDQPPSILDPP-DQVQF
NOV36b  TSEPHAQLEENFCQDPDGDSHGPWCYTMDPRTPFDYCALRRCADDQPPSILDPP-DQVQF
NOV36c  TSEPHAQLEENFCRDPDGDSYGPWCYTMDPRTPFDYCALRRCADDQPPSILDPP-DQVQF
NOV36d  TSEPHAQLEENFCQDPDGDSHGPWCYTMDPRTPFDYCALRRCADDQPPSILDPPTDQVQF
NOV36e  TSEPHAQLEENFCRNPDGDSHGPWCYTMDPRTPFDYCALRRC------------------
NOV36f  TSEPHAQLEENFCQTPDGDSHGPWCYTMDPRTPFDYCALRRCADDQPPSILDPP-DQVQF

NOV36a  EKCGKRVDRLDQRRSKLRVVGGHPGNSPWTVSLRNR-QGQHFCGGSLVKEQWILTARQCF
NOV36b  EKCGKRVDRLDQRCSKLRVAGGHPGNSPWTVSLRNR-QGQHFCGGSLVKEQWILTARQCF
NOV36c  EKCGKRVDRLDQRCSKLRVAGGHPGNSPWTVSLRNR-QGQHFCGGSLVKEQWILTARQCF
NOV36d  EKCGKRVDRLDQRRSKLRVAGGHPGNSPWTVSLGNRRQGQHFCGGSLVKEQWILTARQCF
NOV36e  ------------------------------------------------------------
NOV36f  EKCGKRVDRLDQRRSKLRVAGGHPGNSPWTVSLGNR-QGQHFCGGSLVKEQWILTARQCF

NOV36a  SSCHMPLTGYEVWLGTLFQNPQHGEPSLQRVPVAKMVCGPSGSQLVLLKLERSVTLNQRV
NOV36b  SSSHMPLTGYEVWLGTLFQNPQHGEPGLQRVPVAKMLCGPSGSQLVLLKLERSVTLNQRV
NOV36c  SSSHMPLTGYEVWLGTLFQNPQHGEPGLQRVPVAKMLCGPSGSQLVLLKLERSVTLNQRV
NOV36d  SS-HMPLTGYEVWLGTLFQNPQHGEPGLQRVPVAKMLCGPSGSQLVLLKLERSVTLNQRV
NOV36e  ------------------------------------------------------------
NOV36f  SSQHMPLTGYEVWLGTLFQNPQHGEPGLQRVPVAKMLCGPSGSQLVLLKLERSVTLNQRV

NOV36a  ALICLPPEWYVVPPGTKCEG------------------------------------------
NOV36b  ALICLPPE--------------------------------------------------------
NOV36c  ALICLPPEWYVVPPGTKCEIAGRGETKGTGNDTVLNVALLNVISNQECNIKHRGHVRESE
NOV36d  ALICLPPE--------------------------------------------------------
NOV36e  ------------------------------------------------------------
NOV36f  ALICLPPE--------------------------------------------------------

NOV36a  ----------------DYGGPLACFTHNCWVLEGIIIPNRVCARSRWPAVFTRVSVFVDW
NOV36b  ------------------------------------------------------------
NOV36c  MCTEGLLAPVGACEGGDYGGPLACFTHNCWVLEGIRIPNRVCARSRWPAVFTRVSVFVDW
NOV36d  ------------------------------------------------------------
NOV36e  ------------------------------------------------------------
NOV36f  ------------------------------------------------------------

NOV36a  IHKVMRLG
NOV36b  --------
NOV36c  IHKVMRLG
```

```
NOV36d  --------
NOV36e  --------
NOV36f  --------

NOV36a  (SEQ ID NO:  446)
NOV36b  (SEQ ID NO:  448)
NOV36c  (SEQ ID NO:  450)
NOV36d  (SEQ ID NO:  452)
NOV36e  (SEQ ID NO:  454)
NOV36f  (SEQ ID NO:  456)
```

Further analysis of the NOV36a protein yielded the following properties shown in Table 36C.

Table 36C. Protein Sequence Properties NOV36a

SignalP analysis:	Cleavage site between residues 30 and 31

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 12;  peak value   9.09
      PSG score:   4.69

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   6.16
      possible cleavage site: between 29 and 30

>>> Seems to have a cleavable signal peptide (1 to 29)

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 30
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  3.07 (at 623)
      ALOM score:   3.07  (number of TMSs: 0)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 14
      Charge difference: -3.0   C( 0.0) - N( 3.0)
      N >= C: N-terminal side will be inside

MITDISC: discrimination of mitochondrial targeting seq
      R content:       3       Hyd Moment(75):  1.11
      Hyd Moment(95):  1.26    G content:       4
      D/E content:     1       S/T content:     3
      Score: -4.52

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 44  QRS|PL

NUCDISC: discrimination of nuclear localization signals
      pat4: none
```

```
      pat7: PHTRLRR (3) at  102
      bipartite: none
      content of basic residues:  11.8%
      NLS Score: -0.22

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 55.5

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       66.7 %: extracellular, including cell wall
       22.2 %: mitochondrial
       11.1 %: vacuolar

>> prediction for CG54479-06 is exc (k=9)
```

A search of the NOV36a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 36D.

Table 36D. Geneseq Results for NOV36a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW14270	Human growth factor L5/3 - Homo sapiens, 711 aa. [US5606029-A, 25-FEB-1997]	15..670 1..711	637/711 (89%) 641/711 (89%)	0.0
AAR66602	Human L5/3 tumour suppressor protein - Homo sapiens, 711 aa. [US5315000-A, 24-MAY-1994]	15..670 1..711	637/711 (89%) 641/711 (89%)	0.0
AAY31157	Human macrophage stimulating protein - Homo sapiens, 711 aa. [US5948892-A, 07-SEP-1999]	15..670 1..711	639/716 (89%) 642/716 (89%)	0.0
ABB82662	Human hepatocyte growth factor-like protein (HGFL) - Homo sapiens, 711 aa. [WO200283074-A2, 24-OCT-2002]	15..670 1..711	636/711 (89%) 640/711 (89%)	0.0
AAB84520	Amino acid sequence of human macrophage stimulating protein (MSP) - Homo sapiens, 710 aa. [WO200144294-A2, 21-JUN-2001]	15..670 1..710	638/716 (89%) 641/716 (89%)	0.0

In a BLAST search of public sequence databases, the NOV36a protein was found to have homology to the proteins shown in the BLASTP data in Table 36E.

Table 36E. Public BLASTP Results for NOV36a				
Protein Accession Number	**Protein/Organism/Length**	**NOV36a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAH48330	Macrophage stimulating 1 (Hepatocyte growth factor-like) - Homo sapiens (Human), 711 aa.	15..670 1..711	637/711 (89%) 641/711 (89%)	0.0
P26927	Hepatocyte growth factor-like protein precursor (Macrophage stimulatory protein) (MSP) (Macrophage stimulating protein) - Homo sapiens (Human), 711 aa.	15..670 1..711	636/711 (89%) 640/711 (89%)	0.0
CAD48698	Sequence 61 from Patent WO0229058 - Homo sapiens (Human), 712 aa.	15..670 1..712	612/712 (85%) 626/712 (86%)	0.0
CAD48697	Sequence 59 from Patent WO0229058 - Homo sapiens (Human), 712 aa.	15..670 1..712	611/712 (85%) 625/712 (86%)	0.0
Q13208	Hepatocyte growth factor-like protein homolog - Homo sapiens (Human), 567 aa.	50..606 11..567	533/557 (95%) 543/557 (96%)	0.0

PFam analysis indicates that the NOV36a protein contains the domains shown in the Table 36F.

Table 36F. Domain Analysis of NOV36a			
Pfam Domain	**NOV36a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
PAN	32..120	25/110 (23%) 68/110 (62%)	1.3e-16
kringle	124..200	46/85 (54%) 70/85 (82%)	3.3e-45
kringle	205..282	48/85 (56%) 73/85 (86%)	2e-48
kringle	297..375	44/85 (52%) 73/85 (86%)	3.6e-49
kringle	384..462	44/85 (52%) 76/85 (89%)	1.5e-50
trypsin	498..663	73/263 (28%) 126/263 (48%)	7.3e-21

Example 37.

The NOV37 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 37A.

Table 37A. NOV37 Sequence Analysis		
NOV37a, CG54539-02 DNA Sequence	SEQ ID NO: 457	1567 bp
	ORF Start: ATG at 2	ORF Stop: TGA at 1523

TATGGGGTGTTGGGGTCGGAACCGGGGCCGGCTGCTGTGCATGCTGGCGCTGACCTTCATGTTCATGG
TGCTGGAGGTGGTGGTGAGCCGGGTGACCTCGTCGCTGGCGATGCTCTCCGACTCCTTCCACATGCTG
TCGGACGTGCTGGCGCTGGTGGTGGCGCTGGTGGCCGAGCGCTTCGCCCGGCGGACCCACGCCACCCA
GAAGAACACGTTCGGCTGGATCCGAGCCGAGGTAATGGGGGCTCTGGTGAACGCCATCTTCCTGACTG
GCCTCTGTTTCGCCATCCTGCTGGAGGCCATCGAGCGCTTCATCGAGCCGCACGAGATGCAGCAGCCG
CTGGTGGTCCTTGGGGTCGGCGTGGCCGGGCTGCTGGTCAACGTGCTGGGGCTCTGCCTCTTCCACCA
TCACAGCGGCTTCAGCCAGGACTCCGGCCACGGCCACTCGCACGGGGGTCACGGCTACGGCCACGGCC
TCCCCAAGGGGCCTCGCGTTAAGAGCACCCGCCCCGGGAGCAGCGACATCAACGTGGCCCCGGGCGAG
CAGGGTCCCGACCAGGAGGAGACCAACACCCTGGTGGCCAATACCAGCAACTCCAACGGGCTGAAATT
GGACCCCGCAGACCCAGAAAACCCCAGAAGTGGTGATACAGTGGAAGTACAAGTGAATGGAAATCTTG
TCAGAGAACCTGACCATATGGAACTGGAAGAAGATAGGGCTGGACAACTTAACATGCGTGGAGTTTTT
CTGCATGTCCTTGGAGATGCCTTGGGTTCAGTGATTGTAGTAGTAAATGCCTTAGTCTTTTACTTTTC
TTGGAAAGGTTGTTCTGAAGGGGATTTTTGTGTGAATCCATGTTTCCCTGACCCCTGCAAAGCATTTG
TAGAAATAATTAATAGTACTCATGCATCAGTTTATGAGGCTGGTCCTTGCTGGGTGCTATATTTAGAT
CCAACTCTTTGTGTTGTAATGGTTTGTATACTTCTTTACACAACCTATCCATTACTTAAGGAATCTGC
TCTTATTCTTCTACAAACTGTTCCTAAACAAATTGATATCAGAAATTTGATAAAAGAACTTCGAAATG
TTGAAGGAGTTGAGGAAGTTCATGAATTACATGTTTGGCAACTTGCTGGAAGCAGAATCATTGCCACT
GCTCACATAAAATGTGAAGATCCAACATCATACATGGAGGTGGCTAAAACCATTAAAGACGTTTTTCA
TAATCACGGAATTCACGCTACTACCATTCAGCCTGAATTTGCTAGTGTAGGCTCTAAATCAAGTGTAG

TTCCGTGTGAACTTGCCTGCAGAACCCAGTGTGCTTTGAAGCAATGTTGTGGGACACTACCACAAGCC
CCTTCTGGAAAGGATGCAGAAAAGACCCCAGCAGTTAGCATTTCTTGTTTAGAACTTAGTAACAATCT
AGAGAAGAAGCCCAGGAGGACTAAAGCTGAAAACATCCCTGCTGTTGTGATAGAGATTAAAAACATGC
CAAACAAACAACCTGAATCATCTTTGTGAGTCTTGAAAAAGATGTGATATTTGACTTTTGCTTTAAAC
TGC

NOV37a, CG54539-02 Protein Sequence	SEQ ID NO: 458	507 aa	MW at 55325.3kD

MGCWGRNRGRLLCMLALTFMFMVLEVVVSRVTSSLAMLSDSFHMLSDVLALVVALVAERFARRTHATQ
KNTFGWIRAEVMGALVNAIFLTGLCFAILLEAIERFIEPHEMQQPLVVLGVGVAGLLVNVLGLCLFHH
HSGFSQDSGHGHSHGGHGYGHGLPKGPRVKSTRPGSSDINVAPGEQGPDQEETNTLVANTSNSNGLKL
DPADPENPRSGDTVEVQVNGNLVREPDHMELEEDRAGQLNMRGVFLHVLGDALGSVIVVVNALVFYFS
WKGCSEGDFCVNPCFPDPCKAFVEIINSTHASVYEAGPCWVLYLDPTLCVVMVCILLYTTYPLLKESA
LILLQTVPKQIDIRNLIKELRNVEGVEEVHELHVWQLAGSRIIATAHIKCEDPTSYMEVAKTIKDVFH
NHGIHATTIQPEFASVGSKSSVVPCELACRTQCALKQCCGTLPQAPSGKDAEKTPAVSISCLELSNNL
EKKPRRTKAENIPAVVIEIKNMPNKQPESSL

NOV37b, CG54539-01 DNA Sequence	SEQ ID NO: 459	1665 bp
	ORF Start: ATG at 77	ORF Stop: TGA at 1598

CGACCCTCCGCGTCCCGCCAACGCCGCCGCTGCACCAGTCTCCGGGCCGGGCTCGGCGGGCCCCGCAG
CCGCAGCCATGGGGTGTTGGGGTCGGAACCGGGGCCGGCTGCTGTGCATGCTGGCGCTGACCTTCATG
TTCATGGTGCTGGAGGTGGTGGTGAGCCGGGTGACCTCGTCGCTGGCGATGCTCTCCGACTCCTTCCA
CATGCTGTCGGACGTGCTGGCGCTGGTGGTGGCGCTGGTGGCCGAGCGCTTCGCCCGGCGGACCCACG
CCACCCAGAAGAACACGTTCGGCTGGATCCGAGCCGAGGTAATGGGGGCTCTGGTGAACGCCATCTTC
CTGACTGGCCTCTGTTTCGCCATCCTGCTGGAGGCCATCGAGCGCTTCATCGAGCCGCACGAGATGCA
GCAGCCGCTGGTGGTCCTTGGGGTCGGCGTGGCCGGGCTGCTGGTCAACGTGCTGGGGCTCTGCCTCT
TCCACCATCACAGCGGCTTCAGCCAGGACTCCGGCCACGGCCACTCGCACGGGGGTCACGGCCACGGC
CACGGCCTCCCCAAGGGGCCTCGCGTTAAGAGCACCCGCCCCGGGAGCAGCGACATCAACGTGGCCCC
GGGCGAGCAGGGTCCCGACCAGGAGGAGACCAACACCCTGGTGGCCAATACCAGCAACTCCAACGGGC
TGAAATTGGACCCCGCGGACCCAGAAAACCCCAGAAGTGGTGATACAGTGGAAGTACAAGTGAATGGA
AATCTTGTCAGAGAACCTGACCATATGGAACTGGAAGAAGATAGGGCTGGACAACTTAACATGCGTGG
AGTTTTTCTGCATGTCCTTGGAGATGCCTTGGGTTCAGTGATTGTAGTAGTAAATGCCTTAGTCTTTT
ACTTTTCTTGGAAAGGTTGTTCTGAAGGGGATTTTTGTGTGAATCCATGTTTCCCTGACCCCTGCAAA
GCATTTGTAGAAATAATTAATAGTACTCATGCATCAGTTTATGAGGCTGGTCCTTGCTGGGTGCTATA
TTTAGATCCAACTCTTTGTGTTGTAATGGTTTGTATACTTCTTTACACAACCTATCCATTACTTAAGG
AATCTGCTCTTATTCTTCTACAAACTGTTCCTAAACAAATTGATATCAGAAATTTGATAAAAGAACTT
CGAAATGTTGAAGGAGTTGAGGAAGTTCATGAATTACATGTTTGGCAACTTGCTGGAAGCAGAATCAT
TGCCACTGCTCACATAAAATGTGAAGATCCAACATCATACATGGAGGTGGCTAAAACCATTAAAGACG
TTTTTCATAATCACGGAATTCACGCTACTACCATTCAGCCTGAATTTGCTAGTGTAGGCTCTAAATCA
AGTGTAGTTCCGTGTGAACTTGCCTGCAGAACCCAGTGTGCTTTGAAGCAATGTTGTGGGACACTACC
ACAAGCCCCTTCTGGAAAGGATGCAGAAAAGACCCCAGCAGTTAGCATTTCTTGTTTAGAACTTAGTA
ACAATCTAGAGAAGAAGCCCAGGAGGACTAAAGCTGAAAACATCCCTGCTGTTGTGATAGAGATTAAA
AACATGCCAAACAAACAACCTGAATCATCTTTGTGAGTCTTGAAAAAGATGTGATATTTGACTTTTGC
TTTAAACTGCAAGAGGAAAAAGACTCCACTGAA

NOV37b, CG54539-01 Protein Sequence	SEQ ID NO: 460	507 aa	MW at 55299.3kD

MGCWGRNRGRLLCMLALTFMFMVLEVVVSRVTSSLAMLSDSFHMLSDVLALVVALVAERFARRTHATQ
KNTFGWIRAEVMGALVNAIFLTGLCFAILLEAIERFIEPHEMQQPLVVLGVGVAGLLVNVLGLCLFHH
HSGFSQDSGHGHSHGGHGHGHGLPKGPRVKSTRPGSSDINVAPGEQGPDQEETNTLVANTSNSNGLKL
DPADPENPRSGDTVEVQVNGNLVREPDHMELEEDRAGQLNMRGVFLHVLGDALGSVIVVVNALVFYFS
WKGCSEGDFCVNPCFPDPCKAFVEIINSTHASVYEAGPCWVLYLDPTLCVVMVCILLYTTYPLLKESA
LILLQTVPKQIDIRNLIKELRNVEGVEEVHELHVWQLAGSRIIATAHIKCEDPTSYMEVAKTIKDVFH
NHGIHATTIQPEFASVGSKSSVVPCELACRTQCALKQCCGTLPQAPSGKDAEKTPAVSISCLELSNNL
EKKPRRTKAENIPAVVIEIKNMPNKQPESSL

NOV37c, SNP13382438 of CG54539-02, DNA Sequence	SEQ ID NO: 461	1567 bp
	ORF Start: ATG at 2	ORF Stop: TGA at 1523
	SNP Pos: 464	SNP Change: T to C

TATGGGGTGTTGGGGTCGGAACCGGGGCCGGCTGCTGTGCATGCTGGCGCTGACCTTCATGTTCATGG
TGCTGGAGGTGGTGGTGAGCCGGGTGACCTCGTCGCTGGCGATGCTCTCCGACTCCTTCCACATGCTG
TCGGACGTGCTGGCGCTGGTGGTGGCGCTGGTGGCCGAGCGCTTCGCCCGGCGGACCCACGCCACCCA
GAAGAACACGTTCGGCTGGATCCGAGCCGAGGTAATGGGGGCTCTGGTGAACGCCATCTTCCTGACTG
GCCTCTGTTTCGCCATCCTGCTGGAGGCCATCGAGCGCTTCATCGAGCCGCACGAGATGCAGCAGCCG
CTGGTGGTCCTTGGGGTCGGCGTGGCCGGGCTGCTGGTCAACGTGCTGGGGCTCTGCCTCTTCCACCA
TCACAGCGGCTTCAGCCAGGACTCCGGCCACGGCCACTCGCACGGGGGTCACGGCCACGGCCACGGCC
TCCCCAAGGGGCCTCGCGTTAAGAGCACCCGCCCCGGGAGCAGCGACATCAACGTGGCCCCGGGCGAG
CAGGGTCCCGACCAGGAGGAGACCAACACCCTGGTGGCCAATACCAGCAACTCCAACGGGCTGAAATT
GGACCCCGCAGACCCAGAAAACCCCAGAAGTGGTGATACAGTGGAAGTACAAGTGAATGGAAATCTTG
TCAGAGAACCTGACCATATGGAACTGGAAGAAGATAGGGCTGGACAACTTAACATGCGTGGAGTTTTT
CTGCATGTCCTTGGAGATGCCTTGGGTTCAGTGATTGTAGTAGTAAATGCCTTAGTCTTTTACTTTTC
TTGGAAAGGTTGTTCTGAAGGGGATTTTTGTGTGAATCCATGTTTCCCTGACCCCTGCAAAGCATTTG
TAGAAATAATTAATAGTACTCATGCATCAGTTTATGAGGCTGGTCCTTGCTGGGTGCTATATTTAGAT
CCAACTCTTTGTGTTGTAATGGTTTGTATACTTCTTTACACAACCTATCCATTACTTAAGGAATCTGC
TCTTATTCTTCTACAAACTGTTCCTAAACAAATTGATATCAGAAATTTGATAAAAGAACTTCGAAATG
TTGAAGGAGTTGAGGAAGTTCATGAATTACATGTTTGGCAACTTGCTGGAAGCAGAATCATTGCCACT
GCTCACATAAAATGTGAAGATCCAACATCATACATGGAGGTGGCTAAAACCATTAAAGACGTTTTTCA
TAATCACGGAATTCACGCTACTACCATTCAGCCTGAATTTGCTAGTGTAGGCTCTAAATCAAGTGTAG
TTCCGTGTGAACTTGCCTGCAGAACCCAGTGTGCTTTGAAGCAATGTTGTGGGACACTACCACAAGCC
CCTTCTGGAAAGGATGCAGAAAAGACCCCAGCAGTTAGCATTTCTTGTTTAGAACTTAGTAACAATCT
AGAGAAGAAGCCCAGGAGGACTAAAGCTGAAAACATCCCTGCTGTTGTGATAGAGATTAAAAACATGC
CAAACAAACAACCTGAATCATCTTTGTGAGTCTTGAAAAAGATGTGATATTTGACTTTTGCTTTAAAC
TGC

NOV37c, SNP13382438 of CG54539-02, Protein Sequence	SEQ ID NO: 462	507 aa	MW at 55299.3kD
	SNP Pos: 155		SNP Change: Tyr to His

MGCWGRNRGRLLCMLALTFMFMVLEVVVSRVTSSLAMLSDSFHMLSDVLALVVALVAERFARRTHATQ
KNTFGWIRAEVMGALVNAIFLTGLCFAILLEAIERFIEPHEMQQPLVVLGVGVAGLLVNVLGLCLFHH
HSGFSQDSGHGHSHGGHGHGHGLPKGPRVKSTRPGSSDINVAPGEQGPDQEETNTLVANTSNSNGLKL
DPADPENPRSGDTVEVQVNGNLVREPDHMELEEDRAGQLNMRGVFLHVLGDALGSVIVVVNALVFYFS
WKGCSEGDFCVNPCFPDPCKAFVEIINSTHASVYEAGPCWVLYLDPTLCVVMVCILLYTTYPLLKESA
LILLQTVPKQIDIRNLIKELRNVEGVEEVHELHVWQLAGSRIIATAHIKCEDPTSYMEVAKTIKDVFH
NHGIHATTIQPEFASVGSKSSVVPCELACRTQCALKQCCGTLPQAPSGKDAEKTPAVSISCLELSNNL
EKKPRRTKAENIPAVVIEIKNMPNKQPESSL

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 37B.

Table 37B. Comparison of the NOV37 protein sequences.

```
NOV37a  MGCWGRNRGRLLCMLALTFMFMVLEVVVSRVTSSLAMLSDSFHMLSDVLALVVALVAERF
NOV37b  MGCWGRNRGRLLCMLALTFMFMVLEVVVSRVTSSLAMLSDSFHMLSDVLALVVALVAERF

NOV37a  ARRTHATQKNTFGWIRAEVMGALVNAIFLTGLCFAILLEAIERFIEPHEMQQPLVVLGVG
NOV37b  ARRTHATQKNTFGWIRAEVMGALVNAIFLTGLCFAILLEAIERFIEPHEMQQPLVVLGVG

NOV37a  VAGLLVNVLGLCLFHHHSGFSQDSGHGHSHGGHGYGHGLPKGPRVKSTRPGSSDINVAPG
NOV37b  VAGLLVNVLGLCLFHHHSGFSQDSGHGHSHGGHGHGHGLPKGPRVKSTRPGSSDINVAPG
```

```
NOV37a  EQGPDQEETNTLVANTSNSNGLKLDPADPENPRSGDTVEVQVNGNLVREPDHMELEEDRA
NOV37b  EQGPDQEETNTLVANTSNSNGLKLDPADPENPRSGDTVEVQVNGNLVREPDHMELEEDRA

NOV37a  GQLNMRGVFLHVLGDALGSVIVVVNALVFYFSWKGCSEGDFCVNPCFPDPCKAFVEIINS
NOV37b  GQLNMRGVFLHVLGDALGSVIVVVNALVFYFSWKGCSEGDFCVNPCFPDPCKAFVEIINS

NOV37a  THASVYEAGPCWVLYLDPTLCVVMVCILLYTTYPLLKESALILLQTVPKQIDIRNLIKEL
NOV37b  THASVYEAGPCWVLYLDPTLCVVMVCILLYTTYPLLKESALILLQTVPKQIDIRNLIKEL

NOV37a  RNVEGVEEVHELHVWQLAGSRIIATAHIKCEDPTSYMEVAKTIKDVFHNHGIHATTIQPE
NOV37b  RNVEGVEEVHELHVWQLAGSRIIATAHIKCEDPTSYMEVAKTIKDVFHNHGIHATTIQPE

NOV37a  FASVGSKSSVVPCELACRTQCALKQCCGTLPQAPSGKDAEKTPAVSISCLELSNNLEKKP
NOV37b  FASVGSKSSVVPCELACRTQCALKQCCGTLPQAPSGKDAEKTPAVSISCLELSNNLEKKP

NOV37a  RRTKAENIPAVVIEIKNMPNKQPESSL
NOV37b  RRTKAENIPAVVIEIKNMPNKQPESSL

NOV37a  (SEQ ID NO:  458)
NOV37b  (SEQ ID NO:  460)
```

Further analysis of the NOV37a protein yielded the following properties shown in Table 37C.

Table 37C. Protein Sequence Properties NOV37a

SignalP analysis:	Cleavage site between residues 30 and 31

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 10;  pos.chg 3;  neg.chg 0
      H-region:  length 14;  peak value  12.34
      PSG score:   7.94

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -2.29
      possible cleavage site: between 29 and 30

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood = -8.76   Transmembrane   12 -  28
      INTEGRAL    Likelihood = -3.66   Transmembrane   41 -  57
      INTEGRAL    Likelihood = -7.54   Transmembrane   82 -  98
      INTEGRAL    Likelihood = -7.43   Transmembrane  114 - 130
      INTEGRAL    Likelihood = -5.57   Transmembrane  248 - 264
      INTEGRAL    Likelihood = -7.01   Transmembrane  313 - 329
      PERIPHERAL  Likelihood =  6.21 (at 421)
      ALOM score:  -8.76  (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
```

```
      Center position for calculation: 19
      Charge difference: -4.0   C( 0.0) - N( 4.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       3       Hyd Moment(75):  5.99
      Hyd Moment(95):  2.66    G content:       3
      D/E content:     1       S/T content:     1
      Score: -3.81

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 40  SRV|TS

NUCDISC: discrimination of nuclear localization signals
      pat4: KKPR (4) at  478
      pat4: KPRR (4) at  479
      pat7: PRRTKAE (5) at  480
      bipartite: none
      content of basic residues:   8.3%
      NLS Score:  0.46

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none
```

```
NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       22.2 %: mitochondrial
       11.1 %: nuclear

>> prediction for CG54539-02 is end (k=9)
```

A search of the NOV37a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 37D.

Table 37D. Geneseq Results for NOV37a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV37a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABG73033	Human transporter polypeptide - Homo sapiens, 507 aa. [US2002106721-A1, 08-AUG-2002]	1..507 1..507	506/507 (99%) 507/507 (99%)	0.0
ABG73033	Human transporter polypeptide - Homo sapiens, 507 aa. [US2002106721-A1, 08-AUG-2002]	1..507 1..507	506/507 (99%) 507/507 (99%)	0.0
AAG67549	Amino acid sequence of a human transporter protein - Homo sapiens, 507 aa. [WO200164878-A2, 07-SEP-2001]	1..507 1..507	504/507 (99%) 506/507 (99%)	0.0
AAE16348	Human zinc transporter like protein, POLY12 - Homo sapiens, 512 aa. [WO200185767-A2, 15-NOV-2001]	1..507 1..512	506/512 (98%) 507/512 (98%)	0.0
AAY86241	Human secreted protein HOABR60, SEQ ID NO:156 - Homo sapiens, 490 aa. [WO9966041-A1, 23-DEC-1999]	14..501 1..488	482/488 (98%) 484/488 (98%)	0.0

In a BLAST search of public sequence databases, the NOV37a protein was found to have homology to the proteins shown in the BLASTP data in Table 37E.

Table 37E. Public BLASTP Results for NOV37a				
Protein Accession Number	**Protein/Organism/Length**	**NOV37a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9Y6M5	Zinc transporter 1 (ZnT-1) - Homo sapiens (Human), 507 aa.	1..507 1..507	505/507 (99%) 506/507 (99%)	0.0
Q62720	Zinc transporter 1 (ZnT-1) - Rattus norvegicus (Rat), 507 aa.	1..507 1..507	437/511 (85%) 468/511 (91%)	0.0
Q60738	Zinc transporter 1 (ZnT-1) - Mus musculus (Mouse), 503 aa.	1..507 1..503	431/509 (84%) 464/509 (90%)	0.0
AAH46675	Similar to solute carrier family 30 (zinc transporter), member 1 - Xenopus laevis (African clawed frog), 494 aa.	4..500 2..493	297/508 (58%) 366/508 (71%)	e-156
CAD58282	Sequence 16 from Patent WO0240656 - Homo sapiens (Human), 485 aa.	1..485 1..454	183/503 (36%) 254/503 (50%)	2e-69

PFam analysis indicates that the NOV37a protein contains the domains shown in the Table 37F.

Table 37F. Domain Analysis of NOV37a			
Pfam Domain	**NOV37a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Cation_efflux	11..426	102/435 (23%) 323/435 (74%)	4.8e-56

Example 38.

The NOV38 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 38A.

Table 38A. NOV38 Sequence Analysis		
NOV38a, CG54683-05 DNA Sequence	SEQ ID NO: 463	1416 bp
	ORF Start: ATG at 4	ORF Stop: TGA at 1402
GAGATGGTCCTGGCTTTCCAGTTAGTCTCCTTCACCTACATCTGGATCATATTGAAACCAAATGTTTG TGCTGCTTCTAACATCAAGATGACACACCAGCGGTGCTCCTCTTCAATGAAACAAACCCAAGAAACTA GAATGAAGAAAGATGACAGTACCAAAGCGCGGCCTCAGAAATATGAGCAACTTCTCCATATAGAGGAC		

AACGATTTCGCAATGAGACCTGGATTTGGAGGGTCTCCAGTGCCAGTAGGTATAGATGTCCATGTTGA
AAGCATTGACAGCATTTCAGAGACTAACATGGTAGACTTTACAATGACTTTTTATCTCAGGCATTACT
GGAAAGACGAGAGGCTCTCCTTTCCTAGCACAGCAAACAAAAGCATGACATTTGATCATAGATTGACC
AGAAAGATCTGGGTGCCTGATATCTTTTTTGTCCACTCTAAAAGATCCTTCATCCATGATACAACTAT
GGAGAATATCATGCTGCGCGTACACCCTGATGGAAACGTCCTCCTAAGTCTCAGGATAACGGTTTCGG
CCATGTGCTTTATGGATTTCAGCAGGTTTCCTCTTGACACTCAAAATTGTTCTCTTGAACTGGAAAGC
GCCTACAATGAGGATGACCTAATGCTATACTGGAAACACGGAAACAAGTCCTTAAATACTGAAGAACA
TATGTCCCTTTCTCAGTTCTTCATTGAAGACTTCAGTGCATCTAGTGGATTAGCTTTCTATAGCAGCA
CAGGTTGGTACAATAGGCTTTTCATCAACTTTGTGCTAAGGAGGCATGTTTTCTTCTTTGTGCTGCAA
ACCTATTTCCCAGCCATATTGATGGTGATGCTTTCATGGGTTTCATTTTGGATTGACCGAAGAGCTGT
TCCTGCAAGAGTTTCCCTGGGTGGAATCACCACAGTGCTGACCATGTCCACAATCATCACTGCTGTGA
GCGCCTCCATGCCCCAGGTGTCCTACCTCAAGGCTGTGGATGTGTACCTGTGGGTCAGCTCCCTCTTT
GTGTTCCTGTCAGTCATTGAGTATGCAGCTGTGAACTACCTCACCACAGTGGAAGAGCGGAAACAATT
CAAGAAGACAGGAAAGATTTCTAGGATGTACAATATTGATGCAGTTCAAGCTATGGCCTTTGATGGTT
GTTACCATGACAGCGAGATTGACATGGACCAGACTTCCCTCTCTCTAAACTCAGAAGACTTCATGAGA
AGAAAATCGATATGCAGCCCCAGCACCGATTCATCTCGGATAAAGAGAAGAAAATCCCTAGGAGGACA
TGTTGGTAGAATCATTCTGGAAAACAACCATGTCATTGACACCTATTCTAGGATTNTATTCCCCATTG
TGTATATCTTATTTAATTTGTTTTACTGGGGTGTATATGTA**TGA**AGGGGAATTTCA

NOV38a, CG54683-05 Protein Sequence	SEQ ID NO: 464	466 aa	MW at 53919.5kD

MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTQETRMKKDDSTKARPQKYEQLLHIEDN
DFAMRPGFGGSPVPVGIDVHVESIDSISETNMVDFTMTFYLRHYWKDERLSFPSTANKSMTFDHRLTR
KIWVPDIFFVHSKRSFIHDTTMENIMLRVHPDGNVLLSLRITVSAMCFMDFSRFPLDTQNCSLELESA
YNEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDFSASSGLAFYSSTGWYNRLFINFVLRRHVFFFVLQT
YFPAILMVMLSWVSFWIDRRAVPARVSLGGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFV
FLSVIEYAAVNYLTTVEERKQFKKTGKISRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDFMRR
KSICSPSTDSSRIKRRKSLGGHVGRIILENNHVIDTYSRIXFPIVYILFNLFYWGVYV

NOV38b, CG54683-01 DNA Sequence	SEQ ID NO: 465	1875 bp
	ORF Start: ATG at 10	ORF Stop: TGA at 1411

TTGGAAGAG**ATG**GTCCTGGCTTTCCAGTTAGTCTCCTTCACCTACATCTGGATCATATTGAAACCAAA
TGTTTGTGCTGCTTCTAACATCAAGATGACACACCAGCGGTGCTCCTCTTCAATGAAACAAACCTGGA
TGCAAGAAACTAGAATGAAGAAAGATGACAGTACCAAAGCGCGGCCTCAGAAATATGAGCAACTTCTC
CATATAGAGGACAACGATTTCGCAATGAGACCTGGATTTGGAGGTTCTCCAGTGCCAGTAGGTATAGA
TGTCCATGTTGAAAGCATTGACAGCATTTCAGAGACTAACATGGACTTTACAATGACTTTTTATCTCA
GGCATTACTGGAAAGACGAGAGGCTCTCCTTTCCTAGCACAGCAAACAAAAGCATGACATTTGATCAT
AGACACTTGCGGTATTCGTTATTCATCAGAAGGCTGTATCTGTTATACTGCCAGAGGTCTTTCTTCTC
ACCCTCATCCATACTTCCCTCATCTCCAGACATCCATGCACCTGGTACATCTAAAAGCAGTTTGTCTG
ATAGCCTTGTATGTATATCTGAAAAAAACTTGCCAGGACACAGTAAAAACACACCTCTTGCAATGTCA
GATGTAGCCTACAATGAGGATGACCTAATGCTATACTGGAAACACGGAAACAAGTCCTTAAATACTGA
AGAACATATGTCCCTTTCTCAGTTCTTCATTGAAGACTTCAGTGCATCTAGTGGATTAGCTTTCTATA
GCAGCACAGGTACAGCATTTTACATGGGTGATTCATCAGCATTTATTGGACATCTACTGTTTTTGATC
TGGAGTTCCAGGAAAAGACCAGGTTTAGAGATGTTGGGTTTGGGAATTCTCAGAATCTGGGTAATAAC
TAGAGCCATGGATAAGAAAATGGAAATGGGAATCACCACAGTGCTGACCATGTCCACAATCATCACTG
CTGTGAGCGCCTCCATGCCCCAGGTGTCCTACCTCAAGGCTGTGGATGTGTACCTGTGGGTCAGCTCC
CTCTTTGTGTTCCTGTCAGTCATTGAGTATGCAGCTGTGAACTACCTCACCACAGTGGAAGAGCGGAA
ACAATTCAAAAAAAGTTTTTCAAAGATTTCTAGGATGTACAATATTGATGCAGTTCAAGCTATGGCCT
TTGATGGTTGTTACCATGACAGCGAGATTGACATGGACCAGACTTCCCTCTCTCTAAACTCAGAAGAC
TTCATGAGAAGAAAATCGATATGCAGCCCCAGCACCGATTCATCTCGGATAAAGAGAAGAAAATCCCT
AGGAGGACATGTTGGTAGAATCATTCTGGAAAACAACCATGTCATTGACACCTATTCTAGGATTTTAT
TCCCCATTGTGTATATCTTTTTTAATTTGTTTTACTGGGGTGTATATGTA**TGA**AGGGGAATTTCAAAT
GTATACAACTTTAAAGCCAGATGATGTTTAAAAACAAAACTCTTGAATATGAGTTGGATAGTCCTAGA
TGGAACTGGGAAAGAGCAAGTCACCTCTCCTGCCCTAATGAAAATTTGAAAGCTGTCTGATTTACATC
TAAGAAAGAGTTTAGGTCCTAGAAAAGTTTGACTCCATAAATAAGAGTCATAGGCATGTGTATTATGG
GAAAAACAGTTTTCCATTGGGAAGGGCTTTATAACTACTTCATCTGAACCCTCCTTCTTTCTTAATGA
AATGTTCTTTATTTAACTAGGGAAGAAAGCTGGACTATAACAATAATTCAAAGATATTTTGTTTCTTA
GTGCCAGCCAAGTGCCTGGTTATCTACCAGAGCTCAACCGTCCTAGGCAAGAACATCCACATAGAGGT
GGTATCATCCACACTCACACAGCTGAGAATCCTATGAAG

NOV38b, CG54683-01 Protein Sequence	SEQ ID NO: 466	467 aa	MW at 53522.9kD

MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTWMQETRMKKDDSTKARPQKYEQLLHIE
DNDFAMRPGFGGSPVPVGIDVHVESIDSISETNMDFTMTFYLRHYWKDERLSFPSTANKSMTFDHRHL
RYSLFIRRLYLLYCQRSFFSPSSILPSSPDIHAPGTSKSSLSDSLVCISEKNLPGHSKNTPLAMSDVA
YNEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDFSASSGLAFYSSTGTAFYMGDSSAFIGHLLFLIWSS
RKRPGLEMLGLGILRIWVITRAMDKKMEMGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFV
FLSVIEYAAVNYLTTVEERKQFKKSFSKISRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDFMR
RKSICSPSTDSSRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYIFFNLFYWGVYV

NOV38c, CG54683-02 DNA Sequence	SEQ ID NO: 467	1444 bp
	ORF Start: ATG at 21	ORF Stop: TGA at 1425

GTTTTTTTGTTTTGGAAGAG**ATG**GTCCTGGCTTTCCAGTTAGTCTCCTTCACCTACATCTGGATCATA
TTGAAACCAAATGTTTGTGCTGCTTCTAACATCAAGATGACACACCAGCGGTGCTCCTCTTCAATGAA
ACAAACCTGCAAACAAGAAACTAGAATGAAGAAAGATGACAGTACCAAAGCGCGGCCTCAGAAATATG
AGCAACTTCTCCATATAGAGGACAACGATTTCGCAATGAGACCTGGATTTGGAGGGTCTCCAGTGCCA
GTAGGTATAGATGCCCATGTTGAAAGCATTGACAGCATTTCAGAGACTAACATGGACTTTACAATGAC
TTTTTATCTCAGGCATTACTGGAAAGACGAGAGGCTCTCCTTTCCTAGCACAGCAAACAAAAGCATGA
CATTTGATCATAGATTGACCAGAAAGATCTGGGTGCCTGATATCTTTTTTGTCCACTCTAAAAGATCC
TTCATCCATGATACAACTATGGAGAATATCATGCTGCGCGTACACCCTGATGGAAACGTCCTCCTAAG
TCTCAGGATAACGGTTTCGGCCATGTGCTTTATGGATTTCAGCAGGTTTCCTCTTGACGACACTCAAA
ATTGTTCTCTTGAACTGGAAAGCTGTGCCTACAATGAGGATGACCTAATGCTATACTGGAAACACGGA
AACAAGTCCTTAAATACTGAAGAACATATGTCCCTTTCTCAGTTCTTCATTGAAGACTTCAGTGCATC
TAGTGGATTAGCTTTCTATAGCAGCACAGGTTGGTACAATAGGCTTTTCATCAACTTTGTGCTAAGGA
GGCATGTTTTCTTCTTTGTGCTGCAAACCTATTTCCCAGCCATATTGATGGTGATGCTTTCATGGGTT
TCATTTTGGATTGACCGAAGAGCTGTTCCTGCAAGAGTTTCCCTGGGTATCACCACAGTGCTGACCAT
GTCCACAATCATCACTGCTGTGAGCGCCTCCATGCCCCAGGTGTCCTACCTCAAGGCTGTGGATGTGT
ACCTGTGGGTCAGCTCCCTCTTTGTGTTCCTGTCAGTCATTGAGTATGCAGCTGTGAACTACCTCACC
ACAGTGGAAGAGCGGAAACAATTCAAGAAGACAGGAAAGGTATCTAGGATGTACAATATTGATGCAGT
TCAAGCTATGGCCTTTGATGGTTGTTACCATGACAGCGAGATTGACATGGACCAGACTTCCCTCTCTC
TAAACTCAGAAGACTTCATGAGAAGAAAATCGATATGCAGCCCCAGCACCGATTCATCTCGGATAAAG
AGAAGAAAATCCCTAGGAGGACATGTTGGTAGAATCATTCTGGAAAACAACCATGTCATTGACACCTA
TTCTAGGATTTTATTCCCCATTGTGTATATTTTATTTAATTTGTTTTACTGGGGTGTATATGTATGAA
GGGGAATTTCAAATGT

NOV38c, CG54683-02 Protein Sequence	SEQ ID NO: 468	468 aa	MW at 54283.9kD

MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTCKQETRMKKDDSTKARPQKYEQLLHIE
DNDFAMRPGFGGSPVPVGIDAHVESIDSISETNMDFTMTFYLRHYWKDERLSFPSTANKSMTFDHRLT
RKIWVPDIFFVHSKRSFIHDTTMENIMLRVHPDGNVLLSLRITVSAMCFMDFSRFPLDDTQNCSLELE
SCAYNEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDFSASSGLAFYSSTGWYNRLFINFVLRRHVFFFV
LQTYFPAILMVMLSWVSFWIDRRAVPARVSLGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSL
FVFLSVIEYAAVNYLTTVEERKQFKKTGKVSRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDFM
RRKSICSPSTDSSRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYILFNLFYWGVYV

NOV38d, CG54683-03 DNA Sequence	SEQ ID NO: 469	1438 bp
	ORF Start: ATG at 21	ORF Stop: TGA at 1419

GTTTTTTTGTTTTGGAAGAG**ATG**GTCCTGGCTTTCCAGTTAGTCTCCTTCACCTACATCTGGATCATA
TTGAAACCAAATGTTTGTGCTGCTTCTAACATCAAGATGACACACCAGCGGTGCTCCTCTTCAATGAA
ACAAACCTGCAAACAAGAAACTAGAATGAAGAAAGATGACAGTACCAAAGCGCGGCCTCAGAAATATG
AGCAACTTCTCCATATAGAGGACAACGATTTCGCAATGAGACCTGGATTTGGAGGGTCTCCAGTGCCA
GTAGGTATAGATGTCCATGTTGAAAGCATTGACAGCATTTCAGAGACTAACATGGACTTTACAATGAC
TTTTTATCTCAGGCATTACTGGAAAGACGAGAGGCTCTCCTTTCCTAGCACAGCAAACAAAAGCATGA
CATTTGATCATAGATTGACCAGAAAGATCTGGGTGCCTGATATCTTTTTTGTCCACTCTAAAAGATCC
TTCATCCATGATACAACTATGGAGAATATCATGCTGCGCGTACACCCTGATGGAAACGTCCTCCTAAG
TCTCAGGATAACGGTTTCGGCCATGTGCTTTATGGATTTCAGCAGGTTTCCTCTGACTCAAAATTGTT
CTCTTGAACTGGAAAGCTGTGCCTACAATGAGGATGACCTAATGCTATACTGGAAACACGGAAACAAG
TCCTTAAATACTGAAGAACATATGTCCCTTTCTCAGTTCTTCATTGAAGACTTCAGTGCATCTAGTGG

ATTAGCTTTCTATAGCAGCACAGGTTGGTACAATAGGCTTTTCATCAACTTTGTGCTAAGGAGGCATG
TTTTCTTCTTTGTGCTGCAAACCTATTTCCCAGCCATATTGATGGTGATGCTTTCATGGGTTTCATTT
TGGATTGACCGAAGAGCTGTTCCTGCAAGAGTTTCCCTGGGTATCACCACAGTGCTGACCATGTCCAC
AATCATCACTGCTGTGAGCGCCTCCATGCCCCAGGTGTCCTACCTCAAGGCTGTGGATGTGTACCTGT
GGGTCAGCTCCCTCTTTGTGTTCCTGTCAGTCATTGAGTATGCAGCTGTGAACTACCTCACCACAGTG
GAAGAGCGGAAACAATTCAAGAAGACAGGAAAGGTATCTAGGATGTACAATATTGATGCAGTTCAAGC
TATGGCCTTTGATGGTTGTTACCATGACAGCGAGATTGACATGGACCAGACTTCCCTCTCTCTAAACT
CAGAAGACTTCATGAGAAGAAAATCGATATGCAGCCCCAGCACCGATTCATCTCGGATAAAGAGAAGA
AAATCCCTAGGAGGACATGTTGGTAGAATCATTCTGGAAAACAACCATGTCATTGACACCTATTCTAG
GATTTTATTCCCCATTGTGTATATTTTATTTAATTTGTTTTACTGGGGTGTATATGTATGAAGGGGAA
TTTCAAATGT

NOV38d, CG54683-03 Protein Sequence	SEQ ID NO: 470	466 aa	MW at 54081.8kD

MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTCKQETRMKKDDSTKARPQKYEQLLHIE
DNDFAMRPGFGGSPVPVGIDVHVESIDSISETNMDFTMTFYLRHYWKDERLSFPSTANKSMTFDHRLT
RKIWVPDIFFVHSKRSFIHDTTMENIMLRVHPDGNVLLSLRITVSAMCFMDFSRFPLTQNCSLELESC
AYNEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDFSASSGLAFYSSTGWYNRLFINFVLRRHVFFFVLQ
TYFPAILMVMLSWVSFWIDRRAVPARVSLGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFV
FLSVIEYAAVNYLTTVEERKQFKKTGKVSRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDFMRR
KSICSPSTDSSRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYILFNLFYWGVYV

NOV38e, CG54683-04 DNA Sequence	SEQ ID NO: 471	1799 bp
	ORF Start: ATG at 71	ORF Stop: TGA at 1445

AAGAAGAAACTGTGATCACAGTATTGGTTGCGTTCACCTGCATCCTTTCTGTTTTTTTGTTTTGGAAG
AGATGGTCCTGGCTTTCCAGTTAGTCTCCTTCACCTACATCTGGATCATATTGGTTTGTGCTGCTTCT
AACATCAAGATGACACACCAGCGGTGCTCCTCTTCAATGAAACAAACCGTAAGATGCTCAATGAAGAA
AGATGACAGTACCAAAGCGCGGCCTCAGAAATATGAGCAACTTCTCCATATAGAGGACAACGATTTCG
CAATGAGACCTGGATTTGGAGGTTCTCCAGTGCCAGTAGGTATAGATGTCCATGTTGAAAGCATTGAC
AGCATTTCAGAGACTAACATGGACTTTACAATGACTTTTTATCTCAGGCATTACTGGAAAGACGAGAG
GCTCTCCTTTCCTAGCACAGCAAACAAAAGCATGACATTTGATCATAGAAAGAGTATCCCCCGCCCTG
AACACTTGCGGTATTCGTTATTCATCAGAAGGCTGTATCTGTTATACTGCCAGAGGTCTTTCTTCTCA
CCCTCATCCATACTTCCCTCATCTCCAGACATCCATGCACCTGGTACATCTAAAAGCAGTTTGTCTGA
TAGCCTTGTATGTATATCTGAAAAAAACTTGCCAGGACACAGTAAAAACACACCTCTTGCAATGGCCT
ACAATGAGGATGACCTAATGCTATACTGGAAACACGGAAACAAGTCCTTAAATACTGAAGAACATATG
TCCCTTTCTCAGTTCTTCATTGAAGACTTCAGTGCATCTAGTGGATTAGCTTTCTATAGCAGCACAGG
TACAGCATTTTACATGGGTGATTCATCAGCATTTATTGGACATCTACTGTTTTTAAATAGACATTTAC
ATTTCTTCATCATAAATTTTGAAATTACTCAAATATTGATGATTGGAATCACCACAGTGCTGACCATG
TCCACAATCATCACTGCTGTGAGCGCCTCCATGCCCCAGGTGTCCTACCTCAAGGCTGTGGATGTGTA
CCTGTGGGTCAGCTCCCTCTTTGTGTTCCTGTCAGTCATTGAGTATGCAGCTGTGAACTACCTCACCA
CAGTGGAAGAGCGGAAACAATTCAAGAAGACAGGAAAGGTACAGATTTCTAGGATGTACAATATTGAT
GCAGTTCAAGCTATGGCCTTTGATGGTTGTTACCATGACAGCGAGATTGACATGGACCAGACTTCCCT
CTCTCTAAACTCAGAAGACTTCATGAGAAGAAAATCGATATGCAGCCCCAGCACCGATTCATCTCGGA
TAAAGAGAAGAAAATCCCTAGGAGGACATGTTGGTAGAATCATTCTGGAAAACAACCATGTCATTGAC
ACCTATTCTAGGATTTTATTCCCCATTGTGTATATCCCATTGTGTATATCTTTATTTAATTTGTTTTA
CTGGGGTGTATATGTATGAAGGGGAATTTCAAATGTATACAACTTTAAAGCCAGATGATGTTTAAAAA
CAAAACTCTTGAATATGAGTTGGAATTGAAGACTTCAGTGCATCTAGTGGATTAGCTTTCTATAGCAG
CACAGGTACAGCATTTTACATGGGTGATTCATCAGCATTTATTGGACATCTACTGTTTTACTTTTGGT
CTTTGATGATGGTGATGTACAGATGGGTTGGAATCACCACAGTGCTGACCATGTCCACAATCATCACT
GCTGTGAGCGCCTCCATGCCCCAGGTGTCCTACCTCAAGGCTGTGGATGTGTACCTGTGGGTCAGCTC
CCTCTTTGTGTTCCTGTCAGTCATTGAGTAT

NOV38e, CG54683-04 Protein Sequence	SEQ ID NO: 472	458 aa	MW at 52330.5kD

MVLAFQLVSFTYIWIILVCAASNIKMTHQRCSSSMKQTVRCSMKKDDSTKARPQKYEQLLHIEDNDFA
MRPGFGGSPVPVGIDVHVESIDSISETNMDFTMTFYLRHYWKDERLSFPSTANKSMTFDHRKSIPRPE
HLRYSLFIRRLYLLYCQRSFFSPSSILPSSPDIHAPGTSKSSLSDSLVCISEKNLPGHSKNTPLAMAY
NEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDFSASSGLAFYSSTGTAFYMGDSSAFIGHLLFLNRHLH
FFIINFEITQILMIGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEYAAVNYLTT

VEERKQFKKTGKVQISRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDFMRRKSICSPSTDSSRI
KRRKSLGGHVGRIILENNHVIDTYSRILFPIVYIPLCISLFNLFYWGVYV

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 38B.

Table 38B. Comparison of the NOV38 protein sequences.

```
NOV38a  MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQT--QETRMKKDDSTKARPQK
NOV38b  MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTWMQETRMKKDDSTKARPQK
NOV38c  MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTCKQETRMKKDDSTKARPQK
NOV38d  MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTCKQETRMKKDDSTKARPQK
NOV38e  MVLAFQLVSFTYIWIIL---VCAASNIKMTHQRCSSSMKQTVRCS--MKKDDSTKARPQK

NOV38a  YEQLLHIEDNDFAMRPGFGGSPVPVGIDVHVESIDSISETNMVDFTMTFYLRHYWKDERL
NOV38b  YEQLLHIEDNDFAMRPGFGGSPVPVGIDVHVESIDSISETN-MDFTMTFYLRHYWKDERL
NOV38c  YEQLLHIEDNDFAMRPGFGGSPVPVGIDAHVESIDSISETN-MDFTMTFYLRHYWKDERL
NOV38d  YEQLLHIEDNDFAMRPGFGGSPVPVGIDVHVESIDSISETN-MDFTMTFYLRHYWKDERL
NOV38e  YEQLLHIEDNDFAMRPGFGGSPVPVGIDVHVESIDSISETN-MDFTMTFYLRHYWKDERL

NOV38a  SFPSTANKSMTFDHR--------LTRKIWVPDIFFVHSKRSFIHDTTMENIMLRVHPDGN
NOV38b  SFPSTANKSMTFDHR-------HLRYSLFIRRLYLLYCQRSFFSPSSILPSSPDIHAPG-
NOV38c  SFPSTANKSMTFDHR--------LTRKIWVPDIFFVHSKRSFIHDTTMENIMLRVHPDGN
NOV38d  SFPSTANKSMTFDHR--------LTRKIWVPDIFFVHSKRSFIHDTTMENIMLRVHPDGN
NOV38e  SFPSTANKSMTFDHRKSIPRPEHLRYSLFIRRLYLLYCQRSFFSPSSILPSSPDIHAPG-

NOV38a  VLLSLRITVSAMCFMDFSRFPLD-TQNCSLELES-AYNEDDLMLYWKHGNKSLNTEEHMS
NOV38b  --TSKSSLSDSLVCISEKNLPG-HSKNTPLAMSDVAYNEDDLMLYWKHGNKSLNTEEHMS
NOV38c  VLLSLRITVSAMCFMDFSRFPLDDTQNCSLELESCAYNEDDLMLYWKHGNKSLNTEEHMS
NOV38d  VLLSLRITVSAMCFMDFSRFPL--TQNCSLELESCAYNEDDLMLYWKHGNKSLNTEEHMS
NOV38e  --TSKSSLSDSLVCISEKNLPG-HSKNTPLAM---AYNEDDLMLYWKHGNKSLNTEEHMS

NOV38a  LSQFFIEDFSASSGLAFYSSTGWYNRLFINFVLRRHVFFFVLQT-YFPAILMVMLSWVSF
NOV38b  LSQFFIEDFSASSGLAFYSSTGTAFYMGDSSAFIGHLLFLIWSSRKRPGLEMLGLGILRI
NOV38c  LSQFFIEDFSASSGLAFYSSTGWYNRLFINFVLRRHVFFFVLQT-YFPAILMVMLSWVSF
NOV38d  LSQFFIEDFSASSGLAFYSSTGWYNRLFINFVLRRHVFFFVLQT-YFPAILMVMLSWVSF
NOV38e  LSQFFIEDFSASSGLAFYSSTGTAFYMGDSSAFIGHLLFLN-----R---------HLHF

NOV38a  WIDRRAVPARVSLGGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEY
NOV38b  WVITRAMDKKMEMG-ITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEY
NOV38c  WIDRRAVPARVSLG-ITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEY
NOV38d  WIDRRAVPARVSLG-ITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEY
NOV38e  FIINFEITQILMIG-ITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEY

NOV38a  AAVNYLTTVEERKQFKKT--GKISRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDF
NOV38b  AAVNYLTTVEERKQFKKS-FSKISRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDF
NOV38c  AAVNYLTTVEERKQFKKT--GKVSRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDF
NOV38d  AAVNYLTTVEERKQFKKT--GKVSRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDF
NOV38e  AAVNYLTTVEERKQFKKTGKVQISRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDF

NOV38a  MRRKSICSPSTDSSRIKRRKSLGGHVGRIILENNHVIDTYSRIXFPIVYI-----LFNLF
NOV38b  MRRKSICSPSTDSSRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYI-----FFNLF
NOV38c  MRRKSICSPSTDSSRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYI-----LFNLF
NOV38d  MRRKSICSPSTDSSRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYI-----LFNLF
```

```
NOV38e  MRRKSICSPSTDSSRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYIPLCISLFNLF

NOV38a  YWGVYV
NOV38b  YWGVYV
NOV38c  YWGVYV
NOV38d  YWGVYV
NOV38e  YWGVYV

NOV38a  (SEQ ID NO:  464)
NOV38b  (SEQ ID NO:  466)
NOV38c  (SEQ ID NO:  468)
NOV38d  (SEQ ID NO:  470)
NOV38e  (SEQ ID NO:  472)
```

Further analysis of the NOV38a protein yielded the following properties shown in Table 38C.

Table 38C. Protein Sequence Properties NOV38a

SignalP analysis:	Cleavage site between residues 25 and 26

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 17;  peak value   9.53
      PSG score:   5.12

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -7.09
      possible cleavage site: between 24 and 25

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood = -4.51   Transmembrane    1 -  17
      INTEGRAL    Likelihood = -0.43   Transmembrane  171 - 187
      INTEGRAL    Likelihood = -3.98   Transmembrane  265 - 281
      INTEGRAL    Likelihood = -0.64   Transmembrane  300 - 316
      INTEGRAL    Likelihood = -3.98   Transmembrane  329 - 345
      INTEGRAL    Likelihood = -0.85   Transmembrane  448 - 464
      PERIPHERAL  Likelihood =  7.48 (at 229)
      ALOM score:  -4.51  (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 8
      Charge difference:  1.5   C( 2.5) - N( 1.0)
      C > N: C-terminal side will be inside

>>>Caution: Inconsistent mtop result with signal peptide
>>> membrane topology: type 3b
```

```
MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  2.39
      Hyd Moment(95):  3.90    G content:       0
      D/E content:     1       S/T content:     8
      Score: -2.44

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 43  QRC|SS

NUCDISC: discrimination of nuclear localization signals
      pat4: KRRK (5) at  422
      pat7: none
      bipartite: RRKSICSPSTDSSRIKR at  407
      bipartite: RKSICSPSTDSSRIKRR at  408
      content of basic residues:  10.3%
      NLS Score:  0.99

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
```

```
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       11.1 %: vacuolar
       11.1 %: Golgi
       11.1 %: vesicles of secretory system
       11.1 %: mitochondrial

>> prediction for CG54683-05 is end (k=9)
```

A search of the NOV38a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 38D.

5

Table 38D. Geneseq Results for NOV38a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV38a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE21956	Human transporter protein - Homo sapiens, 467 aa. [US2002028773-A1, 07-MAR-2002]	1..466 1..467	463/469 (98%) 463/469 (98%)	0.0
AAU04467	Human gamma-amino butyric acid (GABA) receptor protein #1 - Homo sapiens, 467 aa. [WO200153489-A1, 26-JUL-2001]	1..466 1..467	463/469 (98%) 463/469 (98%)	0.0
ABU12089	Novel human gamma aminobutyric acid receptor-like protein #2 - Homo sapiens, 468 aa. [US2002123612-A1, 05-SEP-2002]	1..466 1..468	461/470 (98%) 462/470 (98%)	0.0
AAG68256	Human POLY3 protein sequence SEQ ID NO:6 - Homo sapiens, 468 aa. [WO200179294-A2, 25-OCT-2001]	1..466 1..468	461/470 (98%) 462/470 (98%)	0.0
ABU12090	Novel human gamma aminobutyric acid receptor-like protein #2 - Homo sapiens, 466 aa. [US2002123612-A1, 05-SEP-2002]	1..466 1..466	461/469 (98%) 462/469 (98%)	0.0

In a BLAST search of public sequence databases, the NOV38a protein was found to have homology to the proteins shown in the BLASTP data in Table 38E.

Table 38E. Public BLASTP Results for NOV38a

Protein Accession Number	Protein/Organism/Length	NOV38a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P50573	Gamma-aminobutyric-acid receptor rho-3 subunit precursor (GABA(A) receptor) - Rattus norvegicus (Rat), 464 aa.	1..466 1..464	392/470 (83%) 417/470 (88%)	0.0
Q9YGQ2	Gamma-aminobutyric-acid receptor rho-3 subunit - Morone americana (White perch), 470 aa.	1..465 4..469	300/478 (62%) 369/478 (76%)	e-161
P50572	Gamma-aminobutyric-acid receptor rho-1 subunit precursor (GABA(A) receptor) - Rattus norvegicus (Rat), 474 aa.	39..465 50..473	278/428 (64%) 326/428 (75%)	e-155
P56475	Gamma-aminobutyric-acid receptor rho-1 subunit precursor (GABA(A) receptor) - Mus musculus (Mouse), 474 aa.	39..465 50..473	278/428 (64%) 326/428 (75%)	e-154
P24046	Gamma-aminobutyric-acid receptor rho-1 subunit precursor (GABA(A) receptor) - Homo sapiens (Human), 473 aa.	39..465 49..472	276/428 (64%) 325/428 (75%)	e-154

PFam analysis indicates that the NOV38a protein contains the domains shown in the Table 38F.

Table 38F. Domain Analysis of NOV38a

Pfam Domain	NOV38a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Neur_chan_LBD	58..265	74/253 (29%) 176/253 (70%)	1.6e-65
Neur_chan_memb	272..462	45/292 (15%) 145/292 (50%)	9e-37

Example 39.

The NOV39 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 39A.

5

Table 39A. NOV39 Sequence Analysis

NOV39a, CG54692-06 DNA Sequence	SEQ ID NO: 473	1125 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1111

ATGGAGGCCGCTAGCCTTTCAGTGGCCACCGCCGGCGTTGCCCTTGCCCTGGGACCCGAGACCAGCAG
CGGGACCCCAAGCCCGAGAGGGATACTCGGTTCGACCCCGAGCGGCGCCGTCCTGCCGGGCCGAGGGC
CGCCCTTCTCTGTCTTCACGGTCCTGGTGGTGACGCTGCTAGTGCTGCTGATCGCTGCCACTTTCCTG
TGGAACCTGCTGGTTCCGGTCACCATCCCGCGGGTCCGTGCCTTCCACCGCGTGCCGCATAACTTGGT
GGCCTCGACGGCCGTCTCGGACGAACTAGTGGCAGCGCTGGCGATGCCACCGAGCCTGGCGAGTGAGC
TGTCGACCGGGCGACGTCGGCTGCTGGGCCGGAGCCTGTGCCACGTGTGGATCTCCTTCGACGCCCTG
TGCTGCCCCGCCGGCCTCGGGAACGTGGCGGCCATCGCCCTGGGCCGCGACGGGGCCATCACACGGCA
CCTGCAGCACACGCTGCGCACCCGCAGCCGCGCCTCGTTGCTCATGATCGCGCTCGCCCGGGTGCCGT
CGGCGCTCATCGCCCTCGCGCCGCTGCTCTTTGGCCGGGGCGAGGTGTGCGACGCTCGGCTCCAGCGC
TGCCAGGTGAGCCGGGAACCCTCCTATGCCGCCTTCTCCACCCGCGGCGCCTTCCACCTGCCGCTTGG
CGTGGTGCCGTTTGTCTACCGGAAGATCTACGAGGCGGCCAAGTTTCGTTTCGGCCGCCGCCGGAGAG
CTGTGCTGCCGTTGCCGGCCACCATGCAGGTGAAGGAAGCACCTGATGAGGCTGAAGTGGTGTTCACG
GCACATTGCAAAGCAACGGTGTCCTTCCAGGTGAGCGGGGACTCCTGGCGGGAGCAGAAGGAGAGGCG
AGCAGCCATGATGGTGGGAATTCTGATTGGCGTGTTTGTGCTGTGCTGGATCCCCTTCTTCCTGACGG
AACTCATCAGCCCACTCTGTGCCTGCAGCCTGCCCCCCATCTGGAAAAGCATATTTCTGTGGCTTGGC
TACTCCAATTCTTTCTTCAACCCCCTGATTTACACAGCTTTTAACAAGAACTACAACAATGCCTTCAA
GAGCCTCTTTACTAAGCAGAGA**TGA**ACACAGGGGTTA

NOV39a, CG54692-06 Protein Sequence	SEQ ID NO: 474	370 aa	MW at 40243.7kD

MEAASLSVATAGVALALGPETSSGTPSPRGILGSTPSGAVLPGRGPPFSVFTVLVVTLLVLLIAATFL
WNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMPPSLASELSTGRRRLLGRSLCHVWISFDAL
CCPAGLGNVAAIALGRDGAITRHLQHTLRTRSRASLLMIALARVPSALIALAPLLFGRGEVCDARLQR
CQVSREPSYAAFSTRGAFHLPLGVVPFVYRKIYEAAKFRFGRRRRAVLPLPATMQVKEAPDEAEVVFT
AHCKATVSFQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPPIWKSIFLWLG
YSNSFFNPLIYTAFNKNYNNAFKSLFTKQR

NOV39b, CG54692-01 DNA Sequence	SEQ ID NO: 475	1150 bp
	ORF Start: ATG at 24	ORF Stop: TGA at 1134

CTGGAGCTGCGATCCCAAGCGCC**ATG**GAGGCCGCTAGCCTTTCAGTGGCCACCGCCGGCGTTGCCCTT
GCCCTGGGACCCGAGACCAGCAGCGGGACCCCAAGCCCGAGAGGGATACTCGGTTCGACCCCGAGCGG
CGCCGTCCTGCCGGGCCGAGGGCCGCCCTTCTCTGTCTTCACGGTCCTGGTGGTGACGCTGCTAGTGC
TGCTGATCGCTGCCACTTTCCTGTGGAACCTGCTGGTTCCGGTCACCATCCCGCGGGTCCGTGCCTTC
CACCGCGTGCCGCATAACTTGGTGGCCTCGACGGCCGTCTCGGACGAACTAGTGGCAGCGCTGGCGAT
GCCACCGAGCCTGGCGAGTGAGCTGTCGACCGGGCGACGTCGGCTGCTGGGCCGGAGCCTGTGCCACG
TGTGGATCTCCTTCGACGCCCTGTGCTGCCCCGCCGGCCTCGGGAACGTGGCGGCCATCGCCCTGGGC
CGCGACGGGGCCATCACACGGCACCTGCAGCACACGCTGCGCACCCGCAGCCGCGCCTCGTTGCTCAT
GATCGCGCTCGCCCGGGTGCCGTCGGCGCTCATCGCCCTCGCGCCGCTGCTCTTTGGCCGGGGCGAGG
TGTGCGACGCTCGGCTCCAGCGCTGCCAGGTGAGCCGGGAACCCTCCTATGCCGCCTTCTCCACCCGC
GGCGCCTTCCACCTGCCGCTTGGCGTGGTGCCGTTTGTCTACCGGAAGATCTACGAGGCGGCCAAGTT
TCGTTTCGGCCGCCGCCGGAGAGCTGTGCTGCCGTTGCCGGCCACCTCCAAGGTAAAGGAAGCACCTG
ATGAGGCTGAAGTGGTGTTCACGGCACATTGCAAAGCAACGGTGTCCTTCCAGGTGAGCGGGGACTCC
TGGCGGGAGCAGAAGGAGAGGCGAGCAGCCATGATGGTGGGAATTCTGATTGGCGTGTTTGTGCTGTG
CTGGATCCCCTTCTTCCTGACGGAACTCATCAGCCCACTCTGTGCCTGCAGCCTGCCCCCCATCTGGA
AAAGCATATTTCTGTGGCTTGGCTACTCCAATTCTTTCTTCAACCCCCTGATTTACACAGCTTTTAAC
AAGAACTACAACAATGCCTTCAAGAGCCTCTTTACTAAGCAGAGA**TGA**ACACAGGGGTTAGA

NOV39b, CG54692-01 Protein Sequence	SEQ ID NO: 476	370 aa	MW at 40199.7kD

MEAASLSVATAGVALALGPETSSGTPSPRGILGSTPSGAVLPGRGPPFSVFTVLVVTLLVLLIAATFL
WNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMPPSLASELSTGRRRLLGRSLCHVWISFDAL
CCPAGLGNVAAIALGRDGAITRHLQHTLRTRSRASLLMIALARVPSALIALAPLLFGRGEVCDARLQR
CQVSREPSYAAFSTRGAFHLPLGVVPFVYRKIYEAAKFRFGRRRRAVLPLPATSKVKEAPDEAEVVFT
AHCKATVSFQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPPIWKSIFLWLG
YSNSFFNPLIYTAFNKNYNNAFKSLFTKQR

NOV39c, CG54692-02 DNA Sequence	SEQ ID NO: 477	1150 bp
	ORF Start: ATG at 24	ORF Stop: TGA at 1134

CTGGAGCTGCGATCCCAAGCGCC**ATG**GAGGCCGCTAGCCTTTCAGTGGCCACCGCCGGCGTTGCCCTT
GCCCTGGGACCCGAGACCAGCAGCGGGACCCCAAGCCCGAGAGGGATACTCGGTTCGACCCCGAGCGG
CGCCGTCCTGCCGGGCCGAGGGCCGCCCTTCTCTGTCTTCACGGTCCTGGTGGTGACGCTGCTAGTGC
TGCTGATCGCTGCCACTTTCCTGTGGAACCTGCTGGTTCCGGTCACCATCCCGCGGGTCCGTGCCTTC
CACCGCGTGCCGCATAACTTGGTGGCCTCGACGGCCGTCTCGGACGAACTAGTGGCAGCGCTGGCGAT
GCCACCGAGCCTGGCGAGTGAGCTGTCGACCGGGCGACGTCGGCTGCTGGGCCGGAGCCTGTGCCACG
TGTGGATCTCCTTCGACGCCCTGTGCTGCCCCGCCGGCCTCGGGAACGTGGCGGCCATCGCCCTGGGC
CGCGACGGGGCCATCACACGGCACCTGCAGCACACGCTGCGCACCCGCAGCCGCGCCTCGTTGCTCAT
GATCGCGCTCGCCCGGGTGCCGTCGGCGCTCATCGCCCTCGCGCCGCTGCTCTTTGGCCGGGGCGAGG
TGTGCGACGCTCGGCTCCAGCGCTGCCAGGTGAGCCGGGAACCCTCCTATGCCGCCTTCTCCACCCGC
GGCGCCTTCCACCTGCCGCTTGGCGTGGCGCCGTTTGTCTACCGGAAGATCTACGAGGCGGCCAAGTT
TCGTTTCGGCCGCCGCCGGAGAGCTGTGCTGCCGTTGCCGGCCACCATGCAAGTAAAGGAAGCACCTG
ATGAGGCTGAAGTGGTGTTCACGGCACATTGCAAAGCAACGGTGTCCTTCCAGGTGAGCGGGGACTCC
TGGCGGGAGCAGAAGGAGAGGCGAGCAGCCATGATGGTGGGAATTCTGATTGGCGTGTTTGTGCTGTG
CTGGATCCCCTTCTTCCTGACGGAACTCATCAGCCCACTCTGTGCCTGCAGCCTGCCCCCCATCTGGA
AAAGCATATTTCTGTGGCTTGGCTACTCCAATTCTTTCTTCAACCCCCTGATTTACACAGCTTTTAAC
AAGAACTACAACAATGCCTTCAAGAGCCTCTTTACTAAGCAGAGA**TGA**ACACAGGGGTTAGA

NOV39c, CG54692-02 Protein Sequence	SEQ ID NO: 478	370 aa	MW at 40215.7kD

MEAASLSVATAGVALALGPETSSGTPSPRGILGSTPSGAVLPGRGPPFSVFTVLVVTLLVLLIAATFL
WNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMPPSLASELSTGRRRLLGRSLCHVWISFDAL
CCPAGLGNVAAIALGRDGAITRHLQHTLRTRSRASLLMIALARVPSALIALAPLLFGRGEVCDARLQR
CQVSREPSYAAFSTRGAFHLPLGVAPFVYRKIYEAAKFRFGRRRRAVLPLPATMQVKEAPDEAEVVFT
AHCKATVSFQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPPIWKSIFLWLG
YSNSFFNPLIYTAFNKNYNNAFKSLFTKQR

NOV39d, CG54692-03 DNA Sequence	SEQ ID NO: 479	1127 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1117

ATGGAGGCCGCTAGCCTTTCAGTGGCCACCGCCGGCGTTGCCCTTGCCCTGGGACCCGAGACCAGCAG
CGGGACCCCAAGCCCGAGAGGGATACTCGGTTCGACCCCGAGCGGCGCCGTCCTGCCGGGCCGAGGGC
CGCCCTTCTCTGTCTTCACGGTCCTGGTGGTGACGCTGCTAGTGCTGCTGATCGCCGCCACTTTCCTG
TGGAACCTGCTGGTTCCGGTCACCATCCCGCGGGTCCGTGCCTTCCACCGCGTGCCGCATAACTTGGT
GGCCTCGACGGCCGTCTCGGACGAACTAGTGGCAGCGCTGGCGATGCCACCGAGCCTGGCGAGTGAGC
TGTCGACCGGGCGACGTCGGCTGCTGGGCCGGAGCCTGTGCCACGTGTGGATCTCCTTCCACGGCCCA
CGGCTGTGCTGCCCCGCCGGCCTCGGGAACGTGGCGGCCATCGCCCTGGGCCGCGACGGGGCCATCAC
ACGGCACCTGCAGCACACGCTGCGCACCCGCAGCCGCGCCTCGTTGCTCATGATCGCGCTCACCCGGG
TGCCGTCGGCGCTCATCGCCCTCGCGCCGCTGCTCTTTGGCCGGGGCGAGGTGTGCGACGCTCGGCTC
CAGCGCTGCCAGGTGAGCCGGGAACCCTCCTATGCCGCCTTCTCCACCCGCGGCGCCTTCCACCTGCC
GCTTGGCGTGGTGCCGTTTGTCTACCGGAAGATCTACGAGGCGGCCAAGTTTCGTTTCGGCCGCCGCC
GGAGAGCTGTGCTGCCGTTGCCGGCCACCATGCAGGTGAAGGAAGCACCTGATGAGGCTGAAGTGGTG
TTCACGGCACATTGCAAAGCAACGGTGTCCTTCCAGGTGAGCGGGGACTCCTGGCGGGAGCAGAAGGA
GAGGCGAGCAGCCATGATGGTGGGAATTCTGATTGGCGTGTTTGTGCTGTGCTGGATCCCCTTCTTCC
TGACGGAACTCATCAGCCCACTCTGTGCCTGCAGCCTGCCCCCCATCTGGAAAAGCATATTTCTGTGG
CTTGGCTACTCCAATTCTTTCTTCAACCCCCTGATTTACACAGCTTTTAACAAGAACTACAACAATGC
CTTCAAGAGCCTCTTTACTAAGCAGAGA**TGA**ACACAGGG

NOV39d, CG54692-03 Protein Sequence	SEQ ID NO: 480	372 aa	MW at 40535.1kD

MEAASLSVATAGVALALGPETSSGTPSPRGILGSTPSGAVLPGRGPPFSVFTVLVVTLLVLLIAATFL
WNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMPPSLASELSTGRRRLLGRSLCHVWISFHGP
RLCCPAGLGNVAAIALGRDGAITRHLQHTLRTRSRASLLMIALTRVPSALIALAPLLFGRGEVCDARL
QRCQVSREPSYAAFSTRGAFHLPLGVVPFVYRKIYEAAKFRFGRRRRAVLPLPATMQVKEAPDEAEVV
FTAHCKATVSFQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPPIWKSIFLW
LGYSNSFFNPLIYTAFNKNYNNAFKSLFTKQR

NOV39e, CG54692-04 DNA Sequence	SEQ ID NO: 481	1155 bp
	ORF Start: ATG at 5	ORF Stop: TGA at 1145

CGCCATGGAGGCCGCTAGCCTTTCAGTGGCCACCGCCGGCGTTGCCCTTGCCCCCGAGACCAGCAGCC
CGGCGTTGCCCTTGCCCTGGGACCCGAGACCAGCAGCAGGACCCGGGACCCCAAGCCCGAGAGGGATA
CTCGGTTCGACCCCGAGCGGCGCCGTCCTGCCGGGCCGAGGGCCGCCCTTCTCTGTCTTCACGGTCCT
GGTGGTGACGCTGCTAGTGCTGCTGATCGCTGCCACTTTCCTGTGGAACCTGCTGGTTCCGGTCACCA
TCCCGCGGGTCCGTGCCTTCCACCGCGTGCCGCATAACTTGGTGGCCTCGACGGCCGTCTCGGACGAA
CTAGTGGCAGCGCTGGCGATGCCACCGAGCCTGGCGAGTGAGCTGTCGACCGGGCGACGTCGGCTGCT
GGGCCGCCACGTGTGGATCTCCTTCGACGCCCTGTGCTGCCCCGCCGGCCTCGGGAACGTGGCGGCCA
TCGCCCTGGGCCGCGACGGGGCCATCACACGGCACCTGCAGCACACGCTGCGCACCCGCAGCCGCGCC
TCGTTGCTCATGATCGCGCTCGCCCGGGTGCCGTCGGCGCTCATCGCCCTCGCGCCGCTGCTCTTTGG
CCGGGGCGAGGTGTGCGACGCTCGGCTCCAGCGCTGCCAGGTGAGCCGGGAACCCTCCTATGCCGCCT
TCTCCACCCGCGGCGCCTTCCACCTGCCGCTTGGCGTGGTGCCGTTTGTCTACCGGAAGATCTACGAG
GCGGCCAAGTTTCGTTTCGGCCGCCGCCGGAGAGCTGTGCTGCCGTTGCCGGCCACCATGCAGGTGAA
GGAAGCACCTGATGAGGCTGAAGTGGTGTTCACGGCACATTGCAAAGCAACGGTGTCCTTCCAGGTGA
GCGGGGACTCCTGGCGGGAGCAGAAGGAGAGGCGAGCAGCCATGATGGTGGGAATTCTGATTGGCGTG
TTTGTGCTGTGCTGGATCCCCTTCTTCCTGACGGAACTCATCAGCCCACTCTGTGCCTGCAGCCTGCC
CCCCATCTGGAAAAGCATATTTCTGTGGCTTGGCTACTCCAATTCTTTCTTCAACCCCCTGATTTACA
CAGCTTTTAACAAGAACTACAACAATGCCTTCAAGAGCCTCTTTACTAAGCAGAGATGAACACAGGG

NOV39e, CG54692-04 Protein Sequence	SEQ ID NO: 482	380 aa	MW at 41306.9kD

MEAASLSVATAGVALAPETSSPALPLPWDPRPAAGPGTPSPRGILGSTPSGAVLPGRGPPFSVFTVLV
VTLLVLLIAATFLWNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMPPSLASELSTGRRRLLG
RHVWISFDALCCPAGLGNVAAIALGRDGAITRHLQHTLRTRSRASLLMIALARVPSALIALAPLLFGR
GEVCDARLQRCQVSREPSYAAFSTRGAFHLPLGVVPFVYRKIYEAAKFRFGRRRRAVLPLPATMQVKE
APDEAEVVFTAHCKATVSFQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPP
IWKSIFLWLGYSNSFFNPLIYTAFNKNYNNAFKSLFTKQR

NOV39f, CG54692-05 DNA Sequence	SEQ ID NO: 483	1152 bp
	ORF Start: ATG at 5	ORF Stop: TGA at 1142

CGCCATGGAGGCCGCTAGCCTTTCAGTGGCCACCGCCGGCGTTGCCCTTGCCCTGGGACCCGAGACCA
GCAGCGGACCCGGGACCCCAAGCCCGAGAGGGATACTCGGTTCGACCCCGAGCGGCGCCGTCCTGCCG
GGCCGAGGGCCGCCCTTCTCTGTCTTCACGGTCCTGGTGGTGACGCTGCTAGTGCTGCTGATCGCTGC
CACTTTCCTGTGGAACCTGCTGGTTCCGGTCACCATCCCGCGGGTCCGTGCCTTCCACCGCGTGCCGC
ATAACTTGGTGGCCTCGACGGCCGTCTCGGACGAACTAGTGGCAGCGCTGGCGATGCCACCGAGCCTG
GCGAGTGAGCTGTCGACCGGGCGACGTCGGCTGCTGGGCCGGAGCCTGTGCCACGTGTGGATCTCCTT
CGACGCCGGAGCCTGTCTGTGCTGCCCCGCCGGCCTCGGGAACGTGGCGGCCATCGCCCTGGGCCGCG
ACGGGGCCATCACACGGCACCTGCAGCACACGCTGCGCACCCGCAGCCGCGCCTCGTTGCTCATGATC
GCGCTCGCCCGGGTGCCGTCGGCGCTCATCGCCCTCGCGCCGCTGCTCTTTGGCCGGGGCGAGGTGTG
CGACGCTCGGCTCCAGCGCTGCCAGGTGAGCCGGGAACCCTCCTATGCCGCCTTCTCCACCCGCGGCG
CCTTCCACCTGCCGCTTGGCGTGGTGCCGTTTGTCTACCGGAAGATCTACGAGGCGGCCAAGTTTCGT
TTCGGCCGCCGCCGGAGAGCTGTGCTGCCGTTGCCGGCCACCATGCAGGTGAGGTCCAAGGTAAAGGA
AGCACCTGATGAGGCTGAAGTGGTGTTCACGGCACATTGCAAAGCAACGGTGTCCTTCCAGGTGAGCG
GGGACTCCTGGCGGGAGCAGAAGGAGAGGCGAGCAGCCATGATGGTGGGAATTCTGATTGGCGTGTTT
GTGCTGTGCTGGATCCCCTTCTTCCTGACGGAACTCATCAGCCCACTCTGTGCCTGCAGCCTGCCCCC
CATCTGGAAAAGCATATTTCTGTGGCTTGGCTACTCCAATTCTTTCTTCAACCCCCTGATTTACACAG
CTTTTAACAAGAACTACAACAATGCCTTCAAGAGCCTCTTTACTAAGCAGAGATGAACACAGGG

NOV39f, CG54692-05 Protein Sequence	SEQ ID NO: 484	379 aa	MW at 41099.7kD

MEAASLSVATAGVALALGPETSSGPGTPSPRGILGSTPSGAVLPGRGPPFSVFTVLVVTLLVLLIAAT
FLWNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMPPSLASELSTGRRRLLGRSLCHVWISFD
AGACLCCPAGLGNVAAIALGRDGAITRHLQHTLRTRSRASLLMIALARVPSALIALAPLLFGRGEVCD
ARLQRCQVSREPSYAAFSTRGAFHLPLGVVPFVYRKIYEAAKFRFGRRRRAVLPLPATMQVRSKVKEA
PDEAEVVFTAHCKATVSFQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPPI
WKSIFLWLGYSNSFFNPLIYTAFNKNYNNAFKSLFTKQR

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 39B.

Table 39B. Comparison of the NOV39 protein sequences.

```
NOV39a  MEAASLSVATAGVALA----------LGPETSSG--TPSPRGILGSTPSGAVLPGRGPP
NOV39b  MEAASLSVATAGVALA----------LGPETSSG--TPSPRGILGSTPSGAVLPGRGPP
NOV39c  MEAASLSVATAGVALA----------LGPETSSG--TPSPRGILGSTPSGAVLPGRGPP
NOV39d  MEAASLSVATAGVALA----------LGPETSSG--TPSPRGILGSTPSGAVLPGRGPP
NOV39e  MEAASLSVATAGVALAPETSSPALPLPWDPRPAAGPGTPSPRGILGSTPSGAVLPGRGPP
NOV39f  MEAASLSVATAGVALA----------LGPETSSGPGTPSPRGILGSTPSGAVLPGRGPP

NOV39a  FSVFTVLVVTLLVLLIAATFLWNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMP
NOV39b  FSVFTVLVVTLLVLLIAATFLWNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMP
NOV39c  FSVFTVLVVTLLVLLIAATFLWNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMP
NOV39d  FSVFTVLVVTLLVLLIAATFLWNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMP
NOV39e  FSVFTVLVVTLLVLLIAATFLWNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMP
NOV39f  FSVFTVLVVTLLVLLIAATFLWNLLVPVTIPRVRAFHRVPHNLVASTAVSDELVAALAMP

NOV39a  PSLASELSTGRRRLLGRSLCHVWISFDA---LCCPAGLGNVAAIALGRDGAITRHLQHTL
NOV39b  PSLASELSTGRRRLLGRSLCHVWISFDA---LCCPAGLGNVAAIALGRDGAITRHLQHTL
NOV39c  PSLASELSTGRRRLLGRSLCHVWISFDA---LCCPAGLGNVAAIALGRDGAITRHLQHTL
NOV39d  PSLASELSTGRRRLLGRSLCHVWISFHG-PRLCCPAGLGNVAAIALGRDGAITRHLQHTL
NOV39e  PSLASELSTGRRRLLGR---HVWISFDA---LCCPAGLGNVAAIALGRDGAITRHLQHTL
NOV39f  PSLASELSTGRRRLLGRSLCHVWISFDAGACLCCPAGLGNVAAIALGRDGAITRHLQHTL

NOV39a  RTRSRASLLMIALARVPSALIALAPLLFGRGEVCDARLQRCQVSREPSYAAFSTRGAFHL
NOV39b  RTRSRASLLMIALARVPSALIALAPLLFGRGEVCDARLQRCQVSREPSYAAFSTRGAFHL
NOV39c  RTRSRASLLMIALARVPSALIALAPLLFGRGEVCDARLQRCQVSREPSYAAFSTRGAFHL
NOV39d  RTRSRASLLMIALTRVPSALIALAPLLFGRGEVCDARLQRCQVSREPSYAAFSTRGAFHL
NOV39e  RTRSRASLLMIALARVPSALIALAPLLFGRGEVCDARLQRCQVSREPSYAAFSTRGAFHL
NOV39f  RTRSRASLLMIALARVPSALIALAPLLFGRGEVCDARLQRCQVSREPSYAAFSTRGAFHL

NOV39a  PLGVVPFVYRKIYEAAKFRFGRRRRAVLPLPAT----MQVKEAPDEAEVVFTAHCKATVS
NOV39b  PLGVVPFVYRKIYEAAKFRFGRRRRAVLPLPAT----SKVKEAPDEAEVVFTAHCKATVS
NOV39c  PLGVAPFVYRKIYEAAKFRFGRRRRAVLPLPAT----MQVKEAPDEAEVVFTAHCKATVS
NOV39d  PLGVVPFVYRKIYEAAKFRFGRRRRAVLPLPAT----MQVKEAPDEAEVVFTAHCKATVS
NOV39e  PLGVVPFVYRKIYEAAKFRFGRRRRAVLPLPAT----MQVKEAPDEAEVVFTAHCKATVS
NOV39f  PLGVVPFVYRKIYEAAKFRFGRRRRAVLPLPATMQVRSKVKEAPDEAEVVFTAHCKATVS

NOV39a  FQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPPIWKSIFLWLG
NOV39b  FQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPPIWKSIFLWLG
NOV39c  FQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPPIWKSIFLWLG
NOV39d  FQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPPIWKSIFLWLG
```

```
NOV39e  FQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPPIWKSIFLWLG
NOV39f  FQVSGDSWREQKERRAAMMVGILIGVFVLCWIPFFLTELISPLCACSLPPIWKSIFLWLG

NOV39a  YSNSFFNPLIYTAFNKNYNNAFKSLFTKQR
NOV39b  YSNSFFNPLIYTAFNKNYNNAFKSLFTKQR
NOV39c  YSNSFFNPLIYTAFNKNYNNAFKSLFTKQR
NOV39d  YSNSFFNPLIYTAFNKNYNNAFKSLFTKQR
NOV39e  YSNSFFNPLIYTAFNKNYNNAFKSLFTKQR
NOV39f  YSNSFFNPLIYTAFNKNYNNAFKSLFTKQR

NOV39a  (SEQ ID NO:  474)
NOV39b  (SEQ ID NO:  476)
NOV39c  (SEQ ID NO:  478)
NOV39d  (SEQ ID NO:  480)
NOV39e  (SEQ ID NO:  482)
NOV39f  (SEQ ID NO:  484)
```

Further analysis of the NOV39a protein yielded the following properties shown in Table 39C.

Table 39C. Protein Sequence Properties NOV39a

SignalP analysis:	Cleavage site between residues 25 and 26

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 2;  pos.chg 0;  neg.chg 1
      H-region:  length 17;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -4.24
      possible cleavage site: between 23 and 24

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   5
      INTEGRAL    Likelihood =  0.32   Transmembrane    1 -  17
      INTEGRAL    Likelihood =-12.10   Transmembrane   48 -  64
      INTEGRAL    Likelihood = -0.48   Transmembrane  135 - 151
      INTEGRAL    Likelihood = -4.94   Transmembrane  172 - 188
      INTEGRAL    Likelihood = -9.66   Transmembrane  300 - 316
      PERIPHERAL  Likelihood =  0.69 (at 90)
      ALOM score: -12.10  (number of TMSs: 5)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 8
      Charge difference:  0.0   C( 0.0) - N( 0.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a
```

```
MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  3.56
      Hyd Moment(95):  5.35     G content:       2
      D/E content:     2        S/T content:     3
      Score: -7.53

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 95  HRV|PH

NUCDISC: discrimination of nuclear localization signals
      pat4: RRRR (5) at  246
      pat7: none
      bipartite: RKIYEAAKFRFGRRRRA at  234
      content of basic residues:  10.5%
      NLS Score:  0.50

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: FTKQ

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1
```

```
COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       22.2 %: mitochondrial
       11.1 %: nuclear

>> prediction for CG54692-06 is end (k=9)
```

A search of the NOV39a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 39D.

Table 39D. Geneseq Results for NOV39a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE18654	Human G-protein coupled receptor (GCREC-15) - Homo sapiens, 370 aa. [WO200210387-A2, 07-FEB-2002]	1..370 1..370	370/370 (100%) 370/370 (100%)	0.0
AAM47212	Human NOV5b protein - Homo sapiens, 370 aa. [WO200174851-A2, 11-OCT-2001]	1..370 1..370	369/370 (99%) 369/370 (99%)	0.0
AAM47211	Human NOV5a protein - Homo sapiens, 370 aa. [WO200174851-A2, 11-OCT-2001]	1..370 1..370	368/370 (99%) 369/370 (99%)	0.0
AAE15638	Human G-protein coupled receptor-8 (GCREC-8) protein - Homo sapiens, 372 aa. [WO200198351-A2, 27-DEC-2001]	1..370 1..372	367/372 (98%) 367/372 (98%)	0.0
ABB78809	Human NOV5 protein sequence SEQ ID NO:16 - Homo sapiens, 379 aa. [WO200230974-A2, 18-APR-2002]	1..370 1..379	370/379 (97%) 370/379 (97%)	0.0

In a BLAST search of public sequence databases, the NOV39a protein was found to have homology to the proteins shown in the BLASTP data in Table 39E.

Table 39E. Public BLASTP Results for NOV39a

Protein Accession Number	Protein/Organism/Length	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD13088	Sequence 11 from Patent WO0174851 - Homo sapiens (Human), 370 aa.	1..370 1..370	369/370 (99%) 369/370 (99%)	0.0
CAD13087	Sequence 9 from Patent WO0174851 - Homo sapiens (Human), 370 aa.	1..370 1..370	368/370 (99%) 369/370 (99%)	0.0
P35365	5-hydroxytryptamine 5B receptor (5-HT-5B) (Serotonin receptor) (MR22) - Rattus norvegicus (Rat), 370 aa.	1..370 1..370	296/370 (80%) 316/370 (85%)	e-165
P31387	5-hydroxytryptamine 5B receptor (5-HT-5B) (Serotonin receptor) - Mus musculus (Mouse), 370 aa.	1..370 1..370	298/370 (80%) 317/370 (85%)	e-165
S38744	serotonin receptor 5B - rat, 369 aa.	1..370 1..369	296/370 (80%) 316/370 (85%)	e-164

PFam analysis indicates that the NOV39a protein contains the domains shown in the Table 39F.

Table 39F. Domain Analysis of NOV39a

Pfam Domain	NOV39a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1	69..351	75/314 (24%) 197/314 (63%)	1.2e-43

Example 40.

The NOV40 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 40A.

Table 40A. NOV40 Sequence Analysis

NOV40a, CG55069-01 DNA Sequence	SEQ ID NO: 485	8657 bp
	ORF Start: ATG at 151	ORF Stop: TAA at 8326

TTTGGCCTCGGGCCAGAATTCGGCACGAGGGGTCTGGAGCTTGGAGGAGAAGTCTGAACTAAGGATAA
ACTAAAGAGAGGCCAATGAGACTTGAACCCTGAGCCTAAGTTGTCACCAGCAGGACTGATGTGCACAC
AGAAGGAATGAAGTATGGATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAG
AAGGAACGGCGCTACACAAATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCACACAGAAGTCCTA
CAGTTCCAGCGAGACATTGAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGA
AGGATTTGGTTCACAGAGAAGCAGACGAGTTCACTAGACAAGGACAGAATTTTACCCTAAGGCAGTTA
GGAGTTTGTGAACCAGCAACTCGAAGAGGACTGGCATTTTGTGCGGAAATGGGGCTCCCTCACAGAGG
TTACTCTATCAGTGCAGGGTCAGATGCTGATACTGAAAATGAAGCAGTGATGTCCCCAGAGCATGCCA
TGAGACTTTGGGGCAGGGGGGTCAAATCAGGCCGCAGCTCCTGCCTGTCAAGTCGGTCCAACTCAGCC
CTCACCCTGACAGATACGGAGCACGAAAACAAGTCCGACAGTGAGAATGAGCAACCTGCAAGCAATCA
AGGCCAGTCTACCCTGCAGCCCTTGCCGCCTTCCCATAAGCAGCACTCTGCACAGCATCATCCATCCA
TCACTTCTCTCAACAGAAACTCCCTGACCAATAGAAGGAACCAGAGTCCGGCCCCGCCGGCTGCTTTG
CCCGCCGAGCTGCAAACCACACCCGAGTCCGTCCAGCTGCAGGACAGCTGGGTCCTTGGCAGTAATGT
ACCACTGGAAAGCAGGCATTTCCTATTCAAAACAGGAACAGGTACAACGCCACTGTTCAGTACTGCAA
CCCCAGGATACACAATGGCATCTGGCTCTGTTTATTCACCACCTACTCGGCCACTACCTAGAAACACC
CTATCAAGAAGTGCTTTTAAATTCAAGAAGTCTTCAAAGTACTGTAGCTGGAAATGCACTGCACTGTG
TGCCGTAGGGGTCTCGGTGCTCCTGGCAATACTCCTGTCTTATTTTATAGCAATGCATCTCTTTGGCC
TCAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATACCATG
CCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAAATAA
CACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATCTTCT
GGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGATGCA
TTGATTGGAGTATATGGCCGGAAGAAGTTACCGCCTTCCCATACTCAGTCCTCCCCCCAGTATGACTT
CGTGGAGCTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAG
CCGGGCGGCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATC
TGGCATCTGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGA
GTCTGTGGTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTT
TTCCAGGATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAG
TACTCCAAGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCA
GTGTATTGACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAGCTCAGGAT
ACAAAGGAGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATC
CACGGGGAATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCC
AGACCAGTGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGA
CTGGCCCAGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGG
ACGTGTCGCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGC
CGAGCACGGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTA
TCGCTCACTATTTGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGC
AACAGCAATGGAAGATGTACCCTGGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGG
AGCAGGCTGTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGGGATGGACTCA
TTGACTGCATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTG
CCGGATCCTCAGGACATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTA
TGATCGAATCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATA
AGAGCCTTGCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAAT
GTCTCGTTTTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGT
GGCAAATGGTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTG
TGTGGATTCCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGGAGAAAGAAGAGAATGACATT
CCCAGCTGTGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTT
TTTCAGATCTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAA
TTCCAGGAACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAG

ATCACCATGACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGG

AAGACTCTTCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAG
ATGCATATAATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCG
TGTTTGGACCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAA
CATGGGTGGCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAA
ACGGGGAAAACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGG
CGCAGCATTTCCTGCCCCAGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCT
AGCTTGTGGGATCGATGGCAGTCTGTACGTAGGCGATTTCAACTACGTGCGGCGGATATTCCCTTCTG
GAAATGTAACAAGTGTCTTAGAACTAAGAAATAAAGATTTTAGACATAGCAGCAACCCAGCTCATAGA
TACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTACGTTTCTGACACAAACACCCGCAGAATTTA
TCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAAAAATGCAGAAGTCGTCGCAGGGACAGGGG
AGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATGGAGGGAAGGCCGTGGAAGCCACACTCATG
AGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATCTACTTTGTTGATGGAACCATGATTAGGAA
AGTTGACCAAAATGGAATCATATCAACTCTTCTGGGCTCTAACGATTTGACTTCAGCCAGACCTTTAA
CTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGGAATGGCCCACTGACCTAGCCATTAACCCT
ATGGATAACTCCATTTATGTCCTGGATAATAATGTAGTTTTACAGATCACTGAAAATCGTCAAGTTCG
CATTGCTGCTGGACGGCCCATGCACTGTCAGGTTCCCGGAGTGGAATATCCTGTGGGGAAGCACGCGG
TGCAGACAACACTGGAATCAGCCACTGCCATTGCTGTGTCCTACAGTGGGGTCCTGTACATTACTGAA
ACTGATGAGAAGAAAATTAACCGGATAAGGCAGGTCACAACAGATGGAGAAATCTCCTTAGTGGCCGG
AATACCTTCAGAGTGTGACTGCAAAAATGATGCCAACTGTGACTGTTACCAGAGTGGAGATGGCTACG
CCAAGGATGCCAAACTCAGTGCCCCATCCTCCCTGGCTGCTTCTCCAGATGGTACACTGTATATTGCA
GATCTAGGGAATATCCGGATCCGGGCTGTGTCAAAGAATAAGCCTTTACTTAACTCTATGAACTTCTA
TGAAGTTGCGTCTCCAACTGATCAAGAACTCTACATCTTTGACATCAATGGTACTCACCAATATACTG
TAAGTTTAGTCACTGGTGATTACCTTTACAATTTTAGCTACAGCAATGACAATGATATTACTGCTGTG
ACAGACAGCAATGGCAACACCCTTAGAATTAGACGGGACCCAAATCGCATGCCAGTTCGAGTGGTGTC
TCCTGATAACCAAGTGATATGGTTGACAATAGGAACAAATGGATGTTTGAAAGGCATGACTGCTCAAG
GACTGGAATTAGTTTTGTTTACTTACCATGGCAATAGTGGCCTTTTAGCCACTAAAAGTGATGAAACT
GGATGGACAACGTTTTTTGACTATGACAGTGAAGGTCGTCTGACAAATGTTACGTTTCCAACTGGAGT
GGTCACAAACCTGCATGGGGACATGGACAAGGCTATCACAGTGGACATTGAGTCATCTAGCCGAGAAG
AAGATGTCAGCATCACTTCAAATCTGTCCTCGATCGATTCTTTCTACACCATGGTTCAAGATCAGTTA
AGAAACAGCTACCAGATTGGTTATGACGGCTCCCTCAGAATTATCTACGCCAGTGGCCTGGACTCACA
CTACCAAACAGAGCCGCACGTTCTGGCTGGCACCGCTAATCCGACGGTTGCCAAAAGAAACATGACTT
TGCCTGGCGAGAACGGTCAAAACTTGGTGGAATGGAGATTCCGAAAAGAGCAAGCCCAAGGGAAAGTC
AATGTCTTTGGCCGCAAGCTCAGGGTTAATGGCAGAAACCTCCTTTCAGTTGACTTTGATCGAACAAC
AAAGACAGAAAAGATCTATGACGACCACCGTAAATTTCTACTGAGGATCGCCTACGACACGTCTGGGC
ACCCGACTCTCTGGCTGCCAAGCAGCAAGCTGATGGCCGTCAATGTCACCTATTCATCCACAGGTCAA
ATTGCCAGCATCCAGCGAGGCACCACTAGCGAGAAAGTAGATTATGACGGACAGGGGAGGATCGTGTC
TCGGGTCTTTGCTGATGGTAAAACATGGAGTTACACATATTTAGAAAAGTCCATGGTTCTTCTGCTTC
ATAGCCAGCGGCAGTACATCTTCGAATACGATATGTGGGACCGCCTGTCTGCCATCACCATGCCCAGT
GTGGCTCGCCACACCATGCAGACCATCCGATCCATTGGCTACTACCGCAACATATACAACCCCCCGGA
AAGCAACGCCTCCATCATCACGGACTACAACGAGGAAGGGCTGCTTCTACAAACAGCTTTCTTGGGTA
CAAGTCGGAGGGTCTTATTCAAATACAGAAGGCAGACTAGGCTCTCAGAAATTTTATATGATAGCACA
AGAGTCAGTTTTACCTATGATGAAACAGCAGGAGTCCTAAAGACAGTAAACCTCCAGAGTGATGGTTT
TATTTGCACCATTAGATACAGGCAAATTGGTCCCCTGATTGACAGGCAGATTTTCCGCTTTAGTGAAG
ATGGGATGGTAAATGCAAGATTTGACTATAGCTATGACAACAGCTTTCGAGTGACCAGCATGCAGGGT
GTGATCAATGAAACGCCACTGCCTATTGATCTGTATCAGTTTGATGACATTTCTGGCAAAGTTGAGCA
GTTTGGAAAGTTTGGAGTTATATATTATGATATTAACCAGATCATTTCTACAGCTGTAATGACCTATA
CGAAGCACTTTGATGCTCATGGCCGTATCAAGGAGATTCAATATGAGATATTCAGGTCGCTCATGTAC
TGGATTACAATTCAGTATGATAACATGGGTCGGGTAACCAAGAGAGAGATTAAAATAGGGCCCTTTGC
CAACACCACCAAATATGCTTATGAATATGATGTTGATGGACAGCTCCAAACAGTTTACCTCAATGAAA

AGATAATGTGGCGGTACAACTACGATCTGAATGGAAACCTCCATTTACTGAACCCAAGTAACAGTGCG
CGTCTGACACCCCTTCGCTATGACCTGCGAGACAGAATCACTCGACTGGGTGATGTTCAATATCGGTT
GGATGAAGATGGTTTCCTACGTCAAAGGGGCACGGAAATCTTTGAATATAGCTCCAAGGGGCTTCTAA
CTCGAGTTTACAGTAAAGGCAGTGGCTGGACAGTGATCTACCGTTATGACGGCCTGGGAAGGCGTGTT
TCTAGCAAAACCAGTCTAGGACAGCACCTGCAGTTTTTTTATGCTGACTTAACTTATCCCACTAGGAT
TACTCATGTCTACAACCATTCGAGTTCAGAAATTACCTCCCTGTATTATGATCTCCAAGGACATCTTT
TTGCCATGGAAATCAGCAGTGGGGATGAATTCTATATTGCATCGGATAACACAGGGACACCACTGGCT
GTGTTCAGTAGCAATGGGCTTATGCTGAAACAGATTCAGTACACTGCATATGGGGAAATCTATTTTGA
CTCTAATATTGACTTTCAACTGGTAATTGGATTTCATGGTGGCCTGTATGACCCACTCACCAAATTAA

TCCACTTTGGAGAAAGAGATTATGACATTTTGGCAGGACGGTGGACAACACCTGACATAGAAATCTGG
AAAAGAATTGGGAAGGACCCAGCTCCTTTTAACTTGTACATGTTTAGGAATAACAACCCTGCAAGCAA
AATCCATGACGTGAAAGATTACATCACAGATGTTAACAGCTGGCTGGTGACATTTGGTTTCCATCTGC
ACAATGCTATTCCTGGATTCCCTGTTCCCAAATTTGATTTAACAGAACCTTCTTACGAACTTGTGAAG
AGTCAGCAGTGGGATGATATACCGCCCATCTTCGGAGTCCAGCAGCAAGTGGCGCGGCAGGCCAAGGC
CTTCCTGTCGCTGGGGAAGATGGCCGAGGTGCAGGTGAGCCGGCGCCGGGCCGGCGGCGCGCAGTCCT
GGCTGTGGTTCGCCACGGTCAAGTCGCTGATCGGCAAGGGCGTCATGCTGGCCGTCAGCCAGGGCCGC
GTGCAGACCAACGTGCTCAACATCGCCAACGAGGACTGCATCAAGGTGGCGGCCGTGCTCAACAACGC
CTTCTACCTGGAGAACCTGCACTTCACCATCGAGGGCAAGGACACGCACTACTTCATCAAGACCACCA
CGCCCGAGAGCGACCTGGGCACGCTGCGGTTGACCAGCGGCCGCAAGGCGCTGGAGAACGGCATCAAC
GTGACGGTGTCGCAGTCCACCACGGTGGTGAACGGCAGGACGCGCAGGTTCGCGGACGTGGAGATGCA
GTTCGGCGCGCTGGCGCTGCACGTGCGCTACGGCATGACCCTGGACGAGGAGAAGGCGCGCATCCTGG
AGCAGGCGCGGCAGCGCGCGCTCGCCCGGGCCTGGGCGCGCGAGCAGCAGCGCGTGCGCGACGGCGAG
GAGGGCGCGCGCCTCTGGACGGAGGGCGAGAAGCGGCAGCTGCTGAGCGCCGGCAAGGTGCAGGGCTA
CGACGGGTACTACGTACTCTCGGTGGAGCAGTACCCCGAGCTGGCCGACAGCGCCAACAACATCCAGT
TCCTGCGGCAGAGCGAGATCGGCAGGAGG**TAA**CGCCCGGGCCGCGCCCGCCGAGCCGCTCACGCCCTG
CCCACATTGTCCTGTGGCACAACCCGAGTGGGACTCTCCAACGCCCAAGAGCCTTCCTCCCGGGGGAA
TGAGACTGCTGTTACGACCCACACCCACACCGCGAAAACAAGGACCGCTTTTTTCCGAATGACCTTAA
AGGTGATCGGCTTTAACGAATATGTTTACATATGCATAGCGCTGCACTCAGTCGGACTGAACGTAGCC
AGAGGAAAAAAAAAATCATCAAGGACAAAGGCCTCGACCTGTTGCGCTGGGCCGTCTGTTCCTTCTAGG
CACTGTATTTAACTAACTTTA

NOV40a, CG55069-01 Protein Sequence	SEQ ID NO: 486	2725 aa	MW at 303959.6kD

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQGQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWG
RGVKSGRSSCLSSRSNSALTLTDTEHENKSDSENEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLN
RNSLTNRRNQSPAPPAALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYT
MASGSVYSPPTRPLPRNTLSRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQL
QQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQL
FIDQPQFLKFNISLQKDALIGVYGRKKLPPSHTQSSPQYDFVELLDGSRLIAREQRSLLETERAGRQA
RSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFL
GPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACSSGYKGES
CEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYL
DKIVKDKIGYKEGCPGLCNSNGRCTLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMD
PDCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLAS
VIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPW
NVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTD
LKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQ
KVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQ
FISQQPPVVSSIMGNGRRRSISCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTS
VLELRNKDFRHSSNPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLP
FDEARCGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTS
MHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEYPVGKHAVQTTL
ESATAIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCKNDANCDCYQSGDGYAKDAK
LSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFYEVASPTDQELYIFDINGTHQYTVSLVT
GDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPNRMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELV
LFTYHGNSGLLATKSDETGWTTFFDYDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSI
TSNLSSIDSFYTMVQDQLRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGEN
GQNLVEWRFRKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLW
LPSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVLLLHSQRQ
YIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEEGLLLQTAFLGTSRRV
LFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLIDRQIFRFSEDGMVN
ARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISGKVEQFGKFGVIYYYDINQIISTAVMTYTKHFD
AHGRIKEIQYEIFRSLMYWITIQYDNMGRVTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWR
YNYDLNGNLHLLNPSNSARLTPLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYS
KGSGWTVIYRYDGLGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEI
SSGDEFYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIHFGE

RDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLVTFGFHLHNAIP
GFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMAEVQVSRRRAGGAQSWLWFA
TVKSLIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNAFYLENLHFTIEGKDTHYFIKTTTPESD
LGTLRLTSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFGALALHVRYGMTLDEEKARILEQARQ
RALARAWAREQQRVRDGEEGARLWTEGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQS
EIGRR

NOV40b, CG55069-04 DNA Sequence	SEQ ID NO: 487	783 bp
	ORF Start: at 7	ORF Stop: at 778

AAGCTTTGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCCAGG
ATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACTCCA
AGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGTATT
GACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAACTCAGGATACAAAGG
AGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACGGGG
AATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGACCAG
TGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGGCCC
AGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGTGTC
GCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAGCAC
GGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGAGGG
TTGTCCTGGTCTGTGCAACAGCAATGGAAGATGTACCCTGGACCAAAATGGCTGGCATTGTGTGTGCC
AGCCTGGATGGAGAGGAGCAGGCTGTGACGTCGAC

NOV40b, CG55069-04 Protein Sequence	SEQ ID NO: 488	257 aa	MW at 26866.7kD

CPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDP
QCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCS
GHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGT
CKDGKCECSQGWNGEHCTIEGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCD

NOV40c, 248993047 DNA Sequence	SEQ ID NO: 489	2448 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGTACCAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATAC
CATGCCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAA
ATAACACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATC
TTCTGGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGA
TGCATTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTGGAGC
TCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGGGCGG
CAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGCATCT
GGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCTGTGG
TGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCCAGGA
TTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACTCCAA
GGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGTATTG
ACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCTTGTGCTTGCAACTCAGGATACAAAGGA
GAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACGGGGA
ATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGACCAGT
GCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGGCCCA
GACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGTGTCG
CTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAGCACG
GGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGAGGGT
TGTCCTGGTCTGTCCAACAGCAATGGAAGATGTACCCTGGACCAAAATGGCTGGCATTGTGTGTGCCA
GCCTGGATGGAGAGGAGCAGGCTGTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATG
AAGGAGATGGACTCATTGACTGCATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCC
TATTGTCGGGGACTGCCGGATCCTCAGGACATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGC
TGCCAAATCCTTTTATGATCGAATCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAG
AAAGTCCTTTCAATAAGAGCCTTGCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCA
CTTATTGGAGTAAATGTCTCGTTTTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGG
AATGTTTGACTTGGTGGCAAATGGTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCA
CTCAGTATCATACTGTGTGGATTCCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGAAGAAA

GAAGAGAATGACATTCCCAGCTGTGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATC
ACCTTTATCCACCTTTTTCAGATCTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCC
ACGAGGAAACTACAATTCCAGGAACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTAT
AAGTCAGTTCTCAAGATCACCATGACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTAT
GGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCA
TATGGGATAAAACAGATGCATATAATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGA
TATGAGTATGAGTCGTGTTTGGACCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGA
ATTGGATGCGTCCAACATGGGTGGCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTA
TACTGTACAAGGGAAACGGGGAAAACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCGTCGAC

NOV40c, 248993047 Protein Sequence	SEQ ID NO: 490	816 aa	MW at 89174.1kD

GTNWQLQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGI
FWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAGR
QARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPG
FLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACNSGYKG
ESCEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGP
DCSNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIEG
CPGLSNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQP
YCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTP
LIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMKK
EENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGY
KSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVG
YEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSVD

NOV40d, 262802488 DNA Sequence	SEQ ID NO: 491	2519 bp
	ORF Start: at 1	ORF Stop: TAG at 2515

GGTACCAACTGGCAGCTACAGCAGACTGAAAATGACGCATTTGAGAATGGAAAAGTGAATTCTGATAC
CATGCCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAA
ATAACACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATC
TTCTGGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGA
TGCATTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTGGAGC
TCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGGGCGG
CAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGCATCT
GGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCTGTGG
TGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCCAGGA
TTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACTCCAA
GGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGTATTG
ACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAACTCAGGATACAAAGGA
GAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACGGGGA
ATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGACCAGT
GCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGGCCCA
GACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGTGCCG
CTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAGCACG
GGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGCTCAC
TATTTGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGCAACAGCAA
TGGAAGATGTACCCTGGACCAAAATGGCTGGCATTGTGCGTGCCAGCCTGGATGGAGAGGAGCAGGCT
GTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGACTGC
ATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCC
TCAGGGCATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAA
TCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTT
GCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTT
TTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATG
GTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATT
CCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTG
TGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGAT
CTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGA
ACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCAT
GACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCT

TCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATAT
AATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGA
CCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTG
GCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAA
AACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCGTCGACCCAGCAGCCTCCAGTCGTGAGTA
GCG

NOV40d, 262802488 Protein Sequence	SEQ ID NO: 492	838 aa	MW at 91545.9kD

GTNWQLQQTENDAFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGI
FWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAGR
QARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPG
FLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACNSGYKG
ESCEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGP
DCSNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAH
YLDKIVKDKIGYKEGCPGLCNSNGRCTLDQNGWHCACQPGWRGAGCDVAMETLCTDSKDNEGDGLIDC
MDPDCCLQSSCQNQPYCRGLPDPQGIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSL
ASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWI
PWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPG
TDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAY
NQKVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGE
NQFISQQPPVVSSVDPAASSRE

NOV40e, 248993606 DNA Sequence	SEQ ID NO: 493	2536 bp
	ORF Start: at 1	ORF Stop: at 2536

GGTACCAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATAC
CATGCCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAA
ATAACACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATC
TTCTGGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGA
TGCATTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTGGAGC
TCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGGGCGG
CAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGCATCT
GGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCTGTGG
TGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCCAGGA
TTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACTCCAA
GGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGTATTG
ACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCTTGTGCTTGCAACTCAGGATACAAAGGA
GAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACGGGGA
ATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGACCAGT
GCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGGCCCA
GACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGTGTCG
CTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAGCACG
GGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGCTCAC
TATTTGGATAAGATAGTTAAAGAGGGTTGTCCTGGTCTGTGCAACAGCAATGGAAGATGTACCCTGGA
CCAAAATGGCTGGCATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGCCATGGAGA
CTCTTTGCACAGATAGTAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCGATTGCTGC
CTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATCATTAGCCA
AAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCTTATAGGAT
CTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCATCAGAGGC
CAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTACCCAGAATA
TGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTCTCTAACTT
TGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATGTCTTTTAT
GTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGTGGATTCGT
GAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGAAGACAGTC
CCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGAAACTCTCC
TACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCTATTATTCC
ATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTGGTTTCCTG
CCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAGTCTATGGT
CTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTGTGGGAAAA

GAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATTAGATAAAC
ATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCATCTCCCAG
CAGCCTCCAGTCGTGAGTAGCGTCGACGCTACACACGACTGGAGGCTGCTGGTCATTTTCAGCTCTGC
AGTAGCTGCCAGTTGGTACC

NOV40e, 248993606 Protein Sequence	SEQ ID NO: 494	845 aa	MW at 92534.0kD

GTNWQLQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGI
FWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAGR
QARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPG
FLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACNSGYKG
ESCEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGP
DCSNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAH
YLDKIVKEGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCC
LQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRG
QVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFY
VMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLS
YLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYG
LSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQ
QPPVVSSVDATHDWRLLVIFSSAVAASWY

NOV40f, 314411758 DNA Sequence	SEQ ID NO: 495	2500 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCAAGCTTAACTGGCAGCTACAGCAGACTGAAAATGACGCATTTGAGAATGGAAAAGTGAATTCTG
ATACCATGCCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAA
GAAAATAACACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGG
GATCTTCTGGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGA
AGGATGCATTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTG
GAGCTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGG
GCGGCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGC
ATCTGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCT
GTGGTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCC
AGGATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACT
CCAAGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGT
ATTGACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAACTCAGGATACAA
AGGAGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACG
GGGAATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGAC
CAGTGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGG
CCCAGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGT
GCCGCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAG
CACGGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGC
TCACTATTTGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGCAACA
GCAATGGAAGATGTACCCTGGACCAAAATGGCTGGCATTGTGCGTGCCAGCCTGGATGGAGAGGAGCA
GGCTGTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGA
CTGCATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGG
ATCCTCAGGGCATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGAT
CGAATCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAG
CCTTGCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCT
CGTTTTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCA
AATGGTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTG
GATTCCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCA
GCTGTGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTC
AGATCTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCC
AGGAACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCA
CCATGACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGA
CTCTTCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGC
ATATAATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTT
TGGACCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATG
GGTGGCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGG

GGAAAACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCCTCGAGGGC

NOV40f, 314411758 Protein Sequence	SEQ ID NO: 496	833 aa	MW at 91116.5kD

TKLNWQLQQTENDAFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPG
IFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAG
RQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFP
GFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACNSGYK
GESCEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTG
PDCSNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIA
HYLDKIVKDKIGYKEGCPGLCNSNGRCTLDQNGWHCACQPGWRGAGCDVAMETLCTDSKDNEGDGLID
CMDPDCCLQSSCQNQPYCRGLPDPQGIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKS
LASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVW
IPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIP
GTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDA
YNQKVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNG
ENQFISQQPPVVSSLEG

NOV40g, 319067006 DNA Sequence	SEQ ID NO: 497	730 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCTCGCGAACAGAACCTTCTTACGAACTTGTGAAGAGTCAGCAGTGGGATGATATACCGCCCATCT
TCGGAGTCCAGCAGCAAGTGGCGCGGCAGGCCAAGGCCTTCCTGTCGCTGGGGAAGATGGCCGAGGTG
CAGGTGAGCCGGCGCCGGGCCGGCGGCGCGCAGTCCTGGCTGTGGTTCGCCACGGTCAAGTCGCTGAT
CGGCAAGGGCGTCATGCTGGCCGTCAGCCAGGGCCGCGTGCAGACCAACGTGCTCAACATCGCCAACG
AGGACTGCATCAAGGTGGCGGCCGTGCTCAACAACGCCTTCTACCTGGAGAACCTGCACTTCACCATC
GAGGGCAAGGACACGCACTACTTCATCAAGACCACCACGCCCGAGAGCGACCTGGGCACGCTGCGGTT
GACCAGCGGCCGCAAGGCGCTGGAGAACGGCATCAACGTGACGGTGTCGCAGTCCACCACGGTGGTGA
ACGGCAGGACGCGCAGGTTCGCGGACGTGGAGATGCAGTTCGGCGCGCTGGCGCTGCACGTGCGCTAC
GGCATGACCCTGGACGAGGAGAAGGCGCGCATCCTGGAGCAGGCGCAGCAGCGCGTGCGCGACGGCAA
GGTGCAGGGCTACGACGGGTACTACGTACTCTCGGTGGAGCAGTACCCCGAGCTGGCCGACAGCGCCA
ACAACATCCAGTTCCTGCGGCAGAGCGAGATCGGCAGGAGGGTCGACGGC

NOV40g, 319067006 Protein Sequence	SEQ ID NO: 498	243 aa	MW at 27175.5kD

TSRTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMAEVQVSRRRAGGAQSWLWFATVKSLI
GKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNAFYLENLHFTIEGKDTHYFIKTTTPESDLGTLRL
TSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFGALALHVRYGMTLDEEKARILEQAQQRVRDGK
VQGYDGYYVLSVEQYPELADSANNIQFLRQSEIGRRVDG

NOV40h, 319506086 DNA Sequence	SEQ ID NO: 499	1178 bp
	ORF Start: at 78	ORF Stop: end of sequence

CACCTCGCGACCAATGAGACTTGAACCCTGAGCCTAAGTTGTCACCAGCAGGACTGATGTGCACACAG
AAGGAATGAAGTATGGATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAGAA
GGAACGGCGCTACACAAATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCACACAGAAGTCCTACA
GTTCCAGCGAGACATTGAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGAAG
GATTTGGTTCACAGAGAAGCAGACGAGTTCACTAGACAAGAGCAACCTGCAAGCAATCAAGGCCAGTC
TACCCTGCAGCCCTTGCCGCCTTCCCATAAGCAGCACTCTGCACAGCATCATCCATCCATCACTTCTC
TCAACAGAAACTCCCTGACCAATAGAAGGAACCAGAGTCCGGCCCCGCCGGCTGCTTTGCCCGCCGAG
CTGCAAACCACACCCGAGTCCGTCCAGCTGCAGGACAGCTGGGTCCTTGGCAGTAATGTACCACTGGA
AAGCAGGCATTTCCTATTCAAAACAGGAACAGGTACAACGCCACTGTTCAGTACTGCAACCCCAGGAT
ACACAATGGCATCTGGCTCTGTTTATTCACCACCTACTCGGCCACTACCTAGAAACACCCTATCAAGA
AGTGCTTTTAAATTCAAGAAGTCTTCAAAGTACTGTAGCTGGAAATGCACTGCACTGTGTGCCGTAGG
GGTCTCGGTGCTCCTGGCAATACTCCTGTCTTATTTTATAGCAATGCATCTCTTTGGCCTCAACTGGC
AGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATACCATGCCAACAAAC
ACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAAATAACACCATAGA
TTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATCTTCTGGAGATCAC
AGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGATGCATTGATTGGA
GTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTGGAGCTCCTGGATGGCAG

CAGGCTGATTGCCCTCGAGGGC			
NOV40h, 319506086 Protein Sequence	SEQ ID NO: 500	367 aa	MW at 40968.5kD

SMDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLV
HREADEFTRQEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLNRNSLTNRRNQSPAPPAALPAELQT
TPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYTMASGSVYSPPTRPLPRNTLSRSAF
KFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQLQQTENDTFENGKVNSDTMPTNTVS
LPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQLFIDQPQFLKFNISLQKDALIGVYG
RKGLPPSHTQYDFVELLDGSRLIALEG

NOV40i, CG55069-03 DNA Sequence	SEQ ID NO: 501	8473 bp
	ORF Start: ATG at 258	ORF Stop: TAA at 8142

TTGACAGAAAAAGGCAGTAAACGGGGAATCTCTTTTTTTGAATAAAGAAGAAGAAGAAATAAAGTACC
TGTCATCTTGACAAGTGGCGGAGCGGAGGAGTCAAGGATTATAAATGATCACAGCCAGGTCCAGCTCG
CCCCGTGATTGGGCTCTCCCGCGATCTGCACCGGGGGAAGCGCATGAGAGGCCAATGAGACTTGAACC
CTGAGCCTAAGTTGTCACCAGCAGGACTGATGTGCACACAGAAGGAATGAAGTATGGATGTGAAAGAA
CGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAGAAGGAACGGCGCTACACAAATTCCTCCGC
AGACAATGAGGAGTGCCGGGTACCCACACAGAAGTCCTACAGTTCCAGCGAGACATTGAAAGCTTTTG
ATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGAAGGATTTGGTTCACAGAGAAGCAGACGAG
TTCACTAGACAAGAGCAACCTGCAAGCAATCAAGGCCAGTCTACCCTGCAGCCCTTGCCGCCTTCCCA
TAAGCAGCACTCTGCACAGCATCATCCATCCATCACTTCTCTCAACAGAAACTCCCTGACCAATAGAA
GGAACCAGAGTCCGGCCCCGCCGGCTGCTTTGCCCGCCGAGCTGCAAACCACACCCGAGTCCGTCCAG
CTGCAGGACAGCTGGGTCCTTGGCAGTAATGTACCACTGGAAAGCAGGCATTTCCTATTCAAAACAGG
AACAGGTACAACGCCACTGTTCAGTACTGCAACCCCAGGATACACAATGGCATCTGGCTCTGTTTATT
CACCACCTACTCGGCCACTACCTAGAAACACCCTATCAAGAAGTGCTTTTAAATTCAAGAAGTCTTCA
AAGTACTGTAGCTGGAAATGCACTGCACTGTGTGCCGTAGGGGTCTCGGTGCTCCTGGCAATACTCCT
GTCTTATTTTATAGCAATGCATCTCTTTGGCCTCAACTGGCAGCTACAGCAGACTGAAAATGACACAT
TTGAGAATGGAAAAGTGAATTCTGATACCATGCCAACAAACACTGTGTCATTACCTTCTGGAGACAAT
GGAAAATTAGGTGGATTTACGCAAGAAAATAACACCATAGATTCCGGAGAACTTGATATTGGCCGAAG
AGCAATTCAAGAGATTCCTCCCGGGATCTTCTGGAGATCACAGCTCTTCATTGATCAGCCACAGTTTC
TTAAATTCAATATCTCTCTTCAGAAGGATGCATTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCT
TCCCATACTCAGTATGACTTCGTGGAGCTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAG
CCTGCTTGAGACGGAGAGAGCCGGGCGGCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCC
AGTACTTGGATTCTGGAATCTGGCATCTGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCT
TTTAATACCATTGTTATAGAGTCTGTGGTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGT
TTCTGGAACTTGCCATTGTTTTCCAGGATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGT
TATGTAGTGGCAACGGGCAGTACTCCAAGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAG
TGTGATGTGCCGACTACCCAGTGTATTGACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTC
TTGTGCTTGCAACTCAGGATACAAAGGAAAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTT
CTAATCATGGTGTGTGTATCCACGGGGAATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAA
ATACTGAAGACCATGTGTCCAGACCAGTGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTG
CACGTGTGACCCTAACTGGACTGGCCCAGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCAC
ACGGCGTTTGCATGGGGGGGACGTGTCGCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGA
GCCTGCCACCCCCGCTGTGCCGAGCACGGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTG
GAATGGAGAGCACTGCACTATCGCTCACTATTTGGATAAGATAGTTAAAGACAAGATAGGATATAAAG
AGGGTTGTCCTGGTCTGTGCAACAGCAATGGAAGATGTACCCTGGACCAAAATGGCGGACATTGTGTG
TGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGA
CAATGAAGGGGATGGACTCATTGACTGCATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATC
AGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATCATTAGCCAAAGCCTTCAATCGCCTTCTCAG
CAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACC
TGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAA
CTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTACCCAGAATATGGATATACTATTACCCGCCAG
GACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATT
CCTCACTCAGTATCATACTGTGTGGATTCCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGG
AGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTG
TCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGT
ACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAG
GGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCAT

CTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATAC
TTTCATATGGGATAAAACAGATGCATATAATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAG
TTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGC

TATGAATTGGATGCGTCCAACATGGGTGGCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAA
CGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCA
TCATGGGCAATGGGCGAAGGCGCAGCATTTCCTGCCCCAGTTGCAATGGTCAAGCTGATGGTAACAAG
TTACTGGCCCCAGTGGCGCTAGCTTGTGGGATCGATGGCAGTCTGTACGTAGGCGATTTCAACTACGT
GCGGCGGATATTCCCTTCTGGAAATGTAACAAGTGTCTTAGAACTAAGAAATAAAGATTTTAGACATA
GCAGCAACCCAGCTCATAGATACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTACGTTTCTGAC
ACAAACACCCGCAGAATTTATCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAAAAATGCAGA
AGTCGTCGCAGGGACAGGGGAGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATGGAGGGAAGG
CCGTGGAAGCCACACTCATGAGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATCTACTTTGTT
GATGGAACCATGATTAGGAAAGTTGACCAAAATGGAATCATATCAACTCTTCTGGGCTCTAACGATTT
GACTTCAGCCAGACCTTTAACTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGGAATGGCCCA
CTGACCTAGCCATTAACCCTATGGATAACTCCATTTATGTCCTGGATAATAATGTAGTTTTACAGATC
ACTGAAAATCGTCAAGTTCGCATTGCTGCTGGACGGCCCATGCACTGTCAGGTTCCCGGAGTGGAATA
TCCTGTGGGGAAGCACGCGGTGCAGACAACACTGGAATCAGCCACTGCCATTGCTGTGTCCTACAGTG
GGGTCCTGTACATTACTGAAACTGATGAGAAGAAAATTAACCGGATAAGGCAGGTCACAACAGATGGA
GAAATCTCCTTAGTGGCCGGAATACCTTCAGAGTGTGACTGCAAAAATGATGCCAACTGTGACTGTTA
CCAGAGTGGAGATGGCTACGCCAAGGATGCCAAACTCAGTGCCCCATCCTCCCTGGCTGCTTCTCCAG
ATGGTACACTGTATATTGCAGATCTAGGGAATATCCGGATCCGGGCTGTGTCAAAGAATAAGCCTTTA
CTTAACTCTATGAACTTCTATGAAGTTGCGTCTCCAACTGATCAAGAACTCTACATCTTTGACATCAA
TGGTACTCACCAATATACTGTAAGTTTAGTCACTGGTGATTACCTTTACAATTTTAGCTACAGCAATG
ACAATGATATTACTGCTGTGACAGACAGCAATGGCAACACCCTTAGAATTAGACGGGACCCAAATCGC
ATGCCAGTTCGAGTGGTGTCTCCTGATAACCAAGTGATATGGTTGACAATAGGAACAAATGGATGTTT
GAAAGGCATGACTGCTCAAGGACTGGAATTAGTTTTGTTTACTTACCATGGCAATAGTGGCCTTTTAG
CCACTAAAAGTGATGAAACTGGATGGACAACGTTTTTTGACTATGACAGTGAAGGTCGTCTGACAAAT
GTTACGTTTCCAACTGGAGTGGTCACAAACCTGCATGGGGACATGGACAAGGCTATCACAGTGGACAT
TGAGTCATCTAGCCGAGAAGAAGATGTCAGCATCACTTCAAATCTGTCCTCGATCGATTCTTTCTACA
CCATGGTTCAAGATCAGTTAAGAAACAGCTACCAGATTGGTTATGACGGCTCCCTCAGAATTATCTAC
GCCAGTGGCCTGGACTCACACTACCAAACAGAGCCGCACGTTCTGGCTGGCACCGCTAATCCGACGGT
TGCCAAAAGAAACATGACTTTGCCTGGCGAGAACGGTCAAAACTTGGTGGAATGGAGATTCCGAAAAG
AGCAAGCCCAAGGGAAAGTCAATGTCTTTGGCCGCAAGCTCAGGGTTAATGGCAGAAACCTCCTTTCA
GTTGACTTTGATCGAACAACAAAGACAGAAAAGATCTATGACGACCACCGTAAATTTCTACTGAGGAT
CGCCTACGACACGTCTGGGCACCCGACTCTCTGGCTGCCAAGCAGCAAGCTGATGGCCGTCAATGTCA
CCTATTCATCCACAGGTCAAATTGCCAGCATCCAGCGAGGCACCACTAGCGAGAAAGTAGATTATGAC
GGACAGGGGAGGATCGTGTCTCGGGTCTTTGCTGATGGTAAAACATGGAGTTACACATATTTAGAAAA
GTCCATGGTTCTTCTGCTTCATAGCCAGCGGCAGTACATCTTCGAATACGATATGTGGGACCGCCTGT
CTGCCATCACCATGCCCAGTGTGGCTCGCCACACCATGCAGACCATCCGATCCATTGGCTACTACCGC
AACATATACAACCCCCCGGAAAGCAACGCCTCCATCATCACGGACTACAACGAGGAAGGGCTGCTTCT
ACAAACAGCTTTCTTGGGTACAAGTCGGAGGGTCTTATTCAAATACAGAAGGCAGACTAGGCTCTCAG
AAATTTTATATGATAGCACAAGAGTCAGTTTTACCTATGATGAAACAGCAGGAGTCCTAAAGACAGTA
AACCTCCAGAGTGATGGTTTTATTTGCACCATTAGATACAGGCAAATTGGTCCCCTGATTGACAGGCA
GATTTTCCGCTTTAGTGAAGATGGGATGGTAAATGCAAGATTTGACTATAGCTATGACAACAGCTTTC
GAGTGACCAGCATGCAGGGTGTGATCAATGAAACGCCACTGCCTATTGATCTGTATCAGTTTGATGAC
ATTTCTGGCAAAGTTGAGCAGTTTGGAAAGTTTGGAGTTATATATTATGATATTAACCAGATCATTTC
TACAGCTGTAATGACCTATACGAAGCACTTTGATGCTCATGGCCGTATCAAGGAGATTCAATATGAGA
TATTCAGGTCGCTCATGTACTGGATTACAATTCAGTATGATAACATGGGTCGGGTAACCAAGAGAGAG
ATTAAAATAGGGCCCTTTGCCAACACCACCAAATATGCTTATGAATATGATGTTGATGGACAGCTCCA
AACAGTTTACCTCAATGAAAAGATAATGTGGCGGTACAACTACGATCTGAATGGAAACCTCCATTTAC
TGAACCCAAGTAACAGTGCGCGTCTGACACCCCTTCGCTATGACCTGCGAGACAGAATCACTCGACTG
GGTGATGTTCAATATCGGTTGGATGAAGATGGTTTCCTACGTCAAAGGGGCACGGAAATCTTTGAATA

TAGCTCCAAGGGGCTTCTAACTCGAGTTTACAGTAAAGGCAGTGGCTGGACAGTGATCTACCGTTATG
ACGGCCTGGGAAGGCGTGTTTCTAGCAAAACCAGTCTAGGACAGCACCTGCAGTTTTTTTATGCTGAC
TTAACTTATCCCACTAGGATTACTCATGTCTACAACCATTCGAGTTCAGAAATTACCTCCCTGTATTA
TGATCTCCAAGGACATCTTTTTGCCATGGAAATCAGCAGTGGGGATGAATTCTATATTGCATCGGATA
ACACAGGGACACCACTGGCTGTGTTCAGTAGCAATGGGCTTATGCTGAAACAGATTCAGTACACTGCA
TATGGGGAAATCTATTTTGACTCTAATATTGACTTTCAACTGGTAATTGGATTTCATGGTGGCCTGTA
TGACCCACTCACCAAATTAATCCACTTTGGAGAAAGAGATTATGACATTTTGGCAGGACGGTGGACAA

CACCTGACATAGAAATCTGGAAAAGAATTGGGAAGGACCCAGCTCCTTTTAACTTGTACATGTTTAGG
AATAACAACCCTGCAAGCAAAATCCATGACGTGAAAGATTACATCACAGATGTTAACAGCTGGCTGGT
GACATTTGGTTTCCATCTGCACAATGCTATTCCTGGATTCCCTGTTCCCAAATTTGATTTAACAGAAC
CTTCTTACGAACTTGTGAAGAGTCAGCAGTGGGATGATATACCGCCCATCTTCGGAGTCCAGCAGCAA
GTGGCGCGGCAGGCCAAGGCCTTCCTGTCGCTGGGGAAGATGGCCGAGGTGCAGGTGAGCCGGCGCCG
GGCCGGCGGCGCGCAGTCCTGGCTGTGGTTCGCCACGGTCAAGTCGCTGATCGGCAAGGGCGTCATGC
TGGCCGTCAGCCAGGGCCGCGTGCAGACCAACGTGCTCAACATCGCCAACGAGGACTGCATCAAGGTG
GCGGCCGTGCTCAACAACGCCTTCTACCTGGAGAACCTGCACTTCACCATCGAGGGCAAGGACACGCA
CTACTTCATCAAGACCACCACGCCCGAGAGCGACCTGGGCACGCTGCGGTTGACCAGCGGCCGCAAGG
CGCTGGAGAACGGCATCAACGTGACGGTGTCGCAGTCCACCACGGTGGTGAACGGCAGGACGCGCAGG
TTCGCGGACGTGGAGATGCAGTTCGGCGCGCTGGCGCTGCACGTGCGCTACGGCATGACCCTGGACGA
GGAGAAGGCGCGCATCCTGGAGCAGGCGCGGCAGCGCGCGCTCGCCCGGGCCTGGGCGCGCGAGCAGC
AGCGCGTGCGCGACGGCGAGGAGGGCGCGCGCCTCTGGACGGAGGGCGAGAAGCGGCAGCTGCTGAGC
GCCGGCAAGGTGCAGGGCTACGACGGGTACTACGTACTCTCGGTGGAGCAGTACCCCGAGCTGGCCGA
CAGCGCCAACAACATCCAGTTCCTGCGGCAGAGCGAGATCGGCAGGAGGTAACGCCCGGGCCGCGCCC
GCCGAGCCGCTCACGCCCTGCCCACATTGTCCTGTGGCACAACCCGAGTGGGACTCTCCAACGCCCAA
GAGCCTTCCTCCCGGGGGAATGAGACTGCTGTTACGACCCACACCCACACCGCGAAAACAAGGACCGC
TTTTTTCCGAATGACCTTAAAGGTGATCGGCTTTAACGAATATGTTTACATATGCATAGCGCTGCACT
CAGTCGGACTGAACGTAGCCAGAGGAAAAAAAAATCATCAAGGACAAAGGCCTCGACCTGTTGCGCTG
GGCCGTCTGTTCCTTCTAGGCACTGTATTTAACTAACTTTA

NOV40i, CG55069-03 Protein Sequence	SEQ ID NO: 502	2628 aa	MW at 293503.3kD

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLNRNSLTNRRNQSPAPPAALPAELQTT
PESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYTMASGSVYSPPTRPLPRNTLSRSAFK
FKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQLQQTENDTFENGKVNSDTMPTNTVSL
PSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGR
KGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKN
AEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSG
WKGTECDVPTTQCIDPQCGGRGICIMGSCACNSGYKGKSCEEADCIDPGCSNHGVCIHGECHCSPGWG
GSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGTCRCEEGWTGP
ACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVKDKIGYKEGCPGLCNSNGRCTLDQN
GGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQPYCRGLPDPQDIISQSL
QSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGY
TITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMEKEENDIPSCDLSGFVRP
NPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFN
LMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSISCPSCNGQ
ADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELRNKDFRHSSNPAHRYYLATDPVTGD
LYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEARCGDGGKAVEATLMSPKGMAVDKNG
LIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIYVLDNN
VVLQITENRQVRIAAGRPMHCQVPGVEYPVGKHAVQTTLESATAIAVSYSGVLYITETDEKKINRIRQ
VTTDGEISLVAGIPSECDCKNDANCDCYQSGDGYAKDAKLSAPSSLAASPDGTLYIADLGNIRIRAVS
KNKPLLNSMNFYEVASPTDQELYIFDINGTHQYTVSLVTGDYLYNFSYSNDNDITAVTDSNGNTLRIR
RDPNRMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELVLFTYHGNSGLLATKSDETGWTTFFDYDSE
GRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSITSNLSSIDSFYTMVQDQLRNSYQIGYDGS
LRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGENGQNLVEWRFRKEQAQGKVNVFGRKLRVNG
RNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLWLPSSKLMAVNVTYSSTGQIASIQRGTTSE
KVDYDGQGRIVSRVFADGKTWSYTYLEKSMVLLLHSQRQYIFEYDMWDRLSAITMPSVARHTMQTIRS
IGYYRNIYNPPESNASIITDYNEEGLLLQTAFLGTSRRVLFKYRRQTRLSEILYDSTRVSFTYDETAG
VLKTVNLQSDGFICTIRYRQIGPLIDRQIFRFSEDGMVNARFDYSYDNSFRVTSMQGVINETPLPIDL
YQFDDISGKVEQFGKFGVIYYDINQIISTAVMTYTKHFDAHGRIKEIQYEIFRSLMYWITIQYDNMGR
VTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWRYNYDLNGNLHLLNPSNSARLTPLRYDLRD
RITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYSKGSGWTVIYRYDGLGRRVSSKTSLGQHLQ
FFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEISSGDEFYIASDNTGTPLAVFSSNGLMLKQ
IQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIHFGERDYDILAGRWTTPDIEIWKRIGKDPAPFN
LYMFRNNNPASKIHDVKDYITDVNSWLVTFGFHLHNAIPGFPVPKFDLTEPSYELVKSQQWDDIPPIF
GVQQQVARQAKAFLSLGKMAEVQVSRRRAGGAQSWLWFATVKSLIGKGVMLAVSQGRVQTNVLNIANE

DCIKVAAVLNNAFYLENLHFTIEGKDTHYFIKTTTPESDLGTLRLTSGRKALENGINVTVSQSTTVVN
GRTRRFADVEMQFGALALHVRYGMTLDEEKARILEQARQRALARAWAREQQRVRDGEEGARLWTEGEK
RQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQSEIGRR

NOV40j, 219937039 DNA Sequence	SEQ ID NO: 503	1854 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGC
CATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCG
ATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATC
ATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCT
TATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCA
TCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTAC
CCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTC
TCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATG
TCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGT
GGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGA
AGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGA
AACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCT
ATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTG
GTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAG
TCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTG
TGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATT
AGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCA
TCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGGCGCAGCATTTCCTGCCCC
AGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCTAGCTTGTGGGATCGATGG
CAGTCTGTACGTAGGCGATTTCAACTATGTGCGGCGGATATTCCCTTCTGGAAATGTAACAAGTGTCT
TAGAACTAAGAAATAAAGATTTTAGACATAGCAGCAACCCAGCTCATAGATACTACCTTGCAACGGAT
CCAGTCACGGGAGATCTGTACGTTTCTGACACAAACACCCGCAGAATTTATCGCCCAAAGTCACTTAC
GGGGGCAAAAGACTTGACTAAAAATGCAGAAGTCGTCGCAGGGACAGGGGAGCAATGCCTTCCGTTTG
ACGAGGCGAGATGTGGGGATGGAGGGAAGGCCGTGGAAGCCACACTCATGAGTCCCAAAGGAATGGCA
GTTGATAAGAATGGATTAATCTACTTTGTTGATGGAACCATGATTAGGAAAGTTGACCAAAATGGAAT
CATATCAACTCTTCTGGGTTCTAACGATTTGACTTCAGCCAGACCTTTAACTTGTGACACCAGCATGC
ACATCAGCCAGGTACGTCTGGAATGGCCCACTGACCTAGCCATTAACCCTATGGATAACTCCATTTAT
GTCCTGGATAATGTCGAC

NOV40j, 219937039 Protein Sequence	SEQ ID NO: 504	618 aa	MW at 68016.2kD

KLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQPYCRGLPDPQDI
ISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHY
PEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLS
GFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQS
IIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTL
WEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSISCP
SCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELRNKDFRHSSNPAHRYYLATD
PVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEARCGDGGKAVEATLMSPKGMA
VDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIY
VLDNVD

NOV40k, 219937046 DNA Sequence	SEQ ID NO: 505	1854 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGC
CATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCG
ATTGTTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATC
ATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCT
TATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCA
TCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTAC
CCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAGTGGTGGGGCCTC
TCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATG

TCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGT
GGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGA
AGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGA
AACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCT
ATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTG
GTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAG
TCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTG
TGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATT
AGATAAACATCACGCGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCA
TCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGGCGCAGCATTTCCTGCCCC
AGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCTAGCTTGTGGGATCGATGG
CAGTCTGTACGTAGGCGATTTCAACTATGTGCGGCGGATATTCCCTTCTGGAAATGTAACAAGTGTCT
TAGAACTAAGAAATAAAGATTTTAGACATAGCAGCAACCCAGCTCATAGATACTACCTTGCAACGGAT
CCAGTCACGGGAGATCTGTACGTTTCTGACACAAACACCCGCAGAATTTATCGCCCAAAGTCACTTAC
GGGGGCAAAAGACTCGACTAAAAATGCAGAAGTCGTCGCAGGGACAGGGGAGCAATGCCTTCCGTTTG
ACGAGGCGAGATGTGGGGATGGAGGGAAGGCCGTGGAAGCCACACTCATGAGTCCCAAAGGAATGGCA
GTTGATAAGAATGGATTAATCTACTTTGTTGATGGAACCATGATTAGGAAAGTTGACCAAAATGGAAT
CATATCAACTCTTCTGGGTTCTAACGATTTGACTTCAGCCAGACCTTTAACTTGTGACACCAGCATGC
ACATCAGCCAGGTACGTCTGGAATGGCCCACTGACCTAGCCATTAACCCTATGGATAACTCCATTTAT
GTCCTGGATAATGTCGAC

NOV40k, 219937046 Protein Sequence	SEQ ID NO: 506	618 aa	MW at 67935.1kD

KLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQPYCRGLPDPQDI
ISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHY
PEYGYTITRQDGMFDLVASGGASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLS
GFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQS
IIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTL
WEKRTAILQGYELDASNMGGWTLDKHHALDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSISCP
SCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELRNKDFRHSSNPAHRYYLATD
PVTGDLYVSDTNTRRIYRPKSLTGAKDSTKNAEVVAGTGEQCLPFDEARCGDGGKAVEATLMSPKGMA
VDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIY
VLDNVD

NOV40l, 219937583 DNA Sequence	SEQ ID NO: 507	1834 bp
	ORF Start: at 2	ORF Stop: end of sequence

TAAGCTTGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAG
CCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCC
GATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACAT
CATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCC
TTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTC
ATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTA
CCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCT
CTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAAT
GTCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAG
TGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTG
AAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTG
AAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTC
TATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGT
GGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAA
GTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCT
GTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACAT
TAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTC
ATCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGGCGCAGCATTTCCTGCCC
CAGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCTAGCTTGTGGGATCGATG
GCAGTCTGTACGTAGGCGATTTCAACTATGTGCGGCGGATATTCCCTTCTGGAAATGTAACAAGTGTC
TTAGAACTAAGCAGCAACCCAGCTCATAGATACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTA
CGTTTCTGACACAAACACCCGCAGAATTTATCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTA
AAAATGCAGAAGTCGTCGCAGGGACAGGGGAGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGAT

GGAGGGAAGGCCGTGGAAGCCACACTCATGAGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAAT
CTACTTTGTTGATGGAACCATGATTAGGAAAGTTGACCAAAATGGAATCATATCAACTCTTCTGGGTT
CTAACGATTTGACTTCAGCCAGACCTTTAACTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTG
GAATGGCCCACTGACCTAGCCATTAACCCTATGGATAACTCCATTTATGTCCTGGATAATGTCGAC

NOV40l, 219937583 Protein Sequence	SEQ ID NO: 508	611 aa	MW at 67062.2kD

KLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQPYCRGLPDPQDI
ISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHY
PEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLS
GFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQS
IIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTL
WEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSISCP
SCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELSSNPAHRYYLATDPVTGDLY
VSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEARCGDGGKAVEATLMSPKGMAVDKNGLI
YFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD

NOV40m, 219937000 DNA Sequence	SEQ ID NO: 509	1833 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGC
CATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCG
ATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATC
ATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCT
TATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCA
TCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTAC
CCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTC
TCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATG
TCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGT
GGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGA
AGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGA
AACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCT
ATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTG
GTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAG
TCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTG
TGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATT
AGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCA
TCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGGCGCAGCATTTCCTGCCCC
AGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCTAGCTTGTGGGATCGATGG
CAGTCTGTACGTAGGCGATTTCAACTATGTGCGGCGGATATTCCCTTCTGGAAATGTAACAAGTGTCT
TAGAACTAAGCAGCAACCCAGCTCATAGATACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTAC
GTTTCTGACACAAACACCCGCAGAATTTATCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAA
AAATGCAGAAGTCGTCGCAGGGACAGGGGAGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATG
GAGGGAAGGCCGTGGAAGCCACACTCATGAGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATC
TACTTTGTTGATGGAACCATGATTAGGAAAGTTGACCAAAATGGAATCATATCAACTCTTCTGGGTTC
TAACGATTTGACTTCAGCCAGACCTTTAACTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGG
AATGGCCCACTGACCTAGCCATTAACCCTATGGATAACTCCATTTATGTCCTGGATAATGTCGAC

NOV40m, 219937000 Protein Sequence	SEQ ID NO: 510	611 aa	MW at 67062.2kD

KLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQPYCRGLPDPQDI
ISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHY
PEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLS
GFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQS
IIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTL
WEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSISCP
SCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELSSNPAHRYYLATDPVTGDLY
VSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEARCGDGGKAVEATLMSPKGMAVDKNGLI
YFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD

NOV40n, 219937005 DNA Sequence	SEQ ID NO: 511	1833 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGC
CATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCG
ATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATC
ATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCT
TATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCA
TCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTAC
CCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTC
TCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATG
TCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGT
GGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGA
AGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGA
AACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCT
ATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTG
GTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAG
TCTATGGTCTATCTGAAGCTGTTGTGTCAGCTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTG
TGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATT
AGATAAACATCACGTGCTGGATGTACGGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCA
TCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGGCGCAGCATTTCCTGCCCC
AGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCTAGCTTGTGGGATCGATGG
CAGTCTGTACGTAGGCGATTTCAACTATGTGCGGCGGATATTCCCTTCTGGAAATGTAACAAGTGTCT
TAGAACTAAGCAGCAACCCAGCTCATAGATACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTAC
GTTTCTGACACAAACACCCGCAGAATTTATCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAA
AAATGCAGAAGTCGTCGCAGGGACAGGGGAGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATG
GAGGGAAGGCCGTGGAAGCCACACTCATGAGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATC
TACTTTGTTGATGGAACCATGATTAGGAAAGTTGACCAAAATGGAATCATATCAACTCTTCTGGGTTC
TAACGATTTGACTTCAGCCAGACCTTTAACTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGG
AATGGCCCACTGACCTAGCCATTAACCCTATGGATAACTCCATTTATGTCCTGGATAATGTCGAC

NOV40n, 219937005 Protein Sequence	SEQ ID NO: 512	611 aa	MW at 67062.2kD

KLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQPYCRGLPDPQDI
ISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHY
PEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLS
GFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQS
IIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSAGYEYESCLDLTL
WEKRTAILQGYELDASNMGGWTLDKHHVLDVRNGILYKGNGENQFISQQPPVVSSIMGNGRRRSISCP
SCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELSSNPAHRYYLATDPVTGDLY
VSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEARCGDGGKAVEATLMSPKGMAVDKNGLI
YFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD

NOV40o, 219937013 DNA Sequence	SEQ ID NO: 513	1833 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGC
CATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCG
ATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATC
ATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCT
TATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCA
TCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTAC
CCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTC
TCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATG
TCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGT
GGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGA
AGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGA
AACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCT
ATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTG
GTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAG

TCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTG
TGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATT
AGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCA
TCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGGCGCAGCATTTCCTGCCCC
AGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCTAGCTTGTGGGATCGATGG
CAGTCTGTACGTAGGCGATTTCAACTATGTGCGGCGGATATTCCCTTCTGGAAATGTAACAAGTGTCT
TAGAACTAAGCAGCAACCCAGCTCATAGATACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTAC
GTTTCTGACACAAACACCCGCAGAATTTATCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAA
AAATGCAGAAGTCGTCGCAGGGACAGGGGAGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATG
GAGGGAAGGCCGCGGAAGCCACACTCATGAGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATC
TACTTTGTTGATGGAACCATGATTAGGAAAGTTGACCAAAATGGAATCATATCAACTCTTCTGGGCTC
TAACGATTTGACTTCAGCCAGACCTTTAACTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGG
AATGGCCCACTGACCTAGCCATTAACCCTATGGATAACTCCATTTATGTCCTGGATAATGTCGAC

NOV40o, 219937013 Protein Sequence	SEQ ID NO: 514	611 aa	MW at 67034.1kD

KLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQPYCRGLPDPQDI
ISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHY
PEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLS
GFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQS
IIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTL
WEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSISCP
SCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELSSNPAHRYYLATDPVTGDLY
VSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEARCGDGGKAAEATLMSPKGMAVDKNGLI
YFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD

NOV40p, 219937063 DNA Sequence	SEQ ID NO: 515	1833 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGC
CATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCG
ATTGCTGCCTACAGAGTTCCTGCCAGAATCAGTCCTATTGTCGGGGACTGCCGGATCCTCAGGACATC
ATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCT
TATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCA
TCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTAC
CCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTC
TCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATG
TCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGT
GGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGA
AGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGA
AACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCT
ATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTG
GTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAG
TCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTG
TGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATT
AGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCA
TCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGGCGCAGCATTTCCTGCCCC
AGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCTAGCTTGTGGGATCGATGG
CAGTCTGCACGTAGGCGATTTCAACTATGTGCGGCGGATATTCCCTTCTGGAAATGTAACAAGTGTCT
TAGAACTAAGCAGCAACCCAGCTCATAGATACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTAC
GTTTCTGACACAAACACCCGCAGAATTTATCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAA
AAATGCAGAAGTCGTCGCAGGGACAGGGGAGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATG
GAGGGAAGGCCGTGGAAGCCACACTCATGAGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATC
TACTTTGTTGATGGAACCATGATTAGGAAAGTTGACCAAAATGGAATCATATCAACTCTTCTGGGTTC
TAACGATTTGACTTCAGCCAGACCTTTAACTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGG
AATGGCCCACTGACCTAGCCATTAACCCTATGGATAACTCCATTTATGTCCTGGATAATGTCGAC

NOV40p, 219937063 Protein Sequence	SEQ ID NO: 516	611 aa	MW at 67026.1kD

KLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQSYCRGLPDPQDI

ISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHY
PEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLS
GFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQS
IIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTL
WEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSISCP
SCNGQADGNKLLAPVALACGIDGSLHVGDFNYVRRIFPSGNVTSVLELSSNPAHRYYLATDPVTGDLY
VSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEARCGDGGKAVEATLMSPKGMAVDKNGLI
YFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD

NOV40q, CG55069-09 DNA Sequence	SEQ ID NO: 517	2536 bp
	ORF Start: at 7	ORF Stop: at 2470

GGTACCAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATAC
CATGCCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAA
ATAACACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATC
TTCTGGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGA
TGCATTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTGGAGC
TCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGGGCGG
CAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGCATCT
GGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCTGTGG
TGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCCAGGA
TTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACTCCAA
GGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGTATTG
ACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCTTGTGCTTGCAACTCAGGATACAAAGGA
GAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACGGGGA
ATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGACCAGT
GCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGGCCCA
GACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGTGTCG
CTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAGCACG
GGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGCTCAC
TATTTGGATAAGATAGTTAAAGAGGGTTGTCCTGGTCTGTGCAACAGCAATGGAAGATGTACCCTGGA
CCAAAATGGCTGGCATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGCCATGGAGA
CTCTTTGCACAGATAGTAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCGATTGCTGC
CTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATCATTAGCCA
AAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCTTATAGGAT
CTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCATCAGAGGC
CAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTACCCAGAATA
TGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTCTCTAACTT
TGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATGTCTTTTAT
GTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGTGGATTCGT
GAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGAAGACAGTC
CCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGAAACTCTCC
TACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCTATTATTCC
ATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTGGTTTCCTG
CCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAGTCTATGGT
CTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTGTGGGAAAA
GAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATTAGATAAAC
ATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCATCTCCCAG
CAGCCTCCAGTCGTGAGTAGCGTCGACGCTACACACGACTGGAGGCTGCTGGTCATTTTCAGCTCTGC
AGTAGCTGCCAGTTGGTACC

NOV40q, CG55069-09 Protein Sequence	SEQ ID NO: 518	821 aa	MW at 89886.1kD

NWQLQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFW
RSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAGRQA
RSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFL
GPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACNSGYKGES
CEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYL
DKIVKEGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQ

SSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQV
LTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVM
DTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYL
SSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLS
EAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQP
PVVSS

NOV40r, 309327410 DNA Sequence	SEQ ID NO: 519	2482 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCTCGCGAAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTG
ATACCATGCCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAA
GAAAATAACACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGG
GATCTTCTGGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGA
AGGATGCATTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTG
GAGCTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGG
GCGGCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGC
ATCTGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCT
GTGGTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCC
AGGATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACT
CCAAGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGT
ATTGACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCTTGTGCTTGCAACTCAGGATACAA
AGGAGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACG
GGGAATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGAC
CAGTGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGG
CCCAGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGT
GTCGCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAG
CACGGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGC
TCACTATTTGGATAAGATAGTTAAAGAGGGTTGTCCTGGTCTGTGCAACAGCAATGGAAGATGTACCC
TGGACCAAAATGGCTGGCATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGCCATG
GAGACTCTTTGCACAGATAGTAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCGATTG
CTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATCATTA
GCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCTTATA
GGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCATCAG
AGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTACCCAG
AATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTCTCTA
ACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATGTCTT
TTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGTGGAT
TCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGAAGAC
AGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGAAACT
CTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCTATTA
TTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTGGTTT
CCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAGTCTA
TGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTGTGGG
AAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATTAGAT
AAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCATCTC
CCAGCAGCCTCCAGTCGTGAGTAGCCTCGAGGGC

NOV40r, 309327410 Protein Sequence	SEQ ID NO: 520	827 aa	MW at 90529.8kD

TSRNWQLQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPG
IFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAG
RQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFP
GFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACNSGYK
GESCEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTG
PDCSNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIA
HYLDKIVKEGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDC
CLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIR
GQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVF
YVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKL

SYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVY
GLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFIS
QQPPVVSSLEG

NOV40s, CG55069-02 DNA Sequence	SEQ ID NO: 521	8645 bp
	ORF Start: ATG at 151	ORF Stop: TAA at 8314

TTTGGCCTCGGGCCAGAATTCGGCACGAGGGGTCTGGAGCTTGGAGGAGAAGTCTGAACTAAGGATAA
ACTAAAGAGAGGCCAATGAGACTTGAACCCTGAGCCTAAGTTGTCACCAGCAGGACTGATGTGCACAC
AGAAGGAATGAAGTATGGATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAG
AAGGAACGGCGCTACACAAATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCACACAGAAGTCCTA
CAGTTCCAGCGAGACATTGAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGA
AGGATTTGGTTCACAGAGAAGCAGACGAGTTCACTAGACAAGGACAGAATTTTACCCTAAGGCAGTTA
GGAGTTTGTGAACCAGCAACTCGAAGAGGACTGGCATTTTGTGCGGAAATGGGGCTCCCTCACAGAGG
TTACTCTATCAGTGCAGGGTCAGATGCTGATACTGAAAATGAAGCAGTGATGTCCCCAGAGCATGCCA
TGAGACTTTGGGGCAGGGGGGTCAAATCAGGCCGCAGCTCCTGCCTGTCAAGTCGGTCCAACTCAGCC
CTCACCCTGACAGATACGGAGCACGAAAACAAGTCCGACAGTGAGAATGAGCAACCTGCAAGCAATCA
AGGCCAGTCTACCCTGCAGCCCTTGCCGCCTTCCCATAAGCAGCACTCTGCACAGCATCATCCATCCA
TCACTTCTCTCAACAGAAACTCCCTGACCAATAGAAGGAACCAGAGTCCGGCCCCGCCGGCTGCTTTG
CCCGCCGAGCTGCAAACCACACCCGAGTCCGTCCAGCTGCAGGACAGCTGGGTCCTTGGCAGTAATGT
ACCACTGGAAAGCAGGCATTTCCTATTCAAAACAGGAACAGGTACAACGCCACTGTTCAGTACTGCAA
CCCCAGGATACACAATGGCATCTGGCTCTGTTTATTCACCACCTACTCGGCCACTACCTAGAAACACC
CTATCAAGAAGTGCTTTTAAATTCAAGAAGTCTTCAAAGTACTGTAGCTGGAAATGCACTGCACTGTG
TGCCGTAGGGGTCTCGGTGCTCCTGGCAATACTCCTGTCTTATTTTATAGCAATGCATCTCTTTGGCC
TCAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATACCATG
CCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAAATAA
CACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATCTTCT
GGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGATGCA
TTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTGGAGCTCCT
GGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGGGCGGCAGG
CGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGCATCTGGCT
TTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCTGTGGTGGA
ATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCCAGGATTTC
TGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACTCCAAGGGC
CGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGTATTGACCC
ACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAGCTCAGGATACAAAGGAGAAA
GTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACGGGGAATGT
CACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGACCAGTGCTC
CGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGGCCCAGACT
GCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGTGTCGCTGT
GAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAGCACGGGAC
CTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGCTCACTATT
TGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGCAACAGCAATGGA
AGATGTACCCTGGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGA
CGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGGGATGGACTCATTGACTGCATGG
ATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAG
GACATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAG
TTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCAT
CTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTC
CATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGG
GGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCAT
GGAATGTCTTTTATGTGATGGATACCCTAGTCATGGAGAAAGAAGAGAATGACATTCCCAGCTGTGAT
CTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTC
TCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAG
ATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACC
CAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCA

AAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATC
AGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTG
ACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTG
GACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACC

AGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGGCGCAGCATTTCC
TGCCCCAGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCTAGCTTGTGGGAT
CGATGGCAGTCTGTACGTAGGCGATTTCAACTACGTGCGGCGGATATTCCCTTCTGGAAATGTAACAA
GTGTCTTAGAACTAAGAAATAAAGATTTTAGACATAGCAGCAACCCAGCTCATAGATACTACCTTGCA
ACGGATCCAGTCACGGGAGATCTGTACGTTTCTGACACAAACACCCGCAGAATTTATCGCCCAAAGTC
ACTTACGGGGGCAAAAGACTTGACTAAAAATGCAGAAGTCGTCGCAGGGACAGGGGAGCAATGCCTTC
CGTTTGACGAGGCGAGATGTGGGGATGGAGGGAAGGCCGTGGAAGCCACACTCATGAGTCCCAAAGGA
ATGGCAGTTGATAAGAATGGATTAATCTACTTTGTTGATGGAACCATGATTAGGAAAGTTGACCAAAA
TGGAATCATATCAACTCTTCTGGGCTCTAACGATTTGACTTCAGCCAGACCTTTAACTTGTGACACCA
GCATGCACATCAGCCAGGTACGTCTGGAATGGCCCACTGACCTAGCCATTAACCCTATGGATAACTCC
ATTTATGTCCTGGATAATAATGTAGTTTTACAGATCACTGAAAATCGTCAAGTTCGCATTGCTGCTGG
ACGGCCCATGCACTGTCAGGTTCCCGGAGTGGAATATCCTGTGGGGAAGCACGCGGTGCAGACAACAC
TGGAATCAGCCACTGCCATTGCTGTGTCCTACAGTGGGGTCCTGTACATTACTGAAACTGATGAGAAG
AAAATTAACCGGATAAGGCAGGTCACAACAGATGGAGAAATCTCCTTAGTGGCCGGAATACCTTCAGA
GTGTGACTGCAAAAATGATGCCAACTGTGACTGTTACCAGAGTGGAGATGGCTACGCCAAGGATGCCA
AACTCAGTGCCCCATCCTCCCTGGCTGCTTCTCCAGATGGTACACTGTATATTGCAGATCTAGGGAAT
ATCCGGATCCGGGCTGTGTCAAAGAATAAGCCTTTACTTAACTCTATGAACTTCTATGAAGTTGCGTC
TCCAACTGATCAAGAACTCTACATCTTTGACATCAATGGTACTCACCAATATACTGTAAGTTTAGTCA
CTGGTGATTACCTTTACAATTTTAGCTACAGCAATGACAATGATATTACTGCTGTGACAGACAGCAAT
GGCAACACCCTTAGAATTAGACGGGACCCAAATCGCATGCCAGTTCGAGTGGTGTCTCCTGATAACCA
AGTGATATGGTTGACAATAGGAACAAATGGATGTTTGAAAGGCATGACTGCTCAAGGACTGGAATTAG
TTTTGTTTACTTACCATGGCAATAGTGGCCTTTTAGCCACTAAAAGTGATGAAACTGGATGGACAACG
TTTTTTGACTATGACAGTGAAGGTCGTCTGACAAATGTTACGTTTCCAACTGGAGTGGTCACAAACCT
GCATGGGGACATGGACAAGGCTATCACAGTGGACATTGAGTCATCTAGCCGAGAAGAAGATGTCAGCA
TCACTTCAAATCTGTCCTCGATCGATTCTTTCTACACCATGGTTCAAGATCAGTTAAGAAACAGCTAC
CAGATTGGTTATGACGGCTCCCTCAGAATTATCTACGCCAGTGGCCTGGACTCACACTACCAAACAGA
GCCGCACGTTCTGGCTGGCACCGCTAATCCGACGGTTGCCAAAAGAAACATGACTTTGCCTGGCGAGA
ACGGTCAAAACTTGGTGGAATGGAGATTCCGAAAAGAGCAAGCCCAAGGGAAAGTCAATGTCTTTGGC
CGCAAGCTCAGGGTTAATGGCAGAAACCTCCTTTCAGTTGACTTTGATCGAACAACAAAGACAGAAAA
GATCTATGACGACCACCGTAAATTTCTACTGAGGATCGCCTACGACACGTCTGGGCACCCGACTCTCT
GGCTGCCAAGCAGCAAGCTGATGGCCGTCAATGTCACCTATTCATCCACAGGTCAAATTGCCAGCATC
CAGCGAGGCACCACTAGCGAGAAAGTAGATTATGACGGACAGGGGAGGATCGTGTCTCGGGTCTTTGC
TGATGGTAAAACATGGAGTTACACATATTTAGAAAAGTCCATGGTTCTTCTGCTTCATAGCCAGCGGC
AGTACATCTTCGAATACGATATGTGGGACCGCCTGTCTGCCATCACCATGCCCAGTGTGGCTCGCCAC
ACCATGCAGACCATCCGATCCATTGGCTACTACCGCAACATATACAACCCCCCGGAAAGCAACGCCTC
CATCATCACGGACTACAACGAGGAAGGGCTGCTTCTACAAACAGCTTTCTTGGGTACAAGTCGGAGGG
TCTTATTCAAATACAGAAGGCAGACTAGGCTCTCAGAAATTTTATATGATAGCACAAGAGTCAGTTTT
ACCTATGATGAAACAGCAGGAGTCCTAAAGACAGTAAACCTCCAGAGTGATGGTTTTATTTGCACCAT
TAGATACAGGCAAATTGGTCCCCTGATTGACAGGCAGATTTTCCGCTTTAGTGAAGATGGGATGGTAA
ATGCAAGATTTGACTATAGCTATGACAACAGCTTTCGAGTGACCAGCATGCAGGGTGTGATCAATGAA
ACGCCACTGCCTATTGATCTGTATCAGTTTGATGACATTTCTGGCAAAGTTGAGCAGTTTGGAAAGTT
TGGAGTTATATATTATGATATTAACCAGATCATTTCTACAGCTGTAATGACCTATACGAAGCACTTTG
ATGCTCATGGCCGTATCAAGGAGATTCAATATGAGATATTCAGGTCGCTCATGTACTGGATTACAATT
CAGTATGATAACATGGGTCGGGTAACCAAGAGAGAGATTAAAATAGGGCCCTTTGCCAACACCACCAA
ATATGCTTATGAATATGATGTTGATGGACAGCTCCAAACAGTTTACCTCAATGAAAAGATAATGTGGC

GGTACAACTACGATCTGAATGGAAACCTCCATTTACTGAACCCAAGTAACAGTGCGCGTCTGACACCC
CTTCGCTATGACCTGCGAGACAGAATCACTCGACTGGGTGATGTTCAATATCGGTTGGATGAAGATGG
TTTCCTACGTCAAAGGGGCACGGAAATCTTTGAATATAGCTCCAAGGGGCTTCTAACTCGAGTTTACA
GTAAAGGCAGTGGCTGGACAGTGATCTACCGTTATGACGGCCTGGGAAGGCGTGTTTCTAGCAAAACC
AGTCTAGGACAGCACCTGCAGTTTTTTTATGCTGACTTAACTTATCCCACTAGGATTACTCATGTCTA
CAACCATTCGAGTTCAGAAATTACCTCCCTGTATTATGATCTCCAAGGACATCTTTTTGCCATGGAAA
TCAGCAGTGGGGATGAATTCTATATTGCATCGGATAACACAGGGACACCACTGGCTGTGTTCAGTAGC
AATGGGCTTATGCTGAAACAGATTCAGTACACTGCATATGGGGAAATCTATTTTGACTCTAATATTGA
CTTTCAACTGGTAATTGGATTTCATGGTGGCCTGTATGACCCACTCACCAAATTAATCCACTTTGGAG
AAAGAGATTATGACATTTTGGCAGGACGGTGGACAACACCTGACATAGAAATCTGGAAAAGAATTGGG
AAGGACCCAGCTCCTTTTAACTTGTACATGTTTAGGAATAACAACCCTGCAAGCAAAATCCATGACGT
GAAAGATTACATCACAGATGTTAACAGCTGGCTGGTGACATTTGGTTTCCATCTGCACAATGCTATTC
CTGGATTCCCTGTTCCCAAATTTGATTTAACAGAACCTTCTTACGAACTTGTGAAGAGTCAGCAGTGG
GATGATATACCGCCCATCTTCGGAGTCCAGCAGCAAGTGGCGCGGCAGGCCAAGGCCTTCCTGTCGCT

GGGGAAGATGGCCGAGGTGCAGGTGAGCCGGCGCCGGGCCGGCGGCGCGCAGTCCTGGCTGTGGTTCG
CCACGGTCAAGTCGCTGATCGGCAAGGGCGTCATGCTGGCCGTCAGCCAGGGCCGCGTGCAGACCAAC
GTGCTCAACATCGCCAACGAGGACTGCATCAAGGTGGCGGCCGTGCTCAACAACGCCTTCTACCTGGA
GAACCTGCACTTCACCATCGAGGGCAAGGACACGCACTACTTCATCAAGACCACCACGCCCGAGAGCG
ACCTGGGCACGCTGCGGTTGACCAGCGGCCGCAAGGCGCTGGAGAACGGCATCAACGTGACGGTGTCG
CAGTCCACCACGGTGGTGAACGGCAGGACGCGCAGGTTCGCGGACGTGGAGATGCAGTTCGGCGCGCT
GGCGCTGCACGTGCGCTACGGCATGACCCTGGACGAGGAGAAGGCGCGCATCCTGGAGCAGGCGCGGC
AGCGCGCGCTCGCCCGGGCCTGGGCGCGCGAGCAGCAGCGCGTGCGCGACGGCGAGGAGGGCGCGCGC
CTCTGGACGGAGGGCGAGAAGCGGCAGCTGCTGAGCGCCGGCAAGGTGCAGGGCTACGACGGGTACTA
CGTACTCTCGGTGGAGCAGTACCCCGAGCTGGCCGACAGCGCCAACAACATCCAGTTCCTGCGGCAGA
GCGAGATCGGCAGGAGG**TAA**CGCCCGGGCCGCGCCCGCCGAGCCGCTCACGCCCTGCCCACATTGTCC
TGTGGCACAACCCGAGTGGGACTCTCCAACGCCCAAGAGCCTTCCTCCCGGGGGAATGAGACTGCTGT
TACGACCCACACCCACACCGCGAAAACAAGGACCGCTTTTTTCCGAATGACCTTAAAGGTGATCGGCT
TTAACGAATATGTTTACATATGCATAGCGCTGCACTCAGTCGGACTGAACGTAGCCAGAGGAAAAAAA
AATCATCAAGGACAAAGGCCTCGACCTGTTGCGCTGGGCCGTCTGTTCCTTCTAGGCACTGTATTTAA
CTAACTTTA

NOV40s, CG55069-02 Protein Sequence	SEQ ID NO: 522	2721 aa	MW at 303489.1kD

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQGQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWG
RGVKSGRSSCLSSRSNSALTLTDTEHENKSDSENEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLN
RNSLTNRRNQSPAPPAALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYT
MASGSVYSPPTRPLPRNTLSRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQL
QQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQL
FIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAGRQARSVS
LHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFLGPDC
SRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACSSGYKGESCEEA
DCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEI
CSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIV
KDKIGYKEGCPGLCNSNGRCTLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCC
LQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRG
QVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFY
VMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLS
YLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYG
LSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQ
QPPVVSSIMGNGRRRSISCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLEL
RNKDFRHSSNPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEA
RCGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTSMHIS
QVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEYPVGKHAVQTTLESAT
AIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCKNDANCDCYQSGDGYAKDAKLSAP
SSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFYEVASPTDQELYIFDINGTHQYTVSLVTGDYL
YNFSYSNDNDITAVTDSNGNTLRIRRDPNRMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELVLFTY
HGNSGLLATKSDETGWTTFFDYDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSITSNL
SSIDSFYTMVQDQLRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGENGQNL
VEWRFRKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLWLPSS
KLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVLLLHSQRQYIFE
YDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEEGLLLQTAFLGTSRRVLFKY
RRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLIDRQIFRFSEDGMVNARFD
YSYDNSFRVTSMQGVINETPLPIDLYQFDDISGKVEQFGKFGVIYYDINQIISTAVMTYTKHFDAHGR
IKEIQYEIFRSLMYWITIQYDNMGRVTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWRYNYD
LNGNLHLLNPSNSARLTPLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYSKGSG
WTVIYRYDGLGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEISSGD
EFYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIHFGERDYD
ILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLVTFGFHLHNAIPGFPV
PKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMAEVQVSRRRAGGAQSWLWFATVKS
LIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNAFYLENLHFTIEGKDTHYFIKTTTPESDLGTL
RLTSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFGALALHVRYGMTLDEEKARILEQARQRALA
RAWAREQQRVRDGEEGARLWTEGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQSEIGR

R		
NOV40t, CG55069-05 DNA Sequence	SEQ ID NO: 523	810 bp
	ORF Start: at 7	ORF Stop: at 805

AAGCTTTGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCCAGG
ATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACTCCA
AGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGTATT
GACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCTTGTGCTTGCAACTCAGGATACAAAGG
AGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACGGGG
AATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGACCAG
TGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGGCCC
AGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGTGTC
GCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAGCAC
GGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGCTCA
CTATTTGGATAAGATAGTTAAAGAGGGTTGTCCTGGTCTGTGCAACAGCAATGGAAGATGTACCCTGG
ACCAAAATGGCTGGCATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTCGAC

NOV40t, CG55069-05 Protein Sequence	SEQ ID NO: 524	266 aa	MW at 27935.0kD

CPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDP
QCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCS
GHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGT
CKDGKCECSQGWNGEHCTIAHYLDKIVKEGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCD

NOV40u, CG55069-06 DNA Sequence	SEQ ID NO: 525	8204 bp
	ORF Start: ATG at 4	ORF Stop: TAA at 8179

AGTATGGATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAGAAGGAACGGCG
CTACACAAATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCACACAGAAGTCCTACAGTTCCAGCG
AGACATTGAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGAAGGATTTGGTT
CACAGAGAAGCAGACGAGTTCACTAGACAAGGACAGAATTTTACCCTAAGGCAGTTAGGAGTTTGTGA
ACCAGCAACTCGAAGAGGACTGGCATTTTGTGCGGAAATGGGGCTCCCTCACAGAGGTTACTCTATCA
GTGCAGGGTCAGATGCTGATACTGAAAATGAAGCAGTGATGTCCCCAGAGCATGCCATGAGACTTTGG
GGCAGGGGGGTCAAATCAGGCCGCAGCTCCTGCCTGTCAAGTCGGTCCAACTCAGCCCTCACCCTGAC
AGATACGGAGCACGAAAACAAGTCCGACAGTGAGAATGAGCAACCTGCAAGCAATCAAGGCCAGTCTA
CCCTGCAGCCCTTGCCGCCTTCCCATAAGCAGCACTCTGCACAGCATCATCCATCCATCACTTCTCTC
AACAGAAACTCCCTGACCAATAGAAGGAACCAGAGTCCGGCCCCGCCGGCTGCTTTGCCCGCCGAGCT
GCAAACCACACCCGAGTCCGTCCAGCTGCAGGACAGCTGGGTCCTTGGCAGTAATGTACCACTGGAAA
GCAGGCATTTCCTATTCAAAACAGGAACAGGTACAACGCCACTGTTCAGTACTGCAACCCCAGGATAC
ACAATGGCATCTGGCTCTGTTTATTCACCACCTACTCGGCCACTACCTAGAAACACCCTATCAAGAAG
TGCTTTTAAATTCAAGAAGTCTTCAAAGTACTGTAGCTGGAAATGCACTGCACTGTGTGCCGTAGGGG
TCTCGGTGCTCCTGGCAATACTCCTGTCTTATTTTATAGCAATGCATCTCTTTGGCCTCAACTGGCAG
CTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATACCATGCCAACAAACAC
TGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAAATAACACCATAGATT
CCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATCTTCTGGAGATCACAG
CTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGATGCATTGATTGGAGT
ATATGGCCGGAAGAAGTTACCGCCTTCCCATACTCAGTCCTCCCCCCAGTATGACTTCGTGGAGCTCC
TGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGGGCGGCAG
GCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGCATCTGGC
TTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCTGTGGTGG
AATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCCAGGATTT
CTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACTCCAAGGG
CCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGTATTGACC
CACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAGCTCAGGATACAAAGGAGAA
AGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACGGGGAATG
TCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGACCAGTGCT
CCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGGCCCAGAC
TGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGTGTCGCTG
TGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAGCACGGGA

CCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGCTCACTAT
TTGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGCAACAGCAATGG
AAGATGTACCCTGGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTG
ACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGGGATGGACTCATTGACTGCATG
GATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCA
GGACATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCA
GTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCA
TCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTT
CCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTG
GGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCA
TGGAATGTCTTTTATGTGATGGATACCCTAGTCATGGAGAAAGAAGAGAATGACATTCCCAGCTGTGA
TCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTT
CTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACA
GATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGAC
CCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCC
AAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAAT
CAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCT

GACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCT
GGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAAC
CAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGGCGCAGCATTTC
CTGCCCCAGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCTAGCTTGTGGGA
TCGATGGCAGTCTGTACGTAGGCGATTTCAACTACGTGCGGCGGATATTCCCTTCTGGAAATGTAACA
AGTGTCTTAGAACTAAGAAATAAAGATTTTAGACATAGCAGCAACCCAGCTCATAGATACTACCTTGC
AACGGATCCAGTCACGGGAGATCTGTACGTTTCTGACACAAACACCCGCAGAATTTATCGCCCAAAGT
CACTTACGGGGGCAAAAGACTTGACTAAAAATGCAGAAGTCGTCGCAGGGACAGGGGAGCAATGCCTT
CCGTTTGACGAGGCGAGATGTGGGGATGGAGGGAAGGCCGTGGAAGCCACACTCATGAGTCCCAAAGG
AATGGCAGTTGATAAGAATGGATTAATCTACTTTGTTGATGGAACCATGATTAGGAAAGTTGACCAAA
ATGGAATCATATCAACTCTTCTGGGCTCTAACGATTTGACTTCAGCCAGACCTTTAACTTGTGACACC
AGCATGCACATCAGCCAGGTACGTCTGGAATGGCCCACTGACCTAGCCATTAACCCTATGGATAACTC
CATTTATGTCCTGGATAATAATGTAGTTTTACAGATCACTGAAAATCGTCAAGTTCGCATTGCTGCTG
GACGGCCCATGCACTGTCAGGTTCCCGGAGTGGAATATCCTGTGGGGAAGCACGCGGTGCAGACAACA
CTGGAATCAGCCACTGCCATTGCTGTGTCCTACAGTGGGGTCCTGTACATTACTGAAACTGATGAGAA
GAAAATTAACCGGATAAGGCAGGTCACAACAGATGGAGAAATCTCCTTAGTGGCCGGAATACCTTCAG
AGTGTGACTGCAAAAATGATGCCAACTGTGACTGTTACCAGAGTGGAGATGGCTACGCCAAGGATGCC
AAACTCAGTGCCCCATCCTCCCTGGCTGCTTCTCCAGATGGTACACTGTATATTGCAGATCTAGGGAA
TATCCGGATCCGGGCTGTGTCAAAGAATAAGCCTTTACTTAACTCTATGAACTTCTATGAAGTTGCGT
CTCCAACTGATCAAGAACTCTACATCTTTGACATCAATGGTACTCACCAATATACTGTAAGTTTAGTC
ACTGGTGATTACCTTTACAATTTTAGCTACAGCAATGACAATGATATTACTGCTGTGACAGACAGCAA
TGGCAACACCCTTAGAATTAGACGGGACCCAAATCGCATGCCAGTTCGAGTGGTGTCTCCTGATAACC
AAGTGATATGGTTGACAATAGGAACAAATGGATGTTTGAAAGGCATGACTGCTCAAGGACTGGAATTA
GTTTTGTTTACTTACCATGGCAATAGTGGCCTTTTAGCCACTAAAAGTGATGAAACTGGATGGACAAC
GTTTTTTGACTATGACAGTGAAGGTCGTCTGACAAATGTTACGTTTCCAACTGGAGTGGTCACAAACC
TGCATGGGGACATGGACAAGGCTATCACAGTGGACATTGAGTCATCTAGCCGAGAAGAAGATGTCAGC
ATCACTTCAAATCTGTCCTCGATCGATTCTTTCTACACCATGGTTCAAGATCAGTTAAGAAACAGCTA
CCAGATTGGTTATGACGGCTCCCTCAGAATTATCTACGCCAGTGGCCTGGACTCACACTACCAAACAG
AGCCGCACGTTCTGGCTGGCACCGCTAATCCGACGGTTGCCAAAAGAAACATGACTTTGCCTGGCGAG
AACGGTCAAAACTTGGTGGAATGGAGATTCCGAAAAGAGCAAGCCCAAGGGAAAGTCAATGTCTTTGG
CCGCAAGCTCAGGGTTAATGGCAGAAACCTCCTTTCAGTTGACTTTGATCGAACAACAAAGACAGAAA
AGATCTATGACGACCACCGTAAATTTCTACTGAGGATCGCCTACGACACGTCTGGGCACCCGACTCTC
TGGCTGCCAAGCAGCAAGCTGATGGCCGTCAATGTCACCTATTCATCCACAGGTCAAATTGCCAGCAT
CCAGCGAGGCACCACTAGCGAGAAAGTAGATTATGACGGACAGGGGAGGATCGTGTCTCGGGTCTTTG
CTGATGGTAAAACATGGAGTTACACATATTTAGAAAAGTCCATGGTTCTTCTGCTTCATAGCCAGCGG
CAGTACATCTTCGAATACGATATGTGGGACCGCCTGTCTGCCATCACCATGCCCAGTGTGGCTCGCCA
CACCATGCAGACCATCCGATCCATTGGCTACTACCGCAACATATACAACCCCCCGGAAAGCAACGCCT
CCATCATCACGGACTACAACGAGGAAGGGCTGCTTCTACAAACAGCTTTCTTGGGTACAAGTCGGAGG
GTCTTATTCAAATACAGAAGGCAGACTAGGCTCTCAGAAATTTTATATGATAGCACAAGAGTCAGTTT
TACCTATGATGAAACAGCAGGAGTCCTAAAGACAGTAAACCTCCAGAGTGATGGTTTTATTTGCACCA
TTAGATACAGGCAAATTGGTCCCCTGATTGACAGGCAGATTTTCCGCTTTAGTGAAGATGGGATGGTA
AATGCAAGATTTGACTATAGCTATGACAACAGCTTTCGAGTGACCAGCATGCAGGGTGTGATCAATGA

AACGCCACTGCCTATTGATCTGTATCAGTTTGATGACATTTCTGGCAAAGTTGAGCAGTTTGGAAAGT
TTGGAGTTATATATTATGATATTAACCAGATCATTTCTACAGCTGTAATGACCTATACGAAGCACTTT
GATGCTCATGGCCGTATCAAGGAGATTCAATATGAGATATTCAGGTCGCTCATGTACTGGATTACAAT
TCAGTATGATAACATGGGTCGGGTAACCAAGAGAGAGATTAAAATAGGGCCCTTTGCCAACACCACCA
AATATGCTTATGAATATGATGTTGATGGACAGCTCCAAACAGTTTACCTCAATGAAAAGATAATGTGG
CGGTACAACTACGATCTGAATGGAAACCTCCATTTACTGAACCCAAGTAACAGTGCGCGTCTGACACC
CCTTCGCTATGACCTGCGAGACAGAATCACTCGACTGGGTGATGTTCAATATCGGTTGGATGAAGATG

GTTTCCTACGTCAAAGGGGCACGGAAATCTTTGAATATAGCTCCAAGGGGCTTCTAACTCGAGTTTAC
AGTAAAGGCAGTGGCTGGACAGTGATCTACCGTTATGACGGCCTGGGAAGGCGTGTTTCTAGCAAAAC
CAGTCTAGGACAGCACCTGCAGTTTTTTTATGCTGACTTAACTTATCCCACTAGGATTACTCATGTCT
ACAACCATTCGAGTTCAGAAATTACCTCCCTGTATTATGATCTCCAAGGACATCTTTTTGCCATGGAA
ATCAGCAGTGGGGATGAATTCTATATTGCATCGGATAACACAGGGACACCACTGGCTGTGTTCAGTAG
CAATGGGCTTATGCTGAAACAGATTCAGTACACTGCATATGGGGAAATCTATTTTGACTCTAATATTG
ACTTTCAACTGGTAATTGGATTTCATGGTGGCCTGTATGACCCACTCACCAAATTAATCCACTTTGGA
GAAAGAGATTATGACATTTTGGCAGGACGGTGGACAACACCTGACATAGAAATCTGGAAAAGAATTGG
GAAGGACCCAGCTCCTTTTAACTTGTACATGTTTAGGAATAACAACCCTGCAAGCAAAATCCATGACG
TGAAAGATTACATCACAGATGTTAACAGCTGGCTGGTGACATTTGGTTTCCATCTGCACAATGCTATT
CCTGGATTCCCTGTTCCCAAATTTGATTTAACAGAACCTTCTTACGAACTTGTGAAGAGTCAGCAGTG
GGATGATATACCGCCCATCTTCGGAGTCCAGCAGCAAGTGGCGCGGCAGGCCAAGGCCTTCCTGTCGC
TGGGGAAGATGGCCGAGGTGCAGGTGAGCCGGCGCCGGGCCGGCGGCGCGCAGTCCTGGCTGTGGTTC
GCCACGGTCAAGTCGCTGATCGGCAAGGGCGTCATGCTGGCCGTCAGCCAGGGCCGCGTGCAGACCAA
CGTGCTCAACATCGCCAACGAGGACTGCATCAAGGTGGCGGCCGTGCTCAACAACGCCTTCTACCTGG
AGAACCTGCACTTCACCATCGAGGGCAAGGACACGCACTACTTCATCAAGACCACCACGCCCGAGAGC
GACCTGGGCACGCTGCGGTTGACCAGCGGCCGCAAGGCGCTGGAGAACGGCATCAACGTGACGGTGTC
GCAGTCCACCACGGTGGTGAACGGCAGGACGCGCAGGTTCGCGGACGTGGAGATGCAGTTCGGCGCGC
TGGCGCTGCACGTGCGCTACGGCATGACCCTGGACGAGGAGAAGGCGCGCATCCTGGAGCAGGCGCGG
CAGCGCGCGCTCGCCCGGGCCTGGGCGCGCGAGCAGCAGCGCGTGCGCGACGGCGAGGAGGGCGCGCG
CCTCTGGACGGAGGGCGAGAAGCGGCAGCTGCTGAGCGCCGGCAAGGTGCAGGGCTACGACGGGTACT
ACGTACTCTCGGTGGAGCAGTACCCCGAGCTGGCCGACAGCGCCAACAACATCCAGTTCCTGCGGCAG
AGCGAGATCGGCAGGAGG**TAA**CGCCCGGGCCGCGCCCGCCGAGC

NOV40u, CG55069-06 Protein Sequence	SEQ ID NO: 526	2725 aa	MW at 303959.6kD

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQGQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWG
RGVKSGRSSCLSSRSNSALTLTDTEHENKSDSENEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLN
RNSLTNRRNQSPAPPAALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYT
MASGSVYSPPTRPLPRNTLSRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQL
QQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQL
FIDQPQFLKFNISLQKDALIGVYGRKKLPPSHTQSSPQYDFVELLDGSRLIAREQRSLLETERAGRQA
RSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFL
GPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACSSGYKGES
CEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYL
DKIVKDKIGYKEGCPGLCNSNGRCTLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMD
PDCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLAS
VIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPW
NVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTD
LKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQ
KVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQ
FISQQPPVVSSIMGNGRRRSISCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTS
VLELRNKDFRHSSNPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLP
FDEARCGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTS
MHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEYPVGKHAVQTTL
ESATAIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCKNDANCDCYQSGDGYAKDAK
LSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFYEVASPTDQELYIFDINGTHQYTVSLVT
GDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPNRMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELV
LFTYHGNSGLLATKSDETGWTTFFDYDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSI
TSNLSSIDSFYTMVQDQLRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGEN

GQNLVEWRFRKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLW
LPSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVLLLHSQRQ
YIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEEGLLLQTAFLGTSRRV
LFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLIDRQIFRFSEDGMVN
ARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISGKVEQFGKFGVIYYDINQIISTAVMTYTKHFD
AHGRIKEIQYEIFRSLMYWITIQYDNMGRVTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWR
YNYDLNGNLHLLNPSNSARLTPLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYS
KGSGWTVIYRYDGLGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEI
SSGDEFYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIHFGE
RDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLVTFGFHLHNAIP
GFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMAEVQVSRRRAGGAQSWLWFA
TVKSLIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNAFYLENLHFTIEGKDTHYFIKTTTPESD
LGTLRLTSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFGALALHVRYGMTLDEEKARILEQARQ
RALARAWAREQQRVRDGEEGARLWTEGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQS
EIGRR

NOV40v, CG55069-07 DNA Sequence	SEQ ID NO: 527	1833 bp
	ORF Start: at 7	ORF Stop: at 1828

AAGCTTGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGC
CATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCG
ATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATC
ATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCT
TATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCA
TCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTAC
CCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTC
TCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATG
TCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGT
GGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGA
AGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGA
AACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCT
ATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTG
GTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAG
TCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTG
TGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATT
AGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCA
TCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGGCGCAGCATTTCCTGCCCC
AGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCTAGCTTGTGGGATCGATGG
CAGTCTGTACGTAGGCGATTTCAACTATGTGCGGCGGATATTCCCTTCTGGAAATGTAACAAGTGTCT
TAGAACTAAGCAGCAACCCAGCTCATAGATACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTAC
GTTTCTGACACAAACACCCGCAGAATTTATCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAA
AAATGCAGAAGTCGTCGCAGGGACAGGGGAGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATG
GAGGGAAGGCCGTGGAAGCCACACTCATGAGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATC
TACTTTGTTGATGGAACCATGATTAGGAAAGTTGACCAAAATGGAATCATATCAACTCTTCTGGGTTC
TAACGATTTGACTTCAGCCAGACCTTTAACTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGG
AATGGCCCACTGACCTAGCCATTAACCCTATGGATAACTCCATTTATGTCCTGGATAATGTCGAC

NOV40v, CG55069-07 Protein Sequence	SEQ ID NO: 528	607 aa	MW at 66606.6kD

DQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQPYCRGLPDPQDIIS
QSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHYPE
YGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGF
VRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSII
PFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWE
KRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSISCPSC
NGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELSSNPAHRYYLATDPVTGDLYVS
DTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEARCGDGGKAVEATLMSPKGMAVDKNGLIYF
VDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIYVLDN

NOV40w, CG55069-08 DNA Sequence	SEQ ID NO: 529	8487 bp
	ORF Start: ATG at 299	ORF Stop: TAA at 8138

ACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTCCCATTTGACAGAAAAAGGCAGTAAACGGGGA
ATCTCTTTTTTTGAATAAAGAAGAAGAAGAAATAAAGTACCTGTCATCTTGACAAGTGGCGGAGCGGA
GGAGTCAAGGATTATAAATGATCACAGCCAGGTCCAGCTCGCCCCGTGATTGGGCTCTCCCGCGATCT
GCACCGGGGGAAGCGCATGAGAGGCCAATGAGACTTGAACCCTGAGCCTAAGTTGTCACCAGCAGGAC
TGATGTGCACACAGAAGGAATGAAGTATGGATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAG
AGCAGACGAGAGAAGGAACGGCGCTACACAAATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCAC
ACAGAAGTCCTACAGTTCCAGCGAGACATTGAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACG
GCAACAGAGTGAAGGATTTGGTTCACAGAGAAGCAGACGAGTTCACTAGACAAGAGCAACCTGCAAGC
AATCAAGGCCAGTCTACCCTGCAGCCCTTGCCGCCTTCCCATAAGCAGCACTCTGCACAGCATCATCC
ATCCATCACTTCTCTCAACAGAAACTCCCTGACCAATAGAAGGAACCAGAGTCCGGCCCCGCCGGCTG
CTTTGCCCGCCGAGCTGCAAACCACACCCGAGTCCGTCCAGCTGCAGGACAGCTGGGTCCTTGGCAGT
AATGTACCACTGGAAAGCAGGCATTTCCTATTCAAAACAGGAACAGGTACAACGCCACTGTTCAGTAC
TGCAACCCCAGGATACACAATGGCATCTGGCTCTGTTTATTCACCACCTACTCGGCCACTACCTAGAA
ACACCCTATCAAGAAGTGCTTTTAAATTCAAGAAGTCTTCAAAGTACTGTAGCTGGAAATGCACTGCA
CTGTGTGCCGTAGGGGTCTCGGTGCTCCTGGCAATACTCCTGTCTTATTTTATAGCAATGCATCTCTT
TGGCCTCAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATA
CCATGCCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAA
AATAACACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGAT
CTTCTGGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGG
ATGCATTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTGGAG
CTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGGGCG
GCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGCATC
TGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCTGTG
GTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCCAGG
ATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACTCCA
AGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGTATT
GACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAACTCAGGATACAAAGG
AGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACGGGG
AATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGACCAG
TGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGGCCC
AGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGTGTC
GCTGTGAAGAAGGCTGGACGGGCCCAACCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAGCAC
GGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCATGGCTGGAATGGAGAGCACTGCACTATCGAGGG
TTGTCCTGGTCTGTGCAACAGCAATGGAAGATGTACCCTGGACCAAAATGGCTGGCATTGTGTGTGCC
AGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAAT
GAAGGAGATGGACTCATTGACTGCATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCC
CTATTGTCGGGGACTGCCGGATCCTCAGGACATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAG
CTGCCAAATCCTTTTATGATCGAATCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGA
GAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCC
ACTTATTGGAGTAAATGTCTCGTTTTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACG
GAATGTTTGACTTGGTGGCAAATGGTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTC
ACTCAGTATCATACTGTGTGGATTCCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGGAGAA
AGAAGAGAATGACATTCCCAGCTGTGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCAT
CACCTTTATCCACCTTTTTCAGATCTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTC
CACGAGGAAACTACAATTCCAGGAACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTA
TAAGTCAGTTCTCAAGATCACCATGACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTA
TGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTC
ATATGGGATAAAACAGATGCATATAATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGG
ATATGAGTATGAGTCGTGTTTGGACCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATG

AATTGGATGCGTCCAACATGGGTGGCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGT
ATACTGTACAAGGGAAACGGGGAAAACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCAT
GGGCAATGGGCGAAGGCGCAGCATTTCCTGCCCCAGTTGCAATGGTCAAGCTGATGGTAACAAGTTAC
TGGCCCCAGTGGCGCTAGCTTGTGGGATCGATGGCAGTCTGTACGTAGGCGATTTCAACTACGTGCGG
CGGATATTCCCTTCTGGAAATGTAACAAGTGTCTTAGAACTAAGAAATAAAGATTTTAGACATAGCAG
CAACCCAGCTCATAGATACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTACGTTTCTGACACAA

ACACCCGCAGAATTTATCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAAAAATGCAGAAGTC
GTCGCAGGGACAGGGGAGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATGGAGGGAAGGCCGT
GGAAGCCACACTCATGAGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATCTACTTTGTTGATG
GAACCATGATTAGGAAAGTTGACCAAAATGGAATCATATCAACTCTTCTGGGCTCTAACGATTTGACT
TCAGCCAGACCTTTAACTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGGAATGGCCCACTGA
CCTAGCCATTAACCCTATGGATAACTCCATTTATGTCCTGGATAATAATGTAGTTTTACAGATCACTG
AAAATCGTCAAGTTCGCATTGCTGCTGGACGGCCCATGCACTGTCAGGTTCCCGGAGTGGAATATCCT
GTGGGGAAGCACGCGGTGCAGACAACACTGGAATCAGCCACTGCCATTGCTGTGTCCTACAGTGGGGT
CCTGTACATTACTGAAACTGATGAGAAGAAAATTAACCGGATAAGGCAGGTCACAACAGATGGAGAAA
TCTCCTTAGTGGCCGGAATACCTTCAGAGTGTGACTGCAAAAATGATGCCAACTGTGACTGTTACCAG
AGTGGAGATGGCTACGCCAAGGATGCCAAACTCAGTGCCCCATCCTCCCTGGCTGCTTCTCCAGATGG
TACACTGTATATTGCAGATCTAGGGAATATCCGGATCCGGGCTGTGTCAAAGAATAAGCCTTTACTTA
ACTCTATGAACTTCTATGAAGTTGCGTCTCCAACTGATCAAGAACTCTACATCTTTGACATCAATGGT
ACTCACCAATATACTGTAAGTTTAGTCACTGGTGATTACCTTTACAATTTTAGCTACAGCAATGACAA
TGATATTACTGCTGTGACAGACAGCAATGGCAACACCCTTAGAATTAGACGGGACCCAAATCGCATGC
CAGTTCGAGTGGTGTCTCCTGATAACCAAGTGATATGGTTGACAATAGGAACAAATGGATGTTTGAAA
GGCATGACTGCTCAAGGACTGGAATTAGTTTTGTTTACTTACCATGGCAATAGTGGCCTTTTAGCCAC
TAAAAGTGATGAAACTGGATGGACAACGTTTTTTGACTATGACAGTGAAGGTCGTCTGACAAATGTTA
CGTTTCCAACTGGAGTGGTCACAAACCTGCATGGGGACATGGACAAGGCTATCACAGTGGACATTGAG
TCATCTAGCCGAGAAGAAGATGTCAGCATCACTTCAAATCTGTCCTCGATCGATTCTTTCTACACCAT
GGTTCAAGATCAGTTAAGAAACAGCTACCAGATTGGTTATGACGGCTCCCTCAGAATTATCTACGCCA
GTGGCCTGGACTCACACTACCAAACAGAGCCGCACGTTCTGGCTGGCACCGCTAATCCGACGGTTGCC
AAAAGAAACATGACTTTGCCTGGCGAGAACGGTCAAAACTTGGTGGAATGGAGATTCCGAAAAGAGCA
AGCCCAAGGGAAAGTCAATGTCTTTGGCCGCAAGCTCAGGGTTAATGGCAGAAACCTCCTTTCAGTTG
ACTTTGATCGAACAACAAAGACAGAAAAGATCTATGACGACCACCGTAAATTTCTACTGAGGATCGCC
TACGACACGTCTGGGCACCCGACTCTCTGGCTGCCAAGCAGCAAGCTGATGGCCGTCAATGTCACCTA
TTCATCCACAGGTCAAATTGCCAGCATCCAGCGAGGCACCACTAGCGAGAAAGTAGATTATGACGGAC
AGGGGAGGATCGTGTCTCGGGTCTTTGCTGATGGTAAAACATGGAGTTACACATATTTAGAAAAGTCC
ATGGTTCTTCTGCTTCATAGCCAGCGGCAGTACATCTTCGAATACGATATGTGGGACCGCCTGTCTGC
CATCACCATGCCCAGTGTGGCTCGCCACACCATGCAGACCATCCGATCCATTGGCTACTACCGCAACA
TATACAACCCCCCGGAAAGCAACGCCTCCATCATCACGGACTACAACGAGGAAGGGCTGCTTCTACAA
ACAGCTTTCTTGGGTACAAGTCGGAGGGTCTTATTCAAATACAGAAGGCAGACTAGGCTCTCAGAAAT
TTTATATGATAGCACAAGAGTCAGTTTTACCTATGATGAAACAGCAGGAGTCCTAAAGACAGTAAACC
TCCAGAGTGATGGTTTTATTTGCACCATTAGATACAGGCAAATTGGTCCCCTGATTGACAGGCAGATT
TTCCGCTTTAGTGAAGATGGGATGGTAAATGCAAGATTTGACTATAGCTATGACAACAGCTTTCGAGT
GACCAGCATGCAGGGTGTGATCAATGAAACGCCACTGCCTATTGATCTGTATCAGTTTGATGACATTT
CTGGCAAAGTTGAGCAGTTTGGAAAGTTTGGAGTTATATATTATGATATTAACCAGATCATTTCTACA
GCTGTAATGACCTATACGAAGCACTTTGATGCTCATGGCCGTATCAAGGAGATTCAATATGAGATATT
CAGGTCGCTCATGTACTGGATTACAATTCAGTATGATAACATGGGTCGGGTAACCAAGAGAGAGATTA
AAATAGGGCCCTTTGCCAACACCACCAAATATGCTTATGAATATGATGTTGATGGACAGCTCCAAACA
GTTTACCTCAATGAAAAGATAATGTGGCGGTACAACTACGATCTGAATGGAAACCTCCATTTACTGAA
CCCAAGTAACAGTGCGCGTCTGACACCCCTTCGCTATGACCTGCGAGACAGAATCACTCGACTGGGTG
ATGTTCAATATCGGTTGGATGAAGATGGTTTCCTACGTCAAAGGGGCACGGAAATCTTTGAATATAGC

TCCAAGGGGCTTCTAACTCGAGTTTACAGTAAAGGCAGTGGCTGGACAGTGATCTACCGTTATGACGG
CCTGGGAAGGCGTGTTTCTAGCAAAACCAGTCTAGGACAGCACCTGCAGTTTTTTTATGCTGACTTAA
CTTATCCCACTAGGATTACTCATGTCTACAACCATTCGAGTTCAGAAATTACCTCCCTGTATTATGAT
CTCCAAGGACATCTTTTTGCCATGGAAATCAGCAGTGGGGATGAATTCTATATTGCATCGGATAACAC
AGGGACACCACTGGCTGTGTTCAGTAGCAATGGGCTTATGCTGAAACAGATTCAGTACACTGCATATG
GGGAAATCTATTTTGACTCTAATATTGACTTTCAACTGGTAATTGGATTTCATGGTGGCCTGTATGAC
CCACTCACCAAATTAATCCACTTTGGAGAAAGAGATTATGACATTTTGGCAGGACGGTGGACAACACC
TGACATAGAAATCTGGAAAAGAATTGGGAAGGACCCAGCTCCTTTTAACTTGTACATGTTTAGGAATA
ACAACCCTGCAAGCAAAATCCATGACGTGAAAGATTACATCACAGATGTTAACAGCTGGCTGGTGACA
TTTGGTTTCCATCTGCACAATGCTATTCCTGGATTCCCTGTTCCCAAATTTGATTTAACAGAACCTTC
TTACGAACTTGTGAAGAGTCAGCAGTGGGATGATATACCGCCCATCTTCGGAGTCCAGCAGCAAGTGG
CGCGGCAGGCCAAGGCCTTCCTGTCGCTGGGGAAGATGGCCGAGGTGCAGGTGAGCCGGCGCCGGGCC
GGCGGCGCGCAGTCCTGGCTGTGGTTCGCCACGGTCAAGTCGCTGATCGGCAAGGGCGTCATGCTGGC
CGTCAGCCAGGGCCGCGTGCAGACCAACGTGCTCAACATCGCCAACGAGGACTGCATCAAGGTGGCGG
CCGTGCTCAACAACGCCTTCTACCTGGAGAACCTGCACTTCACCATCGAGGGCAAGGACACGCACTAC
TTCATCAAGACCACCACGCCCGAGAGCGACCTGGGCACGCTGCGGTTGACCAGCGGCCGCAAGGCGCT

GGAGAACGGCATCAACGTGACGGTGTCGCAGTCCACCACGGTGGTGAACGGCAGGACGCGCAGGTTCG
CGGACGTGGAGATGCAGTTCGGCGCGCTGGCGCTGCACGTGCGCTACGGCATGACCCTGGACGAGGAG
AAGGCGCGCATCCTGGAGCAGGCGCGGCAGCGCGCGCTCGCCCGGGCCTGGGCGCGCGAGCAGCAGCG
CGTGCGCGACGGCGAGGAGGGCGCGCGCCTCTGGACGGAGGGCGAGAAGCGGCAGCTGCTGAGCGCCG
GCAAGGTGCAGGGCTACGACGGGTACTACGTACTCTCGGTGGAGCAGTACCCCGAGCTGGCCGACAGC
GCCAACAACATCCAGTTCCTGCGGCAGAGCGAGATCGGCAGGAGG**TAA**CGCCCGGGCCGCGCCCGCCG
AGCCGCTCACGCCCTGCCCACATTGTCCTGTGGCACAACCCGAGTGGGACTCTCCAACGCCCAAGAGC
CTTCCTCCCGGGGGAATGAGACTGCTGTTACGACCCACACCCACACCGCGAAAACAAGGACCGCTTTT
TTCCGAATGACCTTAAAGGTGATCGGCTTTAACGAATATGTTTACATATGCATAGCGCTGCACTCAGT
CGGACTGAACGTAGCCAGAGGAAAAAAAAATCATCAAGGACAAAGGCCTCGACCTGTTGCGCTGGGCC
GTCTGTTCCTTCTAGGCACTGTATTTAACTAACTTTAAAAAAAAAAAAAAAAAAG

NOV40w, CG55069-08 Protein Sequence	SEQ ID NO: 530	2613 aa	MW at 291899.4kD

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLNRNSLTNRRNQSPAPPAALPAELQTT
PESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYTMASGSVYSPPTRPLPRNTLSRSAFK
FKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQLQQTENDTFENGKVNSDTMPTNTVSL
PSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGR
KGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKN
AEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSG
WKGTECDVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPGWG
GSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGTCRCEEGWTGP
TCNQRACHPRCAEHGTCKDGKCECSHGWNGEHCTIEGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGC
DVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRI
SFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANG
GASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRS
SPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLF
QKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGG
WTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSISCPSCNGQADGNKLLAPVALACG
IDGSLYVGDFNYVRRIFPSGNVTSVLELRNKDFRHSSNPAHRYYLATDPVTGDLYVSDTNTRRIYRPK
SLTGAKDLTKNAEVVAGTGEQCLPFDEARCGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQ
NGIISTLLGSNDLTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAA
GRPMHCQVPGVEYPVGKHAVQTTLESATAIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPS
ECDCKNDANCDCYQSGDGYAKDAKLSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFYEVA
SPTDQELYIFDINGTHQYTVSLVTGDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPNRMPVRVVSPDN
QVIWLTIGTNGCLKGMTAQGLELVLFTYHGNSGLLATKSDETGWTTFFDYDSEGRLTNVTFPTGVVTN
LHGDMDKAITVDIESSSREEDVSITSNLSSIDSFYTMVQDQLRNSYQIGYDGSLRIIYASGLDSHYQT
EPHVLAGTANPTVAKRNMTLPGENGQNLVEWRFRKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTE
KIYDDHRKFLLRIAYDTSGHPTLWLPSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVF
ADGKTWSYTYLEKSMVLLLHSQRQYIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNA
SIITDYNEEGLLLQTAFLGTSRRVLFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICT
IRYRQIGPLIDRQIFRFSEDGMVNARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISGKVEQFGK
FGVIYYDINQIISTAVMTYTKHFDAHGRIKEIQYEIFRSLMYWITIQYDNMGRVTKREIKIGPFANTT
KYAYEYDVDGQLQTVYLNEKIMWRYNYDLNGNLHLLNPSNSARLTPLRYDLRDRITRLGDVQYRLDED
GFLRQRGTEIFEYSSKGLLTRVYSKGSGWTVIYRYDGLGRRVSSKTSLGQHLQFFYADLTYPTRITHV
YNHSSSEITSLYYDLQGHLFAMEISSGDEFYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNI
DFQLVIGFHGGLYDPLTKLIHFGERDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHD
VKDYITDVNSWLVTFGFHLHNAIPGFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLS
LGKMAEVQVSRRRAGGAQSWLWFATVKSLIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNAFYL
ENLHFTIEGKDTHYFIKTTTPESDLGTLRLTSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFGA
LALHVRYGMTLDEEKARILEQARQRALARAWAREQQRVRDGEEGARLWTEGEKRQLLSAGKVQGYDGY
YVLSVEQYPELADSANNIQFLRQSEIGRR

NOV40x, CG55069-10 DNA Sequence	SEQ ID NO: 531	2519 bp
	ORF Start: at 7	ORF Stop: at 2488

GGTACCAACTGGCAGCTACAGCAGACTGAAAATGACGCATTTGAGAATGGAAAAGTGAATTCTGATAC
CATGCCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAA
ATAACACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATC

TTCTGGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGA
TGCATTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTGGAGC
TCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGGGCGG
CAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGCATCT
GGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCTGTGG
TGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCCAGGA
TTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACTCCAA
GGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGTATTG
ACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAACTCAGGATACAAAGGA
GAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACGGGGA
ATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGACCAGT
GCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGGCCCA
GACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGTGCCG
CTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAGCACG
GGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGCTCAC
TATTTGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGCAACAGCAA
TGGAAGATGTACCCTGGACCAAAATGGCTGGCATTGTGCGTGCCAGCCTGGATGGAGAGGAGCAGGCT
GTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGACTGC
ATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCC
TCAGGGCATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAA
TCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTT
GCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTT
TTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATG
GTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATT
CCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTG
TGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGAT
CTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGA
ACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCAT
GACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCT
TCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATAT
AATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGA
CCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTG
GCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAA
AACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCGTCGACCCAGCAGCCTCCAGTCGTGAGTA
GCG

NOV40x, CG55069-10 Protein Sequence	SEQ ID NO: 532	827 aa	MW at 90474.8kD

NWQLQQTENDAFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFW
RSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAGRQA
RSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFL
GPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACNSGYKGES
CEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYL
DKIVKDKIGYKEGCPGLCNSNGRCTLDQNGWHCACQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMD
PDCCLQSSCQNQPYCRGLPDPQGIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLAS
VIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPW
NVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTD
LKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQ
KVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQ
FISQQPPVVSS

NOV40y, CG55069-11 DNA Sequence	SEQ ID NO: 533	2482 bp
	ORF Start: at 11	ORF Stop: at 2474

CACCTCGCGAAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTG
ATACCATGCCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAA
GAAAATAACACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGG
GATCTTCTGGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGA
AGGATGCATTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTG

GAGCTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGG
GCGGCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGC
ATCTGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCT
GTGGTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCC
AGGATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACT
CCAAGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGT
ATTGACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCTTGTGCTTGCAACTCAGGATACAA
AGGAGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACG
GGGAATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGAC
CAGTGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGG
CCCAGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGT
GTCGCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAG
CACGGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGC
TCACTATTTGGATAAGATAGTTAAAGAGGGTTGTCCTGGTCTGTGCAACAGCAATGGAAGATGTACCC
TGGACCAAAATGGCTGGCATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGCCATG
GAGACTCTTTGCACAGATAGTAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCGATTG
CTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATCATTA
GCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCTTATA
GGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCATCAG
AGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTACCCAG
AATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTCTCTA
ACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATGTCTT
TTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGTGGAT
TCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGAAGAC
AGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGAAACT
CTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCTATTA
TTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTGGTTT
CCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAGTCTA
TGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTGTGGG
AAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATTAGAT
AAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCATCTC
CCAGCAGCCTCCAGTCGTGAGTAGCCTCGAGGGC

NOV40y, CG55069-11 Protein Sequence	SEQ ID NO: 534	821 aa	MW at 89886.1kD

NWQLQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFW
RSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAGRQA
RSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFL
GPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACNSGYKGES
CEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYL
DKIVKEGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQ
SSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQV
LTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVM
DTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYL
SSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLS
EAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQP
PVVSS

NOV40z, CG55069-12 DNA Sequence	SEQ ID NO: 535	2500 bp
	ORF Start: at 11	ORF Stop: at 2492

CACCAAGCTTAACTGGCAGCTACAGCAGACTGAAAATGACGCATTTGAGAATGGAAAAGTGAATTCTG
ATACCATGCCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAA
GAAAATAACACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGG
GATCTTCTGGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGA
AGGATGCATTGATTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTG
GAGCTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGG
GCGGCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATCTGGC
ATCTGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCT

GTGGTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCC
AGGATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACT
CCAAGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGT
ATTGACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAACTCAGGATACAA
AGGAGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACG
GGGAATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGAC
CAGTGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGG
CCCAGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGT
GCCGCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAG
CACGGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGC
TCACTATTTGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGCAACA
GCAATGGAAGATGTACCCTGGACCAAAATGGCTGGCATTGTGCGTGCCAGCCTGGATGGAGAGGAGCA
GGCTGTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGAGATGGACTCATTGA
CTGCATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGG
ATCCTCAGGGCATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGAT
CGAATCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAG
CCTTGCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCT
CGTTTTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCA
AATGGTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTGTGTG
GATTCCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCA
GCTGTGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTC
AGATCTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAATTCC
AGGAACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCA
CCATGACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGA
CTCTTCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAGATGC
ATATAATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTT
TGGACCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATG
GGTGGCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAAACGG
GGAAAACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCCTCGAGGGC

NOV40z, CG55069-12 Protein Sequence	SEQ ID NO: 536	827 aa	MW at 90474.8kD

NWQLQQTENDAFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFW
RSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSRLIAREQRSLLETERAGRQA
RSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFL
GPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACNSGYKGES
CEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYL
DKIVKDKIGYKEGCPGLCNSNGRCTLDQNGWHCACQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMD
PDCCLQSSCQNQPYCRGLPDPQGIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLAS
VIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPW
NVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTD
LKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQ
KVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQ
FISQQPPVVSS

NOV40aa, CG55069-13 DNA Sequence	SEQ ID NO: 537	2541 bp
	ORF Start: ATG at 16	ORF Stop: TAG at 2530

CACCGCGGCCGCACCATGCTGCCCGGTTTGGCACTGCTCCTGCTGGCCGCCTGGACGGCTCGGGCGAA
CTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATACCATGCCAA
CAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAAATAACACC
ATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATCTTCTGGAG
ATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGATGCATTGA
TTGGAGTATATGGCCGGAAAGGCTTACCGCCTTCCCATACTCAGTATGACTTCGTGGAGCTCCTGGAT
GGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAGCCGGGCGGCAGGCGAG
ATCCGTCAGCCTTCATGAGGCCGGCTCTATCCAGTACTTGGATTCTGGAATCTGGCATCTGGCTTTTT
ATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGAGTCTGTGGTGGAATGT
CCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTTTTCCAGGATTTCTGGG
TCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAGTACTCCAAGGGCCGCT

GCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCAGTGTATTGACCCACAG
TGTGGGGGTCGTGGGATTTGTATCATGGGCTCTTGTGCTTGCAACTCAGGATACAAAGGAGAAAGTTG
TGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATCCACGGGGAATGTCACT
GCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCCAGACCAGTGCTCCGGC
CACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGACTGGCCCAGACTGCTC
AAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGGACGTGTCGCTGTGAAG
AAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGCCGAGCACGGGACCTGC
AAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTATCGCTCACTATTTGGA
TAAGATAGTTAAAGAGGGTTGTCCTGGTCTGTGCAACAGCAATGGAAGATGTACCCTGGACCAAAATG
GCTGGCATTGTGTGTGCCAGCCTGGATGGAGAGGAGCAGGCTGTGACGTAGCCATGGAGACTCTTTGC
ACAGATAGTAAGGACAATGAAGGAGATGGACTCATTGACTGCATGGATCCCGATTGCTGCCTACAGAG
TTCCTGCCAGAATCAGCCCTATTGTCGGGGACTGCCGGATCCTCAGGACATCATTAGCCAAAGCCTTC
AATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCCTTATAGGATCTGATAGC
ACCCATGTTATACCTGGAGAAAGTCCTTTCAATAAGAGCCTTGCATCTGTCATCAGAGGCCAAGTACT
GACTGCTGATGGAACTCCACTTATTGGAGTAAATGTCTCGTTTTTCCATTACCCAGAATATGGATATA
CTATTACCCGCCAGGACGGAATGTTTGACTTGGTGGCAAATGGTGGGGCCTCTCTAACTTTGGTATTT
GAACGATCCCCATTCCTCACTCAGTATCATACTGTGTGGATTCCATGGAATGTCTTTTATGTGATGGA
TACCCTAGTCATGAAGAAAGAAGAGAATGACATTCCCAGCTGTGATCTGAGTGGATTCGTGAGGCCAA
ATCCCATCATTGTGTCATCACCTTTATCCACCTTTTTCAGATCTTCTCCTGAAGACAGTCCCATCATT
CCCGAAACACAGGTACTCCACGAGGAAACTACAATTCCAGGAACAGATTTGAAACTCTCCTACTTGAG
TTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAGATCACCATGACCCAGTCTATTATTCCATTTAATT
TAATGAAGGTTCATCTTATGGTAGCTGTAGTAGGAAGACTCTTCCAAAAGTGGTTTCCTGCCTCACCA
AACTTGGCCTATACTTTCATATGGGATAAAACAGATGCATATAATCAGAAAGTCTATGGTCTATCTGA
AGCTGTTGTGTCAGTTGGATATGAGTATGAGTCGTGTTTGGACCTGACTCTGTGGGAAAAGAGGACTG
CCATTCTGCAGGGCTATGAATTGGATGCGTCCAACATGGGTGGCTGGACATTAGATAAACATCACGTG
CTGGATGTACAGAACGGTATACTGTACAAGGGAAACGGGGAAAACCAGTTCATCTCCCAGCAGCCTCC
AGTCGTGAGTAGC**TAGGTCGACGGC**

NOV40aa, CG55069-13 Protein Sequence	SEQ ID NO: 538	838 aa	MW at 91589.1kD

MLPGLALLLLAAWTARANWQLQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGE
LDIGRRAIQEIPPGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSRLI
AREQRSLLETERAGRQARSVSLHEAGSIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCH
GNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRG
ICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYL
QESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKC
ECSQGWNGEHCTIAHYLDKIVKEGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKD
NEGDGLIDCMDPDCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIP
GESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPF
LTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQV
LHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYT
FIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQN
GILYKGNGENQFISQQPPVVSS

NOV40ab, CG55069-14 DNA Sequence	SEQ ID NO: 539	1755 bp
	ORF Start: ATG at 890	ORF Stop: TAA at 1532

GAGAAAGGAGATTAAAAATAACCTCTGGATATTCCTCTCATGTGATCTTTATTCTGGATGAAGCATTA
GGACAGCTAATAGCCGTGTGTCACTGTGTGATTTCTTCCCTAAGACTAAGGACCCATCATTTTAGTGC
AACCTTCTTCATTTAAATGGAGAGTTGTAATTGCCAATGCTCACAGCTACTCCTGCTCCGGCAATTTG
CTGCCAGAAGTGTGTTTTCCTTTTTAAAAGGCAGTAAATTCAAGATGTTGTGGTGGATGTAGATTTTT
GCTGCAAGGAAATAACAGCTGGTGATGGAATTTCATTCTTTTGACTTCTAGATTGCCTGTGAAGAGCT
GCTTCCTCGGAAGAGCACCCTAAGGCTGGGTGGCCACTATCCTTTGCCTTGGCAGAGCCAGCCAGAAG
GCCTAGGCACAACCCGCTGTGTTTGCTGACAGCCAACCTACCCTGGAGTTCCGGAGCGGCTTCCTAGG
AAGACTGGGGAGCGGTAGAAAAATGGCTCTGCTGAGATGAGCTCTTAATTAATGCACTGAGAGCCTGC
AAGTCCCACCTCTCAACAGGAATGATTGACGTCCAAGGATACATAAATTACACTAACTGAGCTCTGCC
TCTATATAAGCTTTCCACATCCAACTCATCAGAGAAGCTAGGCTTGTACCATAACCAATACCCCTGCT
TGGCAACTCTAATGAGCAAACTGCCGCAAAATTGAGAGAGAACACACCTTTTTGATTTCCTGCTCTTC
TAAGACACAGTGATTTAGAATTTCTGTTCAAGCAAGAGAACTAAAGACTTCTTTAAAGAAGAGAAGAG
AGGCCAATGAGACTTGAACCCTGAGCCTAAGTTGTCACCAGCAGGACTGATGTGCACACAGAAGGAAT

GAAGTATGGATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAGAAGGAACGG
CGCTACACAAATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCACACACAACTCCTACAGTTCCAG
CGAGACATTGAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGAAGGATTTGG
TTCACAGAGAAGCAGACGAGTTCACTAGACAAGGACAGAATTTTACCCTAAGGCAGTTAGGAGTTTGT
GAACCAGCAACTCGAAGAGGACTGGCATTTTGTGCGGAAATGGGGCTCCCTCACAGAGGTTACTCTAT
CAGTGCAGGGTCAGATGCTGATACTGAAAATGAAGCAGTGATGTCCCCAGAGCATGCCATGAGACTTT
GGGGCAGGGGGTTCAAATCAGGCCGCAGCTCCTGCCTGTCAAGTCGGTCCAACTCAGCCCTCACCCTG
ACAGATACGGAGCACGAAAACAAGTCCGACAGTGAGAATGGAGGGTCAAGCAGTTGGTTCGGTTTTCA
TTGGAATTTTTATGTGAGTAAAGCTTCCTGTTTGCTGCGCTTGCCTAGGATTTTCTTATCCCACAACT
ACAATGTGAACAAAGAGATGAGAGAGAAATTATGCTAATGCATTTTGGTGGATCAAATGAGTGTTTCA
TGAGACAACTCAAATTTTTGTTAGCTATATGGTGTTGGAATATAATTTCAAAGACAACTAAGCCCTAA
AATAGGAGATTTATTTAAAACATAACTTTTCCTTGAATGAAAGGATGTTTTTGTTCTTTCTCTGACAA
ATATGATTTGAGAATAAAAGACCTGCCCGGGCAGCCGCTCGAGCCCTATAGTGAG

NOV40ab, CG55069-14 Protein Sequence	SEQ ID NO: 540	214 aa	MW at 24449.8kD

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTHNSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQGQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWG
RGFKSGRSSCLSSRSNSALTLTDTEHENKSDSENGGSSSWFGFHWNFYVSKASCLLRLPRIFLSHNYN
VNKEMREKLC

NOV40ac, CG55069-15 DNA Sequence	SEQ ID NO: 541	768 bp
	ORF Start: ATG at 65	ORF Stop: TAA at 707

AGACTTGAACCCTGAGCCTAAGTTGTCACCAGCAGGACTGATGTGCACACAGAAGGAATGAAGTATGG
ATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAGAAGGAACGGCGCTACACA
AATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCACACAGAAGTCCTACAGTTCCAGCGAGACATT
GAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGAAGGATTTGGTTCACAGAG
AAGCAGACGAGTTCACTAGACAAGGACAGAATTTTACCCTAAGGCAGTTAGGAGTTTGTGAACCAGCA
ACTCGAAGAGGACTGGCATTTTGTGCGGAAATGGGGCTCCCTCACAGAGGTTACTCTATCAGTGCAGG
GTCAGATGCTGATACTGAAAATGAAGCAGTGATGTCCCCAGAGCATGCCATGAGACTTTGGGGCAGGG
GGGTCAAATCAGGCCGCAGCTCCTGCCTGTCAAGTCGGTCCAACTCAGCCCTCACCCTGACAGATACG
GAGCACGAAAACAAGTCCGACAGTGAGAATGGAGGGTCAAGCAGTTGGTTCGGTTTTCATTGGAATTT
TTATGTGGGTAAAGCTTCCTGTTTGCTGCGCTTGCCTAGGATTTTCTTATCCCACAACTACAATGTGA
ACAAAGAGATGAGAGAGAAATTATGCTAATGCATTTTGGTGGATCAATGCTAATGCATTTTGGTGGAT
CAATGCTAATGCATTTTGGT

NOV40ac, CG55069-15 Protein Sequence	SEQ ID NO: 542	214 aa	MW at 24376.8kD

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQGQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWG
RGVKSGRSSCLSSRSNSALTLTDTEHENKSDSENGGSSSWFGFHWNFYVGKASCLLRLPRIFLSHNYN
VNKEMREKLC

NOV40ad, SNP13374479 of CG55069-01, DNA Sequence	SEQ ID NO: 543	8657 bp
	ORF Start: ATG at 151	ORF Stop: TAA at 8326
	SNP Pos: 465	SNP Change: C to T

TTTGGCCTCGGGCCAGAATTCGGCACGAGGGGTCTGGAGCTTGGAGGAGAAGTCTGAACTAAGGATAA
ACTAAAGAGAGGCCAATGAGACTTGAACCCTGAGCCTAAGTTGTCACCAGCAGGACTGATGTGCACAC
AGAAGGAATGAAGTATGGATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAG
AAGGAACGGCGCTACACAAATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCACACAGAAGTCCTA
CAGTTCCAGCGAGACATTGAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGA
AGGATTTGGTTCACAGAGAAGCAGACGAGTTCACTAGACAAGGACAGAATTTTACCCTAAGGCAGTTA
GGAGTTTGTGAACCAGCAACTCGAAGAGGACTGGCATTTTGTGCGGAAATGGGGCTTCCTCACAGAGG
TTACTCTATCAGTGCAGGGTCAGATGCTGATACTGAAAATGAAGCAGTGATGTCCCCAGAGCATGCCA
TGAGACTTTGGGGCAGGGGGGTCAAATCAGGCCGCAGCTCCTGCCTGTCAAGTCGGTCCAACTCAGCC
CTCACCCTGACAGATACGGAGCACGAAAACAAGTCCGACAGTGAGAATGAGCAACCTGCAAGCAATCA
AGGCCAGTCTACCCTGCAGCCCTTGCCGCCTTCCCATAAGCAGCACTCTGCACAGCATCATCCATCCA
TCACTTCTCTCAACAGAAACTCCCTGACCAATAGAAGGAACCAGAGTCCGGCCCCGCCGGCTGCTTTG

CCCGCCGAGCTGCAAACCACACCCGAGTCCGTCCAGCTGCAGGACAGCTGGGTCCTTGGCAGTAATGT
ACCACTGGAAAGCAGGCATTTCCTATTCAAAACAGGAACAGGTACAACGCCACTGTTCAGTACTGCAA
CCCCAGGATACACAATGGCATCTGGCTCTGTTTATTCACCACCTACTCGGCCACTACCTAGAAACACC
CTATCAAGAAGTGCTTTTAAATTCAAGAAGTCTTCAAAGTACTGTAGCTGGAAATGCACTGCACTGTG
TGCCGTAGGGGTCTCGGTGCTCCTGGCAATACTCCTGTCTTATTTTATAGCAATGCATCTCTTTGGCC
TCAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATACCATG
CCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAAATAA
CACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATCTTCT
GGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGATGCA
TTGATTGGAGTATATGGCCGGAAGAAGTTACCGCCTTCCCATACTCAGTCCTCCCCCCAGTATGACTT
CGTGGAGCTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAG
CCGGGCGGCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATC
TGGCATCTGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGA
GTCTGTGGTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTT
TTCCAGGATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAG
TACTCCAAGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCA
GTGTATTGACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAGCTCAGGAT
ACAAAGGAGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATC
CACGGGGAATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCC
AGACCAGTGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGA
CTGGCCCAGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGG
ACGTGTCGCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGC
CGAGCACGGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTA
TCGCTCACTATTTGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGC
AACAGCAATGGAAGATGTACCCTGGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGG
AGCAGGCTGTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGGGATGGACTCA
TTGACTGCATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTG
CCGGATCCTCAGGACATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTA
TGATCGAATCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATA
AGAGCCTTGCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAAT
GTCTCGTTTTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGT
GGCAAATGGTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTG
TGTGGATTCCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGGAGAAAGAAGAGAATGACATT
CCCAGCTGTGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTT
TTTCAGATCTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAA
TTCCAGGAACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAG
ATCACCATGACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGG

AAGACTCTTCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAG
ATGCATATAATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCG
TGTTTGGACCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAA
CATGGGTGGCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAA
ACGGGGAAAACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGG
CGCAGCATTTCCTGCCCCAGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCT
AGCTTGTGGGATCGATGGCAGTCTGTACGTAGGCGATTTCAACTACGTGCGGCGGATATTCCCTTCTG
GAAATGTAACAAGTGTCTTAGAACTAAGAAATAAAGATTTTAGACATAGCAGCAACCCAGCTCATAGA
TACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTACGTTTCTGACACAAACACCCGCAGAATTTA
TCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAAAAATGCAGAAGTCGTCGCAGGGACAGGGG
AGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATGGAGGGAAGGCCGTGGAAGCCACACTCATG
AGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATCTACTTTGTTGATGGAACCATGATTAGGAA
AGTTGACCAAAATGGAATCATATCAACTCTTCTGGGCTCTAACGATTTGACTTCAGCCAGACCTTTAA
CTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGGAATGGCCCACTGACCTAGCCATTAACCCT
ATGGATAACTCCATTTATGTCCTGGATAATAATGTAGTTTTACAGATCACTGAAAATCGTCAAGTTCG
CATTGCTGCTGGACGGCCCATGCACTGTCAGGTTCCCGGAGTGGAATATCCTGTGGGGAAGCACGCGG
TGCAGACAACACTGGAATCAGCCACTGCCATTGCTGTGTCCTACAGTGGGGTCCTGTACATTACTGAA
ACTGATGAGAAGAAAATTAACCGGATAAGGCAGGTCACAACAGATGGAGAAATCTCCTTAGTGGCCGG
AATACCTTCAGAGTGTGACTGCAAAAATGATGCCAACTGTGACTGTTACCAGAGTGGAGATGGCTACG
CCAAGGATGCCAAACTCAGTGCCCCATCCTCCCTGGCTGCTTCTCCAGATGGTACACTGTATATTGCA
GATCTAGGGAATATCCGGATCCGGGCTGTGTCAAAGAATAAGCCTTTACTTAACTCTATGAACTTCTA
TGAAGTTGCGTCTCCAACTGATCAAGAACTCTACATCTTTGACATCAATGGTACTCACCAATATACTG

TAAGTTTAGTCACTGGTGATTACCTTTACAATTTTAGCTACAGCAATGACAATGATATTACTGCTGTG
ACAGACAGCAATGGCAACACCCTTAGAATTAGACGGGACCCAAATCGCATGCCAGTTCGAGTGGTGTC
TCCTGATAACCAAGTGATATGGTTGACAATAGGAACAAATGGATGTTTGAAAGGCATGACTGCTCAAG
GACTGGAATTAGTTTTGTTTACTTACCATGGCAATAGTGGCCTTTTAGCCACTAAAAGTGATGAAACT
GGATGGACAACGTTTTTTGACTATGACAGTGAAGGTCGTCTGACAAATGTTACGTTTCCAACTGGAGT
GGTCACAAACCTGCATGGGGACATGGACAAGGCTATCACAGTGGACATTGAGTCATCTAGCCGAGAAG
AAGATGTCAGCATCACTTCAAATCTGTCCTCGATCGATTCTTTCTACACCATGGTTCAAGATCAGTTA
AGAAACAGCTACCAGATTGGTTATGACGGCTCCCTCAGAATTATCTACGCCAGTGGCCTGGACTCACA
CTACCAAACAGAGCCGCACGTTCTGGCTGGCACCGCTAATCCGACGGTTGCCAAAAGAAACATGACTT
TGCCTGGCGAGAACGGTCAAAACTTGGTGGAATGGAGATTCCGAAAAGAGCAAGCCCAAGGGAAAGTC
AATGTCTTTGGCCGCAAGCTCAGGGTTAATGGCAGAAACCTCCTTTCAGTTGACTTTGATCGAACAAC
AAAGACAGAAAAGATCTATGACGACCACCGTAAATTTCTACTGAGGATCGCCTACGACACGTCTGGGC
ACCCGACTCTCTGGCTGCCAAGCAGCAAGCTGATGGCCGTCAATGTCACCTATTCATCCACAGGTCAA
ATTGCCAGCATCCAGCGAGGCACCACTAGCGAGAAAGTAGATTATGACGGACAGGGGAGGATCGTGTC
TCGGGTCTTTGCTGATGGTAAAACATGGAGTTACACATATTTAGAAAAGTCCATGGTTCTTCTGCTTC
ATAGCCAGCGGCAGTACATCTTCGAATACGATATGTGGGACCGCCTGTCTGCCATCACCATGCCCAGT
GTGGCTCGCCACACCATGCAGACCATCCGATCCATTGGCTACTACCGCAACATATACAACCCCCCGGA
AAGCAACGCCTCCATCATCACGGACTACAACGAGGAAGGGCTGCTTCTACAAACAGCTTTCTTGGGTA
CAAGTCGGAGGGTCTTATTCAAATACAGAAGGCAGACTAGGCTCTCAGAAATTTTATATGATAGCACA
AGAGTCAGTTTTACCTATGATGAAACAGCAGGAGTCCTAAAGACAGTAAACCTCCAGAGTGATGGTTT
TATTTGCACCATTAGATACAGGCAAATTGGTCCCCTGATTGACAGGCAGATTTTCCGCTTTAGTGAAG
ATGGGATGGTAAATGCAAGATTTGACTATAGCTATGACAACAGCTTTCGAGTGACCAGCATGCAGGGT
GTGATCAATGAAACGCCACTGCCTATTGATCTGTATCAGTTTGATGACATTTCTGGCAAAGTTGAGCA
GTTTGGAAAGTTTGGAGTTATATATTATGATATTAACCAGATCATTTCTACAGCTGTAATGACCTATA
CGAAGCACTTTGATGCTCATGGCCGTATCAAGGAGATTCAATATGAGATATTCAGGTCGCTCATGTAC
TGGATTACAATTCAGTATGATAACATGGGTCGGGTAACCAAGAGAGAGATTAAAATAGGGCCCTTTGC
CAACACCACCAAATATGCTTATGAATATGATGTTGATGGACAGCTCCAAACAGTTTACCTCAATGAAA

AGATAATGTGGCGGTACAACTACGATCTGAATGGAAACCTCCATTTACTGAACCCAAGTAACAGTGCG
CGTCTGACACCCCTTCGCTATGACCTGCGAGACAGAATCACTCGACTGGGTGATGTTCAATATCGGTT
GGATGAAGATGGTTTCCTACGTCAAAGGGGCACGGAAATCTTTGAATATAGCTCCAAGGGGCTTCTAA
CTCGAGTTTACAGTAAAGGCAGTGGCTGGACAGTGATCTACCGTTATGACGGCCTGGGAAGGCGTGTT
TCTAGCAAAACCAGTCTAGGACAGCACCTGCAGTTTTTTTATGCTGACTTAACTTATCCCACTAGGAT
TACTCATGTCTACAACCATTCGAGTTCAGAAATTACCTCCCTGTATTATGATCTCCAAGGACATCTTT
TTGCCATGGAAATCAGCAGTGGGGATGAATTCTATATTGCATCGGATAACACAGGGACACCACTGGCT
GTGTTCAGTAGCAATGGGCTTATGCTGAAACAGATTCAGTACACTGCATATGGGGAAATCTATTTTGA
CTCTAATATTGACTTTCAACTGGTAATTGGATTTCATGGTGGCCTGTATGACCCACTCACCAAATTAA
TCCACTTTGGAGAAAGAGATTATGACATTTTGGCAGGACGGTGGACAACACCTGACATAGAAATCTGG
AAAAGAATTGGGAAGGACCCAGCTCCTTTTAACTTGTACATGTTTAGGAATAACAACCCTGCAAGCAA
AATCCATGACGTGAAAGATTACATCACAGATGTTAACAGCTGGCTGGTGACATTTGGTTTCCATCTGC
ACAATGCTATTCCTGGATTCCCTGTTCCCAAATTTGATTTAACAGAACCTTCTTACGAACTTGTGAAG
AGTCAGCAGTGGGATGATATACCGCCCATCTTCGGAGTCCAGCAGCAAGTGGCGCGGCAGGCCAAGGC
CTTCCTGTCGCTGGGGAAGATGGCCGAGGTGCAGGTGAGCCGGCGCCGGGCCGGCGGCGCGCAGTCCT
GGCTGTGGTTCGCCACGGTCAAGTCGCTGATCGGCAAGGGCGTCATGCTGGCCGTCAGCCAGGGCCGC
GTGCAGACCAACGTGCTCAACATCGCCAACGAGGACTGCATCAAGGTGGCGGCCGTGCTCAACAACGC
CTTCTACCTGGAGAACCTGCACTTCACCATCGAGGGCAAGGACACGCACTACTTCATCAAGACCACCA
CGCCCGAGAGCGACCTGGGCACGCTGCGGTTGACCAGCGGCCGCAAGGCGCTGGAGAACGGCATCAAC
GTGACGGTGTCGCAGTCCACCACGGTGGTGAACGGCAGGACGCGCAGGTTCGCGGACGTGGAGATGCA
GTTCGGCGCGCTGGCGCTGCACGTGCGCTACGGCATGACCCTGGACGAGGAGAAGGCGCGCATCCTGG
AGCAGGCGCGGCAGCGCGCGCTCGCCCGGGCCTGGGCGCGCGAGCAGCAGCGCGTGCGCGACGGCGAG
GAGGGCGCGCGCCTCTGGACGGAGGGCGAGAAGCGGCAGCTGCTGAGCGCCGGCAAGGTGCAGGGCTA
CGACGGGTACTACGTACTCTCGGTGGAGCAGTACCCCGAGCTGGCCGACAGCGCCAACAACATCCAGT
TCCTGCGGCAGAGCGAGATCGGCAGGAGG**TAA**CGCCCGGGCCGCGCCCGCCGAGCCGCTCACGCCCTG
CCCACATTGTCCTGTGGCACAACCCGAGTGGGACTCTCCAACGCCCAAGAGCCTTCCTCCCGGGGGAA
TGAGACTGCTGTTACGACCCACACCCACACCGCGAAAACAAGGACCGCTTTTTTCCGAATGACCTTAA
AGGTGATCGGCTTTAACGAATATGTTTACATATGCATAGCGCTGCACTCAGTCGGACTGAACGTAGCC
AGAGGAAAAAAAAAATCATCAAGGACAAAGGCCTCGACCTGTTGCGCTGGGCCGTCTGTTCCTTCTAGG
CACTGTATTTAACTAACTTTA

NOV40ad, SNP13374479 of CG55069-01, Protein Sequence	SEQ ID NO: 544	2725 aa	MW at 303959.6kD
	SNP Pos: 105		SNP Change: Leu to Leu

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQGQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWG
RGVKSGRSSCLSSRSNSALTLTDTEHENKSDSENEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLN
RNSLTNRRNQSPAPPAALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYT
MASGSVYSPPTRPLPRNTLSRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQL
QQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQL
FIDQPQFLKFNISLQKDALIGVYGRKKLPPSHTQSSPQYDFVELLDGSRLIAREQRSLLETERAGRQA
RSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFL
GPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACSSGYKGES
CEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYL
DKIVKDKIGYKEGCPGLCNSNGRCTLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMD
PDCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLAS
VIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPW
NVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTD
LKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQ
KVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQ
FISQQPPVVSSIMGNGRRRSISCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTS
VLELRNKDFRHSSNPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLP
FDEARCGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTS
MHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEYPVGKHAVQTTL
ESATAIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCKNDANCDCYQSGDGYAKDAK
LSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFYEVASPTDQELYIFDINGTHQYTVSLVT
GDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPNRMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELV
LFTYHGNSGLLATKSDETGWTTFFDYDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSI
TSNLSSIDSFYTMVQDQLRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGEN
GQNLVEWRFRKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLW
LPSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVLLLHSQRQ
YIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEEGLLLQTAFLGTSRRV
LFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLIDRQIFRFSEDGMVN
ARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISGKVEQFGKFGVIYYDINQIISTAVMTYTKHFD
AHGRIKEIQYEIFRSLMYWITIQYDNMGRVTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWR
YNYDLNGNLHLLNPSNSARLTPLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYS
KGSGWTVIYRYDGLGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEI
SSGDEFYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIHFGE
RDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLVTFGFHLHNAIP
GFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMAEVQVSRRRAGGAQSWLWFA
TVKSLIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNAFYLENLHFTIEGKDTHYFIKTTTPESD
LGTLRLTSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFGALALHVRYGMTLDEEKARILEQARQ
RALARAWAREQQRVRDGEEGARLWTEGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQS
EIGRR

NOV40ae, SNP13382453 of CG55069-01, DNA Sequence	SEQ ID NO: 545	8657 bp
	ORF Start: ATG at 151	ORF Stop: TAA at 8326
	SNP Pos: 973	SNP Change: T to C

TTTGGCCTCGGGCCAGAATTCGGCACGAGGGGTCTGGAGCTTGGAGGAGAAGTCTGAACTAAGGATAA
ACTAAAGAGAGGCCAATGAGACTTGAACCCTGAGCCTAAGTTGTCACCAGCAGGACTGATGTGCACAC
AGAAGGAATGAAGT**ATG**GATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAG
AAGGAACGGCGCTACACAAATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCACACAGAAGTCCTA
CAGTTCCAGCGAGACATTGAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGA
AGGATTTGGTTCACAGAGAAGCAGACGAGTTCACTAGACAAGGACAGAATTTTACCCTAAGGCAGTTA
GGAGTTTGTGAACCAGCAACTCGAAGAGGACTGGCATTTTGTGCGGAAATGGGGCTCCCTCACAGAGG
TTACTCTATCAGTGCAGGGTCAGATGCTGATACTGAAAATGAAGCAGTGATGTCCCCAGAGCATGCCA
TGAGACTTTGGGGCAGGGGGGTCAAATCAGGCCGCAGCTCCTGCCTGTCAAGTCGGTCCAACTCAGCC

CTCACCCTGACAGATACGGAGCACGAAAACAAGTCCGACAGTGAGAATGAGCAACCTGCAAGCAATCA
AGGCCAGTCTACCCTGCAGCCCTTGCCGCCTTCCCATAAGCAGCACTCTGCACAGCATCATCCATCCA
TCACTTCTCTCAACAGAAACTCCCTGACCAATAGAAGGAACCAGAGTCCGGCCCCGCCGGCTGCTTTG
CCCGCCGAGCTGCAAACCACACCCGAGTCCGTCCAGCTGCAGGACAGCTGGGTCCTTGGCAGTAATGT
ACCACTGGAAAGCAGGCATTTCCTATTCAAAACAGGAACAGGTACAACGCCACTGTTCAGTACTGCAA
CCCCAGGATACACAATGGCACCTGGCTCTGTTTATTCACCACCTACTCGGCCACTACCTAGAAACACC
CTATCAAGAAGTGCTTTTAAATTCAAGAAGTCTTCAAAGTACTGTAGCTGGAAATGCACTGCACTGTG
TGCCGTAGGGGTCTCGGTGCTCCTGGCAATACTCCTGTCTTATTTTATAGCAATGCATCTCTTTGGCC
TCAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATACCATG
CCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAAATAA
CACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATCTTCT
GGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGATGCA
TTGATTGGAGTATATGGCCGGAAGAAGTTACCGCCTTCCCATACTCAGTCCTCCCCCCAGTATGACTT
CGTGGAGCTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAG
CCGGGCGGCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATC
TGGCATCTGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGA
GTCTGTGGTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTT
TTCCAGGATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAG
TACTCCAAGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCA
GTGTATTGACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAGCTCAGGAT
ACAAAGGAGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATC
CACGGGGAATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCC
AGACCAGTGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGA
CTGGCCCAGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGG
ACGTGTCGCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGC
CGAGCACGGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTA
TCGCTCACTATTTGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGC
AACAGCAATGGAAGATGTACCCTGGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGG
AGCAGGCTGTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGGGATGGACTCA
TTGACTGCATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTG
CCGGATCCTCAGGACATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTA
TGATCGAATCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATA
AGAGCCTTGCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAAT
GTCTCGTTTTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGT
GGCAAATGGTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTG
TGTGGATTCCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGGAGAAAGAAGAGAATGACATT
CCCAGCTGTGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTT
TTTCAGATCTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAA
TTCCAGGAACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAG
ATCACCATGACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGG

AAGACTCTTCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAG
ATGCATATAATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCG
TGTTTGGACCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAA
CATGGGTGGCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAA
ACGGGGAAAACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGG
CGCAGCATTTCCTGCCCCAGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCT
AGCTTGTGGGATCGATGGCAGTCTGTACGTAGGCGATTTCAACTACGTGCGGCGGATATTCCCTTCTG
GAAATGTAACAAGTGTCTTAGAACTAAGAAATAAAGATTTTAGACATAGCAGCAACCCAGCTCATAGA
TACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTACGTTTCTGACACAAACACCCGCAGAATTTA
TCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAAAAATGCAGAAGTCGTCGCAGGGACAGGGG
AGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATGGAGGGAAGGCCGTGGAAGCCACACTCATG
AGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATCTACTTTGTTGATGGAACCATGATTAGGAA
AGTTGACCAAAATGGAATCATATCAACTCTTCTGGGCTCTAACGATTTGACTTCAGCCAGACCTTTAA
CTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGGAATGGCCCACTGACCTAGCCATTAACCCT
ATGGATAACTCCATTTATGTCCTGGATAATAATGTAGTTTTACAGATCACTGAAAATCGTCAAGTTCG
CATTGCTGCTGGACGGCCCATGCACTGTCAGGTTCCCGGAGTGGAATATCCTGTGGGGAAGCACGCGG
TGCAGACAACACTGGAATCAGCCACTGCCATTGCTGTGTCCTACAGTGGGGTCCTGTACATTACTGAA
ACTGATGAGAAGAAAATTAACCGGATAAGGCAGGTCACAACAGATGGAGAAATCTCCTTAGTGGCCGG
AATACCTTCAGAGTGTGACTGCAAAAATGATGCCAACTGTGACTGTTACCAGAGTGGAGATGGCTACG

CCAAGGATGCCAAACTCAGTGCCCCATCCTCCCTGGCTGCTTCTCCAGATGGTACACTGTATATTGCA
GATCTAGGGAATATCCGGATCCGGGCTGTGTCAAAGAATAAGCCTTTACTTAACTCTATGAACTTCTA
TGAAGTTGCGTCTCCAACTGATCAAGAACTCTACATCTTTGACATCAATGGTACTCACCAATATACTG
TAAGTTTAGTCACTGGTGATTACCTTTACAATTTTAGCTACAGCAATGACAATGATATTACTGCTGTG
ACAGACAGCAATGGCAACACCCTTAGAATTAGACGGGACCCAAATCGCATGCCAGTTCGAGTGGTGTC
TCCTGATAACCAAGTGATATGGTTGACAATAGGAACAAATGGATGTTTGAAAGGCATGACTGCTCAAG
GACTGGAATTAGTTTTGTTTACTTACCATGGCAATAGTGGCCTTTTAGCCACTAAAAGTGATGAAACT
GGATGGACAACGTTTTTTTGACTATGACAGTGAAGGTCGTCTGACAAATGTTACGTTTCCAACTGGAGT
GGTCACAAACCTGCATGGGGACATGGACAAGGCTATCACAGTGGACATTGAGTCATCTAGCCGAGAAG
AAGATGTCAGCATCACTTCAAATCTGTCCTCGATCGATTCTTTCTACACCATGGTTCAAGATCAGTTA
AGAAACAGCTACCAGATTGGTTATGACGGCTCCCTCAGAATTATCTACGCCAGTGGCCTGGACTCACA
CTACCAAACAGAGCCGCACGTTCTGGCTGGCACCGCTAATCCGACGGTTGCCAAAAGAAACATGACTT
TGCCTGGCGAGAACGGTCAAAACTTGGTGGAATGGAGATTCCGAAAAGAGCAAGCCCAAGGGAAAGTC
AATGTCTTTGGCCGCAAGCTCAGGGTTAATGGCAGAAACCTCCTTTCAGTTGACTTTGATCGAACAAC
AAAGACAGAAAAGATCTATGACGACCACCGTAAATTTCTACTGAGGATCGCCTACGACACGTCTGGGC
ACCCGACTCTCTGGCTGCCAAGCAGCAAGCTGATGGCCGTCAATGTCACCTATTCATCCACAGGTCAA
ATTGCCAGCATCCAGCGAGGCACCACTAGCGAGAAAGTAGATTATGACGGACAGGGGAGGATCGTGTC
TCGGGTCTTTGCTGATGGTAAAACATGGAGTTACACATATTTAGAAAAGTCCATGGTTCTTCTGCTTC
ATAGCCAGCGGCAGTACATCTTCGAATACGATATGTGGGACCGCCTGTCTGCCATCACCATGCCCAGT
GTGGCTCGCCACACCATGCAGACCATCCGATCCATTGGCTACTACCGCAACATATACAACCCCCCGGA
AAGCAACGCCTCCATCATCACGGACTACAACGAGGAAGGGCTGCTTCTACAAACAGCTTTCTTGGGTA
CAAGTCGGAGGGTCTTATTCAAATACAGAAGGCAGACTAGGCTCTCAGAAATTTTATATGATAGCACA
AGAGTCAGTTTTACCTATGATGAAACAGCAGGAGTCCTAAAGACAGTAAACCTCCAGAGTGATGGTTT
TATTTGCACCATTAGATACAGGCAAATTGGTCCCCTGATTGACAGGCAGATTTTCCGCTTTAGTGAAG
ATGGGATGGTAAATGCAAGATTTGACTATAGCTATGACAACAGCTTTCGAGTGACCAGCATGCAGGGT
GTGATCAATGAAACGCCACTGCCTATTGATCTGTATCAGTTTGATGACATTTCTGGCAAAGTTGAGCA
GTTTGGAAAGTTTGGAGTTATATATTATGATATTAACCAGATCATTTCTACAGCTGTAATGACCTATA
CGAAGCACTTTGATGCTCATGGCCGTATCAAGGAGATTCAATATGAGATATTCAGGTCGCTCATGTAC
TGGATTACAATTCAGTATGATAACATGGGTCGGGTAACCAAGAGAGAGATTAAAATAGGGCCCTTTGC
CAACACCACCAAATATGCTTATGAATATGATGTTGATGGACAGCTCCAAACAGTTTACCTCAATGAAA

AGATAATGTGGCGGTACAACTACGATCTGAATGGAAACCTCCATTTACTGAACCCAAGTAACAGTGCG
CGTCTGACACCCCTTCGCTATGACCTGCGAGACAGAATCACTCGACTGGGTGATGTTCAATATCGGTT
GGATGAAGATGGTTTCCTACGTCAAAGGGGCACGGAAATCTTTGAATATAGCTCCAAGGGGCTTCTAA
CTCGAGTTTACAGTAAAGGCAGTGGCTGGACAGTGATCTACCGTTATGACGGCCTGGGAAGGCGTGTT
TCTAGCAAAACCAGTCTAGGACAGCACCTGCAGTTTTTTTATGCTGACTTAACTTATCCCACTAGGAT
TACTCATGTCTACAACCATTCGAGTTCAGAAATTACCTCCCTGTATTATGATCTCCAAGGACATCTTT
TTGCCATGGAAATCAGCAGTGGGGATGAATTCTATATTGCATCGGATAACACAGGGACACCACTGGCT
GTGTTCAGTAGCAATGGGCTTATGCTGAAACAGATTCAGTACACTGCATATGGGGAAATCTATTTTGA
CTCTAATATTGACTTTCAACTGGTAATTGGATTTCATGGTGGCCTGTATGACCCACTCACCAAATTAA
TCCACTTTGGAGAAAGAGATTATGACATTTTGGCAGGACGGTGGACAACACCTGACATAGAAATCTGG
AAAAGAATTGGGAAGGACCCAGCTCCTTTTAACTTGTACATGTTTAGGAATAACAACCCTGCAAGCAA
AATCCATGACGTGAAAGATTACATCACAGATGTTAACAGCTGGCTGGTGACATTTGGTTTCCATCTGC
ACAATGCTATTCCTGGATTCCCTGTTCCCAAATTTGATTTAACAGAACCTTCTTACGAACTTGTGAAG
AGTCAGCAGTGGGATGATATACCGCCCATCTTCGGAGTCCAGCAGCAAGTGGCGCGGCAGGCCAAGGC
CTTCCTGTCGCTGGGGAAGATGGCCGAGGTGCAGGTGAGCCGGCGCCGGGCCGGCGGCGCGCAGTCCT
GGCTGTGGTTCGCCACGGTCAAGTCGCTGATCGGCAAGGGCGTCATGCTGGCCGTCAGCCAGGGCCGC
GTGCAGACCAACGTGCTCAACATCGCCAACGAGGACTGCATCAAGGTGGCGGCCGTGCTCAACAACGC
CTTCTACCTGGAGAACCTGCACTTCACCATCGAGGGCAAGGACACGCACTACTTCATCAAGACCACCA
CGCCCGAGAGCGACCTGGGCACGCTGCGGTTGACCAGCGGCCGCAAGGCGCTGGAGAACGGCATCAAC
GTGACGGTGTCGCAGTCCACCACGGTGGTGAACGGCAGGACGCGCAGGTTCGCGGACGTGGAGATGCA
GTTCGGCGCGCTGGCGCTGCACGTGCGCTACGGCATGACCCTGGACGAGGAGAAGGCGCGCATCCTGG
AGCAGGCGCGGCAGCGCGCGCTCGCCCGGGCCTGGGCGCGCGAGCAGCAGCGCGTGCGCGACGGCGAG
GAGGGCGCGCGCCTCTGGACGGAGGGCGAGAAGCGGCAGCTGCTGAGCGCCGGCAAGGTGCAGGGCTA
CGACGGGTACTACGTACTCTCGGTGGAGCAGTACCCCGAGCTGGCCGACAGCGCCAACAACATCCAGT
TCCTGCGGCAGAGCGAGATCGGCAGGAGG**TAA**CGCCCGGGCCGCGCCCGCCGAGCCGCTCACGCCCTG
CCCACATTGTCCTGTGGCACAACCCGAGTGGGACTCTCCAACGCCCAAGAGCCTTCCTCCCGGGGGAA
TGAGACTGCTGTTACGACCCACACCCACACCGCGAAAACAAGGACCGCTTTTTTCCGAATGACCTTAA
AGGTGATCGGCTTTAACGAATATGTTTACATATGCATAGCGCTGCACTCAGTCGGACTGAACGTAGCC
AGAGGAAAAAAAAAATCATCAAGGACAAAGGCCTCGACCTGTTGCGCTGGGCCGTCTGTTCCTTCTAGG

CACTGTATTTAACTAACTTTA

NOV40ae, SNP13382453 of CG55069-01, Protein Sequence	SEQ ID NO: 546	2725 aa	MW at 303969.7kD
	SNP Pos: 275		SNP Change: Ser to Pro

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQGQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWG
RGVKSGRSSCLSSRSNSALTLTDTEHENKSDSENEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLN
RNSLTNRRNQSPAPPAALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYT
MAPGSVYSPPTRPLPRNTLSRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQL
QQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQL
FIDQPQFLKFNISLQKDALIGVYGRKKLPPSHTQSSPQYDFVELLDGSRLIAREQRSLLETERAGRQA
RSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFL
GPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACSSGYKGES
CEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYL
DKIVKDKIGYKEGCPGLCNSNGRCTLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMD
PDCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLAS
VIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPW
NVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTD
LKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQ
KVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQ
FISQQPPVVSSIMGNGRRRSISCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTS
VLELRNKDFRHSSNPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLP
FDEARCGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTS
MHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEYPVGKHAVQTTL
ESATAIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCKNDANCDCYQSGDGYAKDAK
LSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFYEVASPTDQELYIFDINGTHQYTVSLVT
GDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPNRMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELV
LFTYHGNSGLLATKSDETGWTTFFDYDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSI
TSNLSSIDSFYTMVQDQLRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGEN
GQNLVEWRFRKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLW
LPSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVLLLHSQRQ
YIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEEGLLLQTAFLGTSRRV
LFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLIDRQIFRFSEDGMVN
ARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISGKVEQFGKFGVIYYYDINQIISTAVMTYTKHFD
AHGRIKEIQYEIFRSLMYWITIQYDNMGRVTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWR
YNYDLNGNLHLLNPSNSARLTPLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYS
KGSGWTVIYRYDGLGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEI
SSGDEFYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIHFGE
RDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLVTFGFHLHNAIP
GFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMAEVQVSRRRAGGAQSWLWFA
TVKSLIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNAFYLENLHFTIEGKDTHYFIKTTTPESD
LGTLRLTSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFGALALHVRYGMTLDEEKARILEQARQ
RALARAWAREQQRVRDGEEGARLWTEGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQS
EIGRR

NOV40af, SNP13382454 of CG55069-01, DNA Sequence	SEQ ID NO: 547	8657 bp
	ORF Start: ATG at 151	ORF Stop: TAA at 8326
	SNP Pos: 1953	SNP Change: C to T

TTTGGCCTCGGGCCAGAATTCGGCACGAGGGGTCTGGAGCTTGGAGGAGAAGTCTGAACTAAGGATAA
ACTAAAGAGAGGCCAATGAGACTTGAACCCTGAGCCTAAGTTGTCACCAGCAGGACTGATGTGCACAC
AGAAGGAATGAAGTATGGATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAG
AAGGAACGGCGCTACACAAATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCACACAGAAGTCCTA
CAGTTCCAGCGAGACATTGAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGA
AGGATTTGGTTCACAGAGAAGCAGACGAGTTCACTAGACAAGGACAGAATTTTACCCTAAGGCAGTTA
GGAGTTTGTGAACCAGCAACTCGAAGAGGACTGGCATTTTGTGCGGAAATGGGGCTCCCTCACAGAGG
TTACTCTATCAGTGCAGGGTCAGATGCTGATACTGAAAATGAAGCAGTGATGTCCCCAGAGCATGCCA
TGAGACTTTGGGGCAGGGGGGTCAAATCAGGCCGCAGCTCCTGCCTGTCAAGTCGGTCCAACTCAGCC

CTCACCCTGACAGATACGGAGCACGAAAACAAGTCCGACAGTGAGAATGAGCAACCTGCAAGCAATCA
AGGCCAGTCTACCCTGCAGCCCTTGCCGCCTTCCCATAAGCAGCACTCTGCACAGCATCATCCATCCA
TCACTTCTCTCAACAGAAACTCCCTGACCAATAGAAGGAACCAGAGTCCGGCCCCGCCGGCTGCTTTG
CCCGCCGAGCTGCAAACCACACCCGAGTCCGTCCAGCTGCAGGACAGCTGGGTCCTTGGCAGTAATGT
ACCACTGGAAAGCAGGCATTTCCTATTCAAAACAGGAACAGGTACAACGCCACTGTTCAGTACTGCAA
CCCCAGGATACACAATGGCATCTGGCTCTGTTTATTCACCACCTACTCGGCCACTACCTAGAAACACC
CTATCAAGAAGTGCTTTTAAATTCAAGAAGTCTTCAAAGTACTGTAGCTGGAAATGCACTGCACTGTG
TGCCGTAGGGGTCTCGGTGCTCCTGGCAATACTCCTGTCTTATTTTATAGCAATGCATCTCTTTGGCC
TCAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATACCATG
CCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAAATAA
CACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATCTTCT
GGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGATGCA
TTGATTGGAGTATATGGCCGGAAGAAGTTACCGCCTTCCCATACTCAGTCCTCCCCCCAGTATGACTT
CGTGGAGCTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAG
CCGGGCGGCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATC
TGGCATCTGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGA
GTCTGTGGTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTT
TTCCAGGATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAG
TACTCCAAGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCA
GTGTATTGACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCTTGTGCTTGCAGCTCAGGAT
ACAAAGGAGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATC
CACGGGGAATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCC
AGACCAGTGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGA
CTGGCCCAGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGG
ACGTGTCGCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGC
CGAGCACGGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTA
TCGCTCACTATTTGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGC
AACAGCAATGGAAGATGTACCCTGGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGG
AGCAGGCTGTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGGGATGGACTCA
TTGACTGCATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTG
CCGGATCCTCAGGACATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTA
TGATCGAATCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATA
AGAGCCTTGCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAAT
GTCTCGTTTTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGT
GGCAAATGGTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTG
TGTGGATTCCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGGAGAAAGAAGAGAATGACATT
CCCAGCTGTGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTT
TTTCAGATCTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAA
TTCCAGGAACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAG
ATCACCATGACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGG

AAGACTCTTCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAG
ATGCATATAATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCG
TGTTTGGACCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAA
CATGGGTGGCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAA
ACGGGGAAAACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGG
CGCAGCATTTCCTGCCCCAGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCT
AGCTTGTGGGATCGATGGCAGTCTGTACGTAGGCGATTTCAACTACGTGCGGCGGATATTCCCTTCTG
GAAATGTAACAAGTGTCTTAGAACTAAGAAATAAAGATTTTAGACATAGCAGCAACCCAGCTCATAGA
TACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTACGTTTCTGACACAAACACCCGCAGAATTTA
TCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAAAAATGCAGAAGTCGTCGCAGGGACAGGGG
AGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATGGAGGGAAGGCCGTGGAAGCCACACTCATG
AGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATCTACTTTGTTGATGGAACCATGATTAGGAA
AGTTGACCAAAATGGAATCATATCAACTCTTCTGGGCTCTAACGATTTGACTTCAGCCAGACCTTTAA
CTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGGAATGGCCCACTGACCTAGCCATTAACCCT
ATGGATAACTCCATTTATGTCCTGGATAATAATGTAGTTTTACAGATCACTGAAAATCGTCAAGTTCG
CATTGCTGCTGGACGGCCCATGCACTGTCAGGTTCCCGGAGTGGAATATCCTGTGGGGAAGCACGCGG
TGCAGACAACACTGGAATCAGCCACTGCCATTGCTGTGTCCTACAGTGGGGTCCTGTACATTACTGAA
ACTGATGAGAAGAAAATTAACCGGATAAGGCAGGTCACAACAGATGGAGAAATCTCCTTAGTGGCCGG
AATACCTTCAGAGTGTGACTGCAAAAATGATGCCAACTGTGACTGTTACCAGAGTGGAGATGGCTACG

CCAAGGATGCCAAACTCAGTGCCCCATCCTCCCTGGCTGCTTCTCCAGATGGTACACTGTATATTGCA
GATCTAGGGAATATCCGGATCCGGGCTGTGTCAAAGAATAAGCCTTTACTTAACTCTATGAACTTCTA
TGAAGTTGCGTCTCCAACTGATCAAGAACTCTACATCTTTGACATCAATGGTACTCACCAATATACTG
TAAGTTTAGTCACTGGTGATTACCTTTACAATTTTAGCTACAGCAATGACAATGATATTACTGCTGTG
ACAGACAGCAATGGCAACACCCTTAGAATTAGACGGGACCCAAATCGCATGCCAGTTCGAGTGGTGTC
TCCTGATAACCAAGTGATATGGTTGACAATAGGAACAAATGGATGTTTGAAAGGCATGACTGCTCAAG
GACTGGAATTAGTTTTGTTTACTTACCATGGCAATAGTGGCCTTTTAGCCACTAAAAGTGATGAAACT
GGATGGACAACGTTTTTTGACTATGACAGTGAAGGTCGTCTGACAAATGTTACGTTTCCAACTGGAGT
GGTCACAAACCTGCATGGGGACATGGACAAGGCTATCACAGTGGACATTGAGTCATCTAGCCGAGAAG
AAGATGTCAGCATCACTTCAAATCTGTCCTCGATCGATTCTTTCTACACCATGGTTCAAGATCAGTTA
AGAAACAGCTACCAGATTGGTTATGACGGCTCCCTCAGAATTATCTACGCCAGTGGCCTGGACTCACA
CTACCAAACAGAGCCGCACGTTCTGGCTGGCACCGCTAATCCGACGGTTGCCAAAAGAAACATGACTT
TGCCTGGCGAGAACGGTCAAAACTTGGTGGAATGGAGATTCCGAAAAGAGCAAGCCCAAGGGAAAGTC
AATGTCTTTGGCCGCAAGCTCAGGGTTAATGGCAGAAACCTCCTTTCAGTTGACTTTGATCGAACAAC
AAAGACAGAAAAGATCTATGACGACCACCGTAAATTTCTACTGAGGATCGCCTACGACACGTCTGGGC
ACCCGACTCTCTGGCTGCCAAGCAGCAAGCTGATGGCCGTCAATGTCACCTATTCATCCACAGGTCAA
ATTGCCAGCATCCAGCGAGGCACCACTAGCGAGAAAGTAGATTATGACGGACAGGGGAGGATCGTGTC
TCGGGTCTTTGCTGATGGTAAAACATGGAGTTACACATATTTAGAAAAGTCCATGGTTCTTCTGCTTC
ATAGCCAGCGGCAGTACATCTTCGAATACGATATGTGGGACCGCCTGTCTGCCATCACCATGCCCAGT
GTGGCTCGCCACACCATGCAGACCATCCGATCCATTGGCTACTACCGCAACATATACAACCCCCCGGA
AAGCAACGCCTCCATCATCACGGACTACAACGAGGAAGGGCTGCTTCTACAAACAGCTTTCTTGGGTA
CAAGTCGGAGGGTCTTATTCAAATACAGAAGGCAGACTAGGCTCTCAGAAATTTTATATGATAGCACA
AGAGTCAGTTTTACCTATGATGAAACAGCAGGAGTCCTAAAGACAGTAAACCTCCAGAGTGATGGTTT
TATTTGCACCATTAGATACAGGCAAATTGGTCCCCTGATTGACAGGCAGATTTTCCGCTTTAGTGAAG
ATGGGATGGTAAATGCAAGATTTGACTATAGCTATGACAACAGCTTTCGAGTGACCAGCATGCAGGGT
GTGATCAATGAAACGCCACTGCCTATTGATCTGTATCAGTTTGATGACATTTCTGGCAAAGTTGAGCA
GTTTGGAAAGTTTGGAGTTATATATTATGATATTAACCAGATCATTTCTACAGCTGTAATGACCTATA
CGAAGCACTTTGATGCTCATGGCCGTATCAAGGAGATTCAATATGAGATATTCAGGTCGCTCATGTAC
TGGATTACAATTCAGTATGATAACATGGGTCGGGTAACCAAGAGAGAGATTAAAATAGGGCCCTTTGC
CAACACCACCAAATATGCTTATGAATATGATGTTGATGGACAGCTCCAAACAGTTTACCTCAATGAAA

AGATAATGTGGCGGTACAACTACGATCTGAATGGAAACCTCCATTTACTGAACCCAAGTAACAGTGCG
CGTCTGACACCCCTTCGCTATGACCTGCGAGACAGAATCACTCGACTGGGTGATGTTCAATATCGGTT
GGATGAAGATGGTTTCCTACGTCAAAGGGGCACGGAAATCTTTGAATATAGCTCCAAGGGGCTTCTAA
CTCGAGTTTACAGTAAAGGCAGTGGCTGGACAGTGATCTACCGTTATGACGGCCTGGGAAGGCGTGTT
TCTAGCAAAACCAGTCTAGGACAGCACCTGCAGTTTTTTTATGCTGACTTAACTTATCCCACTAGGAT
TACTCATGTCTACAACCATTCGAGTTCAGAAATTACCTCCCTGTATTATGATCTCCAAGGACATCTTT
TTGCCATGGAAATCAGCAGTGGGGATGAATTCTATATTGCATCGGATAACACAGGGACACCACTGGCT
GTGTTCAGTAGCAATGGGCTTATGCTGAAACAGATTCAGTACACTGCATATGGGGAAATCTATTTTGA
CTCTAATATTGACTTTCAACTGGTAATTGGATTTCATGGTGGCCTGTATGACCCACTCACCAAATTAA
TCCACTTTGGAGAAAGAGATTATGACATTTTGGCAGGACGGTGGACAACACCTGACATAGAAATCTGG
AAAAGAATTGGGAAGGACCCAGCTCCTTTTAACTTGTACATGTTTAGGAATAACAACCCTGCAAGCAA
AATCCATGACGTGAAAGATTACATCACAGATGTTAACAGCTGGCTGGTGACATTTGGTTTCCATCTGC
ACAATGCTATTCCTGGATTCCCTGTTCCCAAATTTGATTTAACAGAACCTTCTTACGAACTTGTGAAG
AGTCAGCAGTGGGATGATATACCGCCCATCTTCGGAGTCCAGCAGCAAGTGGCGCGGCAGGCCAAGGC
CTTCCTGTCGCTGGGGAAGATGGCCGAGGTGCAGGTGAGCCGGCGCCGGGCCGGCGGCGCGCAGTCCT
GGCTGTGGTTCGCCACGGTCAAGTCGCTGATCGGCAAGGGCGTCATGCTGGCCGTCAGCCAGGGCCGC
GTGCAGACCAACGTGCTCAACATCGCCAACGAGGACTGCATCAAGGTGGCGGCCGTGCTCAACAACGC
CTTCTACCTGGAGAACCTGCACTTCACCATCGAGGGCAAGGACACGCACTACTTCATCAAGACCACCA
CGCCCGAGAGCGACCTGGGCACGCTGCGGTTGACCAGCGGCCGCAAGGCGCTGGAGAACGGCATCAAC
GTGACGGTGTCGCAGTCCACCACGGTGGTGAACGGCAGGACGCGCAGGTTCGCGGACGTGGAGATGCA
GTTCGGCGCGCTGGCGCTGCACGTGCGCTACGGCATGACCCTGGACGAGGAGAAGGCGCGCATCCTGG
AGCAGGCGCGGCAGCGCGCGCTCGCCCGGGCCTGGGCGCGCGAGCAGCAGCGCGTGCGCGACGGCGAG
GAGGGCGCGCGCCTCTGGACGGAGGGCGAGAAGCGGCAGCTGCTGAGCGCCGGCAAGGTGCAGGGCTA
CGACGGGTACTACGTACTCTCGGTGGAGCAGTACCCCGAGCTGGCCGACAGCGCCAACAACATCCAGT
TCCTGCGGCAGAGCGAGATCGGCAGGAGG**TAA**CGCCCGGGCCGCGCCCGCCGAGCCGCTCACGCCCTG
CCCACATTGTCCTGTGGCACAACCCGAGTGGGACTCTCCAACGCCCAAGAGCCTTCCTCCCGGGGGAA
TGAGACTGCTGTTACGACCCACACCCACACCGCGAAAACAAGGACCGCTTTTTTCCGAATGACCTTAA
AGGTGATCGGCTTTAACGAATATGTTTACATATGCATAGCGCTGCACTCAGTCGGACTGAACGTAGCC
AGAGGAAAAAAAAAATCATCAAGGACAAAGGCCTCGACCTGTTGCGCTGGGCCGTCTGTTCCTTCTAGG

CACTGTATTTAACTAACTTTA

NOV40af, SNP13382454 of CG55069-01, Protein Sequence	SEQ ID NO: 548	2725 aa	MW at 303959.6kD
	SNP Pos: 601		SNP Change: Ser to Ser

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQGQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWG
RGVKSGRSSCLSSRSNSALTLTDTEHENKSDSENEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLN
RNSLTNRRNQSPAPPAALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYT
MASGSVYSPPTRPLPRNTLSRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQL
QQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQL
FIDQPQFLKFNISLQKDALIGVYGRKKLPPSHTQSSPQYDFVELLDGSRLIAREQRSLLETERAGRQA
RSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFL
GPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACSSGYKGES
CEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYL
DKIVKDKIGYKEGCPGLCNSNGRCTLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMD
PDCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLAS
VIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPW
NVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTD
LKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQ
KVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQ
FISQQPPVVSSIMGNGRRRSISCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTS
VLELRNKDFRHSSNPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLP
FDEARCGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTS
MHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEYPVGKHAVQTTL
ESATAIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCKNDANCDCYQSGDGYAKDAK
LSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFYEVASPTDQELYIFDINGTHQYTVSLVT
GDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPNRMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELV
LFTYHGNSGLLATKSDETGWTTFFDYDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSI
TSNLSSIDSFYTMVQDQLRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGEN
GQNLVEWRFRKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLW
LPSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVLLLHSQRQ
YIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEEGLLLQTAFLGTSRRV
LFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLIDRQIFRFSEDGMVN
ARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISGKVEQFGKFGVIYYYDINQIISTAVMTYTKHFD
AHGRIKEIQYEIFRSLMYWITIQYDNMGRVTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWR
YNYDLNGNLHLLNPSNSARLTPLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYS
KGSGWTVIYRYDGLGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEI
SSGDEFYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIHFGE
RDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLVTFGFHLHNAIP
GFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMAEVQVSRRRAGGAQSWLWFA
TVKSLIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNAFYLENLHFTIEGKDTHYFIKTTTPESD
LGTLRLTSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFGALALHVRYGMTLDEEKARILEQARQ
RALARAWAREQQRVRDGEEGARLWTEGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQS
EIGRR

NOV40ag, SNP13382455 of CG55069-01, DNA Sequence	SEQ ID NO: 549	8657 bp
	ORF Start: ATG at 151	ORF Stop: TAA at 8326
	SNP Pos: 2559	SNP Change: C to T

TTTGGCCTCGGGCCAGAATTCGGCACGAGGGGTCTGGAGCTTGGAGGAGAAGTCTGAACTAAGGATAA
ACTAAAGAGAGGCCAATGAGACTTGAACCCTGAGCCTAAGTTGTCACCAGCAGGACTGATGTGCACAC
AGAAGGAATGAAGTATGGATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAG
AAGGAACGGCGCTACACAAATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCACACAGAAGTCCTA
CAGTTCCAGCGAGACATTGAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGA
AGGATTTGGTTCACAGAGAAGCAGACGAGTTCACTAGACAAGGACAGAATTTTACCCTAAGGCAGTTA
GGAGTTTGTGAACCAGCAACTCGAAGAGGACTGGCATTTTGTGCGGAAATGGGGCTCCCTCACAGAGG
TTACTCTATCAGTGCAGGGTCAGATGCTGATACTGAAAATGAAGCAGTGATGTCCCCAGAGCATGCCA
TGAGACTTTGGGGCAGGGGGGTCAAATCAGGCCGCAGCTCCTGCCTGTCAAGTCGGTCCAACTCAGCC

CTCACCCTGACAGATACGGAGCACGAAAACAAGTCCGACAGTGAGAATGAGCAACCTGCAAGCAATCA
AGGCCAGTCTACCCTGCAGCCCTTGCCGCCTTCCCATAAGCAGCACTCTGCACAGCATCATCCATCCA
TCACTTCTCTCAACAGAAACTCCCTGACCAATAGAAGGAACCAGAGTCCGGCCCCGCCGGCTGCTTTG
CCCGCCGAGCTGCAAACCACACCCGAGTCCGTCCAGCTGCAGGACAGCTGGGTCCTTGGCAGTAATGT
ACCACTGGAAAGCAGGCATTTCCTATTCAAAACAGGAACAGGTACAACGCCACTGTTCAGTACTGCAA
CCCCAGGATACACAATGGCATCTGGCTCTGTTTATTCACCACCTACTCGGCCACTACCTAGAAACACC
CTATCAAGAAGTGCTTTTAAATTCAAGAAGTCTTCAAAGTACTGTAGCTGGAAATGCACTGCACTGTG
TGCCGTAGGGGTCTCGGTGCTCCTGGCAATACTCCTGTCTTATTTTATAGCAATGCATCTCTTTGGCC
TCAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATACCATG
CCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAAATAA
CACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATCTTCT
GGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGATGCA
TTGATTGGAGTATATGGCCGGAAGAAGTTACCGCCTTCCCATACTCAGTCCTCCCCCCAGTATGACTT
CGTGGAGCTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAG
CCGGGCGGCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATC
TGGCATCTGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGA
GTCTGTGGTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTT
TTCCAGGATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAG
TACTCCAAGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCA
GTGTATTGACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAGCTCAGGAT
ACAAAGGAGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATC
CACGGGGAATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCC
AGACCAGTGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGA
CTGGCCCAGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGG
ACGTGTCGCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGC
CGAGCACGGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTA
TCGCTCACTATTTGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGC
AACAGCAATGGAAGATGTACCCTGGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGG
AGCAGGCTGTGACGTAGCCATGGAGACTCTTTGCACAGATAGTAAGGACAATGAAGGGGATGGACTCA
TTGACTGCATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTG
CCGGATCCTCAGGACATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTA
TGATCGAATCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATA
AGAGCCTTGCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAAT
GTCTCGTTTTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGT
GGCAAATGGTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTG
TGTGGATTCCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGGAGAAAGAAGAGAATGACATT
CCCAGCTGTGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTT
TTTCAGATCTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAA
TTCCAGGAACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAG
ATCACCATGACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGG

AAGACTCTTCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAG
ATGCATATAATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCG
TGTTTGGACCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAA
CATGGGTGGCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAA
ACGGGGAAAACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGG
CGCAGCATTTCCTGCCCCAGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCT
AGCTTGTGGGATCGATGGCAGTCTGTACGTAGGCGATTTCAACTACGTGCGGCGGATATTCCCTTCTG
GAAATGTAACAAGTGTCTTAGAACTAAGAAATAAAGATTTTAGACATAGCAGCAACCCAGCTCATAGA
TACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTACGTTTCTGACACAAACACCCGCAGAATTTA
TCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAAAAATGCAGAAGTCGTCGCAGGGACAGGGG
AGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATGGAGGGAAGGCCGTGGAAGCCACACTCATG
AGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATCTACTTTGTTGATGGAACCATGATTAGGAA
AGTTGACCAAAATGGAATCATATCAACTCTTCTGGGCTCTAACGATTTGACTTCAGCCAGACCTTTAA
CTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGGAATGGCCCACTGACCTAGCCATTAACCCT
ATGGATAACTCCATTTATGTCCTGGATAATAATGTAGTTTTACAGATCACTGAAAATCGTCAAGTTCG
CATTGCTGCTGGACGGCCCATGCACTGTCAGGTTCCCGGAGTGGAATATCCTGTGGGGAAGCACGCGG
TGCAGACAACACTGGAATCAGCCACTGCCATTGCTGTGTCCTACAGTGGGGTCCTGTACATTACTGAA
ACTGATGAGAAGAAAATTAACCGGATAAGGCAGGTCACAACAGATGGAGAAATCTCCTTAGTGGCCGG
AATACCTTCAGAGTGTGACTGCAAAAATGATGCCAACTGTGACTGTTACCAGAGTGGAGATGGCTACG

CCAAGGATGCCAAACTCAGTGCCCCATCCTCCCTGGCTGCTTCTCCAGATGGTACACTGTATATTGCA
GATCTAGGGAATATCCGGATCCGGGCTGTGTCAAAGAATAAGCCTTTACTTAACTCTATGAACTTCTA
TGAAGTTGCGTCTCCAACTGATCAAGAACTCTACATCTTTGACATCAATGGTACTCACCAATATACTG
TAAGTTTAGTCACTGGTGATTACCTTTACAATTTTAGCTACAGCAATGACAATGATATTACTGCTGTG
ACAGACAGCAATGGCAACACCCTTAGAATTAGACGGGACCCAAATCGCATGCCAGTTCGAGTGGTGTC
TCCTGATAACCAAGTGATATGGTTGACAATAGGAACAAATGGATGTTTGAAAGGCATGACTGCTCAAG
GACTGGAATTAGTTTTGTTTACTTACCATGGCAATAGTGGCCTTTTAGCCACTAAAAGTGATGAAACT
GGATGGACAACGTTTTTTTGACTATGACAGTGAAGGTCGTCTGACAAATGTTACGTTTCCAACTGGAGT
GGTCACAAACCTGCATGGGGACATGGACAAGGCTATCACAGTGGACATTGAGTCATCTAGCCGAGAAG
AAGATGTCAGCATCACTTCAAATCTGTCCTCGATCGATTCTTTCTACACCATGGTTCAAGATCAGTTA
AGAAACAGCTACCAGATTGGTTATGACGGCTCCCTCAGAATTATCTACGCCAGTGGCCTGGACTCACA
CTACCAAACAGAGCCGCACGTTCTGGCTGGCACCGCTAATCCGACGGTTGCCAAAAGAAACATGACTT
TGCCTGGCGAGAACGGTCAAAACTTGGTGGAATGGAGATTCCGAAAAGAGCAAGCCCAAGGGAAAGTC
AATGTCTTTGGCCGCAAGCTCAGGGTTAATGGCAGAAACCTCCTTTCAGTTGACTTTGATCGAACAAC
AAAGACAGAAAAGATCTATGACGACCACCGTAAATTTCTACTGAGGATCGCCTACGACACGTCTGGGC
ACCCGACTCTCTGGCTGCCAAGCAGCAAGCTGATGGCCGTCAATGTCACCTATTCATCCACAGGTCAA
ATTGCCAGCATCCAGCGAGGCACCACTAGCGAGAAAGTAGATTATGACGGACAGGGGAGGATCGTGTC
TCGGGTCTTTGCTGATGGTAAAACATGGAGTTACACATATTTAGAAAAGTCCATGGTTCTTCTGCTTC
ATAGCCAGCGGCAGTACATCTTCGAATACGATATGTGGGACCGCCTGTCTGCCATCACCATGCCCAGT
GTGGCTCGCCACACCATGCAGACCATCCGATCCATTGGCTACTACCGCAACATATACAACCCCCCGGA
AAGCAACGCCTCCATCATCACGGACTACAACGAGGAAGGGCTGCTTCTACAAACAGCTTTCTTGGGTA
CAAGTCGGAGGGTCTTATTCAAATACAGAAGGCAGACTAGGCTCTCAGAAATTTTATATGATAGCACA
AGAGTCAGTTTTACCTATGATGAAACAGCAGGAGTCCTAAAGACAGTAAACCTCCAGAGTGATGGTTT
TATTTGCACCATTAGATACAGGCAAATTGGTCCCCTGATTGACAGGCAGATTTTCCGCTTTAGTGAAG
ATGGGATGGTAAATGCAAGATTTGACTATAGCTATGACAACAGCTTTCGAGTGACCAGCATGCAGGGT
GTGATCAATGAAACGCCACTGCCTATTGATCTGTATCAGTTTGATGACATTTCTGGCAAAGTTGAGCA
GTTTGGAAAGTTTGGAGTTATATATTATGATATTAACCAGATCATTTCTACAGCTGTAATGACCTATA
CGAAGCACTTTGATGCTCATGGCCGTATCAAGGAGATTCAATATGAGATATTCAGGTCGCTCATGTAC
TGGATTACAATTCAGTATGATAACATGGGTCGGGTAACCAAGAGAGAGATTAAAATAGGGCCCTTTGC
CAACACCACCAAATATGCTTATGAATATGATGTTGATGGACAGCTCCAAACAGTTTACCTCAATGAAA

AGATAATGTGGCGGTACAACTACGATCTGAATGGAAACCTCCATTTACTGAACCCAAGTAACAGTGCG
CGTCTGACACCCCTTCGCTATGACCTGCGAGACAGAATCACTCGACTGGGTGATGTTCAATATCGGTT
GGATGAAGATGGTTTCCTACGTCAAAGGGGCACGGAAATCTTTGAATATAGCTCCAAGGGGCTTCTAA
CTCGAGTTTACAGTAAAGGCAGTGGCTGGACAGTGATCTACCGTTATGACGGCCTGGGAAGGCGTGTT
TCTAGCAAAACCAGTCTAGGACAGCACCTGCAGTTTTTTTATGCTGACTTAACTTATCCCACTAGGAT
TACTCATGTCTACAACCATTCGAGTTCAGAAATTACCTCCCTGTATTATGATCTCCAAGGACATCTTT
TTGCCATGGAAATCAGCAGTGGGGATGAATTCTATATTGCATCGGATAACACAGGGACACCACTGGCT
GTGTTCAGTAGCAATGGGCTTATGCTGAAACAGATTCAGTACACTGCATATGGGGAAATCTATTTTGA
CTCTAATATTGACTTTCAACTGGTAATTGGATTTCATGGTGGCCTGTATGACCCACTCACCAAATTAA
TCCACTTTGGAGAAAGAGATTATGACATTTTGGCAGGACGGTGGACAACACCTGACATAGAAATCTGG
AAAAGAATTGGGAAGGACCCAGCTCCTTTTAACTTGTACATGTTTAGGAATAACAACCCTGCAAGCAA
AATCCATGACGTGAAAGATTACATCACAGATGTTAACAGCTGGCTGGTGACATTTGGTTTCCATCTGC
ACAATGCTATTCCTGGATTCCCTGTTCCCAAATTTGATTTAACAGAACCTTCTTACGAACTTGTGAAG
AGTCAGCAGTGGGATGATATACCGCCCATCTTCGGAGTCCAGCAGCAAGTGGCGCGGCAGGCCAAGGC
CTTCCTGTCGCTGGGGAAGATGGCCGAGGTGCAGGTGAGCCGGCGCCGGGCCGGCGGCGCGCAGTCCT
GGCTGTGGTTCGCCACGGTCAAGTCGCTGATCGGCAAGGGCGTCATGCTGGCCGTCAGCCAGGGCCGC
GTGCAGACCAACGTGCTCAACATCGCCAACGAGGACTGCATCAAGGTGGCGGCCGTGCTCAACAACGC
CTTCTACCTGGAGAACCTGCACTTCACCATCGAGGGCAAGGACACGCACTACTTCATCAAGACCACCA
CGCCCGAGAGCGACCTGGGCACGCTGCGGTTGACCAGCGGCCGCAAGGCGCTGGAGAACGGCATCAAC
GTGACGGTGTCGCAGTCCACCACGGTGGTGAACGGCAGGACGCGCAGGTTCGCGGACGTGGAGATGCA
GTTCGGCGCGCTGGCGCTGCACGTGCGCTACGGCATGACCCTGGACGAGGAGAAGGCGCGCATCCTGG
AGCAGGCGCGGCAGCGCGCGCTCGCCCGGGCCTGGGCGCGCGAGCAGCAGCGCGTGCGCGACGGCGAG
GAGGGCGCGCGCCTCTGGACGGAGGGCGAGAAGCGGCAGCTGCTGAGCGCCGGCAAGGTGCAGGGCTA
CGACGGGTACTACGTACTCTCGGTGGAGCAGTACCCCGAGCTGGCCGACAGCGCCAACAACATCCAGT
TCCTGCGGCAGAGCGAGATCGGCAGGAGGTAACGCCCGGGCCGCGCCCGCCGAGCCGCTCACGCCCTG
CCCACATTGTCCTGTGGCACAACCCGAGTGGGACTCTCCAACGCCCAAGAGCCTTCCTCCCGGGGGAA
TGAGACTGCTGTTACGACCCACACCCACACCGCGAAAACAAGGACCGCTTTTTTCCGAATGACCTTAA
AGGTGATCGGCTTTAACGAATATGTTTACATATGCATAGCGCTGCACTCAGTCGGACTGAACGTAGCC
AGAGGAAAAAAAAAATCATCAAGGACAAAGGCCTCGACCTGTTGCGCTGGGCCGTCTGTTCCTTCTAGG

CACTGTATTTAACTAACTTTA

NOV40ag, SNP13382455 of CG55069-01, Protein Sequence	SEQ ID NO: 550	2725 aa	MW at 303959.6kD
	SNP Pos: 803		SNP Change: Ser to Ser

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQGQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWG
RGVKSGRSSCLSSRSNSALTLTDTEHENKSDSENEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLN
RNSLTNRRNQSPAPPAALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYT
MASGSVYSPPTRPLPRNTLSRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQL
QQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQL
FIDQPQFLKFNISLQKDALIGVYGRKKLPPSHTQSSPQYDFVELLDGSRLIAREQRSLLETERAGRQA
RSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFL
GPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACSSGYKGES
CEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYL
DKIVKDKIGYKEGCPGLCNSNGRCTLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMD
PDCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLAS
VIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPW
NVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTD
LKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQ
KVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQ
FISQQPPVVSSIMGNGRRRSISCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTS
VLELRNKDFRHSSNPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLP
FDEARCGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTS
MHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEYPVGKHAVQTTL
ESATAIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCKNDANCDCYQSGDGYAKDAK
LSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFYEVASPTDQELYIFDINGTHQYTVSLVT
GDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPNRMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELV
LFTYHGNSGLLATKSDETGWTTFFDYDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSI
TSNLSSIDSFYTMVQDQLRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGEN
GQNLVEWRFRKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLW
LPSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVLLLHSQRQ
YIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEEGLLLQTAFLGTSRRV
LFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLIDRQIFRFSEDGMVN
ARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISGKVEQFGKFGVIYYYDINQIISTAVMTYTKHFD
AHGRIKEIQYEIFRSLMYWITIQYDNMGRVTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWR
YNYDLNGNLHLLNPSNSARLTPLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYS
KGSGWTVIYRYDGLGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEI
SSGDEFYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIHFGE
RDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLVTFGFHLHNAIP
GFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMAEVQVSRRRAGGAQSWLWFA
TVKSLIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNAFYLENLHFTIEGKDTHYFIKTTTPESD
LGTLRLTSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFGALALHVRYGMTLDEEKARILEQARQ
RALARAWAREQQRVRDGEEGARLWTEGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQS
EIGRR

NOV40ah, SNP13378354 of CG55069-01, DNA Sequence	SEQ ID NO: 551	8657 bp
	ORF Start: ATG at 151	ORF Stop: TAA at 8326
	SNP Pos: 7770	SNP Change: C to T

TTTGGCCTCGGGCCAGAATTCGGCACGAGGGGTCTGGAGCTTGGAGGAGAAGTCTGAACTAAGGATAA
ACTAAAGAGAGGCCAATGAGACTTGAACCCTGAGCCTAAGTTGTCACCAGCAGGACTGATGTGCACAC
AGAAGGAATGAAGTATGGATGTGAAAGAACGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAG
AAGGAACGGCGCTACACAAATTCCTCCGCAGACAATGAGGAGTGCCGGGTACCCACACAGAAGTCCTA
CAGTTCCAGCGAGACATTGAAAGCTTTTGATCATGATTCCTCGCGGCTGCTTTACGGCAACAGAGTGA
AGGATTTGGTTCACAGAGAAGCAGACGAGTTCACTAGACAAGGACAGAATTTTACCCTAAGGCAGTTA
GGAGTTTGTGAACCAGCAACTCGAAGAGGACTGGCATTTTGTGCGGAAATGGGGCTCCCTCACAGAGG
TTACTCTATCAGTGCAGGGTCAGATGCTGATACTGAAAATGAAGCAGTGATGTCCCCAGAGCATGCCA
TGAGACTTTGGGGCAGGGGGGTCAAATCAGGCCGCAGCTCCTGCCTGTCAAGTCGGTCCAACTCAGCC

CTCACCCTGACAGATACGGAGCACGAAAACAAGTCCGACAGTGAGAATGAGCAACCTGCAAGCAATCA
AGGCCAGTCTACCCTGCAGCCCTTGCCGCCTTCCCATAAGCAGCACTCTGCACAGCATCATCCATCCA
TCACTTCTCTCAACAGAAACTCCCTGACCAATAGAAGGAACCAGAGTCCGGCCCCGCCGGCTGCTTTG
CCCGCCGAGCTGCAAACCACACCCGAGTCCGTCCAGCTGCAGGACAGCTGGGTCCTTGGCAGTAATGT
ACCACTGGAAAGCAGGCATTTCCTATTCAAAACAGGAACAGGTACAACGCCACTGTTCAGTACTGCAA
CCCCAGGATACACAATGGCATCTGGCTCTGTTTATTCACCACCTACTCGGCCACTACCTAGAAACACC
CTATCAAGAAGTGCTTTTAAATTCAAGAAGTCTTCAAAGTACTGTAGCTGGAAATGCACTGCACTGTG
TGCCGTAGGGGTCTCGGTGCTCCTGGCAATACTCCTGTCTTATTTTATAGCAATGCATCTCTTTGGCC
TCAACTGGCAGCTACAGCAGACTGAAAATGACACATTTGAGAATGGAAAAGTGAATTCTGATACCATG
CCAACAAACACTGTGTCATTACCTTCTGGAGACAATGGAAAATTAGGTGGATTTACGCAAGAAAATAA
CACCATAGATTCCGGAGAACTTGATATTGGCCGAAGAGCAATTCAAGAGATTCCTCCCGGGATCTTCT
GGAGATCACAGCTCTTCATTGATCAGCCACAGTTTCTTAAATTCAATATCTCTCTTCAGAAGGATGCA
TTGATTGGAGTATATGGCCGGAAGAAGTTACCGCCTTCCCATACTCAGTCCTCCCCCCAGTATGACTT
CGTGGAGCTCCTGGATGGCAGCAGGCTGATTGCCAGAGAGCAGCGGAGCCTGCTTGAGACGGAGAGAG
CCGGGCGGCAGGCGAGATCCGTCAGCCTTCATGAGGCCGGCTTTATCCAGTACTTGGATTCTGGAATC
TGGCATCTGGCTTTTTATAATGATGGGAAAAATGCAGAGCAGGTGTCTTTTAATACCATTGTTATAGA
GTCTGTGGTGGAATGTCCCCGAAATTGCCATGGAAATGGAGAATGCGTTTCTGGAACTTGCCATTGTT
TTCCAGGATTTCTGGGTCCGGATTGTTCAAGAGCCGCCTGTCCAGTGTTATGTAGTGGCAACGGGCAG
TACTCCAAGGGCCGCTGCCTGTGTTTCAGCGGCTGGAAGGGCACCGAGTGTGATGTGCCGACTACCCA
GTGTATTGACCCACAGTGTGGGGGTCGTGGGATTTGTATCATGGGCTCCTGTGCTTGCAGCTCAGGAT
ACAAAGGAGAAAGTTGTGAAGAAGCTGACTGTATAGACCCTGGGTGTTCTAATCATGGTGTGTGTATC
CACGGGGAATGTCACTGCAGTCCAGGATGGGGAGGTAGCAATTGTGAAATACTGAAGACCATGTGTCC
AGACCAGTGCTCCGGCCACGGAACGTATCTTCAAGAAAGTGGCTCCTGCACGTGTGACCCTAACTGGA
CTGGCCCAGACTGCTCAAACGAAATATGTTCTGTGGACTGTGGCTCACACGGCGTTTGCATGGGGGGG
ACGTGTCGCTGTGAAGAAGGCTGGACGGGCCCAGCCTGTAATCAGAGAGCCTGCCACCCCCGCTGTGC
CGAGCACGGGACCTGCAAGGATGGCAAGTGTGAATGCAGCCAGGGCTGGAATGGAGAGCACTGCACTA
TCGCTCACTATTTGGATAAGATAGTTAAAGACAAGATAGGATATAAAGAGGGTTGTCCTGGTCTGTGC
AACAGCAATGGAAGATGTACCCTGGACCAAAATGGCGGACATTGTGTGTGCCAGCCTGGATGGAGAGG
AGCAGGCTGTGACGTAGCCATGGAGACTCTTTGCACAGATAGCAAGGACAATGAAGGGGATGGACTCA
TTGACTGCATGGATCCCGATTGCTGCCTACAGAGTTCCTGCCAGAATCAGCCCTATTGTCGGGGACTG
CCGGATCCTCAGGACATCATTAGCCAAAGCCTTCAATCGCCTTCTCAGCAAGCTGCCAAATCCTTTTA
TGATCGAATCAGTTTCCTTATAGGATCTGATAGCACCCATGTTATACCTGGAGAAAGTCCTTTCAATA
AGAGCCTTGCATCTGTCATCAGAGGCCAAGTACTGACTGCTGATGGAACTCCACTTATTGGAGTAAAT
GTCTCGTTTTTCCATTACCCAGAATATGGATATACTATTACCCGCCAGGACGGAATGTTTGACTTGGT
GGCAAATGGTGGGGCCTCTCTAACTTTGGTATTTGAACGATCCCCATTCCTCACTCAGTATCATACTG
TGTGGATTCCATGGAATGTCTTTTATGTGATGGATACCCTAGTCATGGAGAAAGAAGAGAATGACATT
CCCAGCTGTGATCTGAGTGGATTCGTGAGGCCAAATCCCATCATTGTGTCATCACCTTTATCCACCTT
TTTCAGATCTTCTCCTGAAGACAGTCCCATCATTCCCGAAACACAGGTACTCCACGAGGAAACTACAA
TTCCAGGAACAGATTTGAAACTCTCCTACTTGAGTTCCAGAGCTGCAGGGTATAAGTCAGTTCTCAAG
ATCACCATGACCCAGTCTATTATTCCATTTAATTTAATGAAGGTTCATCTTATGGTAGCTGTAGTAGG

AAGACTCTTCCAAAAGTGGTTTCCTGCCTCACCAAACTTGGCCTATACTTTCATATGGGATAAAACAG
ATGCATATAATCAGAAAGTCTATGGTCTATCTGAAGCTGTTGTGTCAGTTGGATATGAGTATGAGTCG
TGTTTGGACCTGACTCTGTGGGAAAAGAGGACTGCCATTCTGCAGGGCTATGAATTGGATGCGTCCAA
CATGGGTGGCTGGACATTAGATAAACATCACGTGCTGGATGTACAGAACGGTATACTGTACAAGGGAA
ACGGGGAAAACCAGTTCATCTCCCAGCAGCCTCCAGTCGTGAGTAGCATCATGGGCAATGGGCGAAGG
CGCAGCATTTCCTGCCCCAGTTGCAATGGTCAAGCTGATGGTAACAAGTTACTGGCCCCAGTGGCGCT
AGCTTGTGGGATCGATGGCAGTCTGTACGTAGGCGATTTCAACTACGTGCGGCGGATATTCCCTTCTG
GAAATGTAACAAGTGTCTTAGAACTAAGAAATAAAGATTTTAGACATAGCAGCAACCCAGCTCATAGA
TACTACCTTGCAACGGATCCAGTCACGGGAGATCTGTACGTTTCTGACACAAACACCCGCAGAATTTA
TCGCCCAAAGTCACTTACGGGGGCAAAAGACTTGACTAAAAATGCAGAAGTCGTCGCAGGGACAGGGG
AGCAATGCCTTCCGTTTGACGAGGCGAGATGTGGGGATGGAGGGAAGGCCGTGGAAGCCACACTCATG
AGTCCCAAAGGAATGGCAGTTGATAAGAATGGATTAATCTACTTTGTTGATGGAACCATGATTAGGAA
AGTTGACCAAAATGGAATCATATCAACTCTTCTGGGCTCTAACGATTTGACTTCAGCCAGACCTTTAA
CTTGTGACACCAGCATGCACATCAGCCAGGTACGTCTGGAATGGCCCACTGACCTAGCCATTAACCCT
ATGGATAACTCCATTTATGTCCTGGATAATAATGTAGTTTTACAGATCACTGAAAATCGTCAAGTTCG
CATTGCTGCTGGACGGCCCATGCACTGTCAGGTTCCCGGAGTGGAATATCCTGTGGGGAAGCACGCGG
TGCAGACAACACTGGAATCAGCCACTGCCATTGCTGTGTCCTACAGTGGGGTCCTGTACATTACTGAA
ACTGATGAGAAGAAAATTAACCGGATAAGGCAGGTCACAACAGATGGAGAAATCTCCTTAGTGGCCGG
AATACCTTCAGAGTGTGACTGCAAAAATGATGCCAACTGTGACTGTTACCAGAGTGGAGATGGCTACG

CCAAGGATGCCAAACTCAGTGCCCCATCCTCCCTGGCTGCTTCTCCAGATGGTACACTGTATATTGCA
GATCTAGGGAATATCCGGATCCGGGCTGTGTCAAAGAATAAGCCTTTACTTAACTCTATGAACTTCTA
TGAAGTTGCGTCTCCAACTGATCAAGAACTCTACATCTTTGACATCAATGGTACTCACCAATATACTG
TAAGTTTAGTCACTGGTGATTACCTTTACAATTTTAGCTACAGCAATGACAATGATATTACTGCTGTG
ACAGACAGCAATGGCAACACCCTTAGAATTAGACGGGACCCAAATCGCATGCCAGTTCGAGTGGTGTC
TCCTGATAACCAAGTGATATGGTTGACAATAGGAACAAATGGATGTTTGAAAGGCATGACTGCTCAAG
GACTGGAATTAGTTTTGTTTACTTACCATGGCAATAGTGGCCTTTTAGCCACTAAAAGTGATGAAACT
GGATGGACAACGTTTTTTGACTATGACAGTGAAGGTCGTCTGACAAATGTTACGTTTCCAACTGGAGT
GGTCACAAACCTGCATGGGGACATGGACAAGGCTATCACAGTGGACATTGAGTCATCTAGCCGAGAAG
AAGATGTCAGCATCACTTCAAATCTGTCCTCGATCGATTCTTTCTACACCATGGTTCAAGATCAGTTA
AGAAACAGCTACCAGATTGGTTATGACGGCTCCCTCAGAATTATCTACGCCAGTGGCCTGGACTCACA
CTACCAAACAGAGCCGCACGTTCTGGCTGGCACCGCTAATCCGACGGTTGCCAAAAGAAACATGACTT
TGCCTGGCGAGAACGGTCAAAACTTGGTGGAATGGAGATTCCGAAAAGAGCAAGCCCAAGGGAAAGTC
AATGTCTTTGGCCGCAAGCTCAGGGTTAATGGCAGAAACCTCCTTTCAGTTGACTTTGATCGAACAAC
AAAGACAGAAAAGATCTATGACGACCACCGTAAATTTCTACTGAGGATCGCCTACGACACGTCTGGGC
ACCCGACTCTCTGGCTGCCAAGCAGCAAGCTGATGGCCGTCAATGTCACCTATTCATCCACAGGTCAA
ATTGCCAGCATCCAGCGAGGCACCACTAGCGAGAAAGTAGATTATGACGGACAGGGGAGGATCGTGTC
TCGGGTCTTTGCTGATGGTAAAACATGGAGTTACACATATTTAGAAAAGTCCATGGTTCTTCTGCTTC
ATAGCCAGCGGCAGTACATCTTCGAATACGATATGTGGGACCGCCTGTCTGCCATCACCATGCCCAGT
GTGGCTCGCCACACCATGCAGACCATCCGATCCATTGGCTACTACCGCAACATATACAACCCCCCGGA
AAGCAACGCCTCCATCATCACGGACTACAACGAGGAAGGGCTGCTTCTACAAACAGCTTTCTTGGGTA
CAAGTCGGAGGGTCTTATTCAAATACAGAAGGCAGACTAGGCTCTCAGAAATTTTATATGATAGCACA
AGAGTCAGTTTTACCTATGATGAAACAGCAGGAGTCCTAAAGACAGTAAACCTCCAGAGTGATGGTTT
TATTTGCACCATTAGATACAGGCAAATTGGTCCCCTGATTGACAGGCAGATTTTCCGCTTTAGTGAAG
ATGGGATGGTAAATGCAAGATTTGACTATAGCTATGACAACAGCTTTCGAGTGACCAGCATGCAGGGT
GTGATCAATGAAACGCCACTGCCTATTGATCTGTATCAGTTTGATGACATTTCTGGCAAAGTTGAGCA
GTTTGGAAAGTTTGGAGTTATATATTATGATATTAACCAGATCATTTCTACAGCTGTAATGACCTATA
CGAAGCACTTTGATGCTCATGGCCGTATCAAGGAGATTCAATATGAGATATTCAGGTCGCTCATGTAC
TGGATTACAATTCAGTATGATAACATGGGTCGGGTAACCAAGAGAGAGATTAAAATAGGGCCCTTTGC
CAACACCACCAAATATGCTTATGAATATGATGTTGATGGACAGCTCCAAACAGTTTACCTCAATGAAA

AGATAATGTGGCGGTACAACTACGATCTGAATGGAAACCTCCATTTACTGAACCCAAGTAACAGTGCG
CGTCTGACACCCCTTCGCTATGACCTGCGAGACAGAATCACTCGACTGGGTGATGTTCAATATCGGTT
GGATGAAGATGGTTTCCTACGTCAAAGGGGCACGGAAATCTTTGAATATAGCTCCAAGGGGCTTCTAA
CTCGAGTTTACAGTAAAGGCAGTGGCTGGACAGTGATCTACCGTTATGACGGCCTGGGAAGGCGTGTT
TCTAGCAAAACCAGTCTAGGACAGCACCTGCAGTTTTTTTATGCTGACTTAACTTATCCCACTAGGAT
TACTCATGTCTACAACCATTCGAGTTCAGAAATTACCTCCCTGTATTATGATCTCCAAGGACATCTTT
TTGCCATGGAAATCAGCAGTGGGGATGAATTCTATATTGCATCGGATAACACAGGGACACCACTGGCT
GTGTTCAGTAGCAATGGGCTTATGCTGAAACAGATTCAGTACACTGCATATGGGGAAATCTATTTTGA
CTCTAATATTGACTTTCAACTGGTAATTGGATTTCATGGTGGCCTGTATGACCCACTCACCAAATTAA
TCCACTTTGGAGAAAGAGATTATGACATTTTGGCAGGACGGTGGACAACACCTGACATAGAAATCTGG
AAAAGAATTGGGAAGGACCCAGCTCCTTTTAACTTGTACATGTTTAGGAATAACAACCCTGCAAGCAA
AATCCATGACGTGAAAGATTACATCACAGATGTTAACAGCTGGCTGGTGACATTTGGTTTCCATCTGC
ACAATGCTATTCCTGGATTCCCTGTTCCCAAATTTGATTTAACAGAACCTTCTTACGAACTTGTGAAG
AGTCAGCAGTGGGATGATATACCGCCCATCTTCGGAGTCCAGCAGCAAGTGGCGCGGCAGGCCAAGGC
CTTCCTGTCGCTGGGGAAGATGGCCGAGGTGCAGGTGAGCCGGCGCCGGGCCGGCGGCGCGCAGTCCT
GGCTGTGGTTCGCCACGGTCAAGTCGCTGATCGGCAAGGGCGTCATGCTGGCCGTCAGCCAGGGCCGC
GTGCAGACCAACGTGCTTAACATCGCCAACGAGGACTGCATCAAGGTGGCGGCCGTGCTCAACAACGC
CTTCTACCTGGAGAACCTGCACTTCACCATCGAGGGCAAGGACACGCACTACTTCATCAAGACCACCA
CGCCCGAGAGCGACCTGGGCACGCTGCGGTTGACCAGCGGCCGCAAGGCGCTGGAGAACGGCATCAAC
GTGACGGTGTCGCAGTCCACCACGGTGGTGAACGGCAGGACGCGCAGGTTCGCGGACGTGGAGATGCA
GTTCGGCGCGCTGGCGCTGCACGTGCGCTACGGCATGACCCTGGACGAGGAGAAGGCGCGCATCCTGG
AGCAGGCGCGGCAGCGCGCGCTCGCCCGGGCCTGGGCGCGCGAGCAGCAGCGCGTGCGCGACGGCGAG
GAGGGCGCGCGCCTCTGGACGGAGGGCGAGAAGCGGCAGCTGCTGAGCGCCGGCAAGGTGCAGGGCTA
CGACGGGTACTACGTACTCTCGGTGGAGCAGTACCCCGAGCTGGCCGACAGCGCCAACAACATCCAGT
TCCTGCGGCAGAGCGAGATCGGCAGGAGGTAACGCCCGGGCCGCGCCCGCCGAGCCGCTCACGCCCTG
CCCACATTGTCCTGTGGCACAACCCGAGTGGGACTCTCCAACGCCCAAGAGCCTTCCTCCCGGGGGAA
TGAGACTGCTGTTACGACCCACACCCACACCGCGAAAACAAGGACCGCTTTTTTCCGAATGACCTTAA
AGGTGATCGGCTTTAACGAATATGTTTACATATGCATAGCGCTGCACTCAGTCGGACTGAACGTAGCC
AGAGGAAAAAAAAATCATCAAGGACAAAGGCCTCGACCTGTTGCGCTGGGCCGTCTGTTCCTTCTAGG

CACTGTATTTAACTAACTTTA			
NOV40ah, SNP13378354 of CG55069-01, Protein Sequence	SEQ ID NO: 552	2725 aa	MW at 303959.6kD
	SNP Pos: 2540		SNP Change: Leu to Leu

MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVH
READEFTRQGQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWG
RGVKSGRSSCLSSRSNSALTLTDTEHENKSDSENEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLN
RNSLTNRRNQSPAPPAALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYT
MASGSVYSPPTRPLPRNTLSRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQL
QQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQL
FIDQPQFLKFNISLQKDALIGVYGRKKLPPSHTQSSPQYDFVELLDGSRLIAREQRSLLETERAGRQA
RSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFL
GPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACSSGYKGES
CEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYL
DKIVKDKIGYKEGCPGLCNSNGRCTLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMD
PDCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLAS
VIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPW
NVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTD
LKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQ
KVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQ
FISQQPPVVSSIMGNGRRRSISCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTS
VLELRNKDFRHSSNPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLP
FDEARCGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTS
MHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEYPVGKHAVQTTL
ESATAIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCKNDANCDCYQSGDGYAKDAK
LSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFYEVASPTDQELYIFDINGTHQYTVSLVT
GDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPNRMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELV
LFTYHGNSGLLATKSDETGWTTFFDYDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSI
TSNLSSIDSFYTMVQDQLRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGEN
GQNLVEWRFRKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLW
LPSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVLLLHSQRQ
YIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEEGLLLQTAFLGTSRRV
LFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLIDRQIFRFSEDGMVN
ARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISGKVEQFGKFGVIYYDINQIISTAVMTYTKHFD
AHGRIKEIQYEIFRSLMYWITIQYDNMGRVTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWR
YNYDLNGNLHLLNPSNSARLTPLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYS
KGSGWTVIYRYDGLGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEI
SSGDEFYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIHFGE
RDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLVTFGFHLHNAIP
GFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMAEVQVSRRRAGGAQSWLWFA
TVKSLIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNAFYLENLHFTIEGKDTHYFIKTTTPESD
LGTLRLTSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFGALALHVRYGMTLDEEKARILEQARQ
RALARAWAREQQRVRDGEEGARLWTEGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQS
EIGRR

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 40B.

Table 40B. Comparison of the NOV40 protein sequences.

```
NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
```

```
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYG
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  NRVKDLVHREADEFTRQEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLNRNSLTNRRN
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------
```

```
NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  QSPAPPAALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYTM
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ASGSVYSPPTRPLPRNTLSRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHL
NOV40x  ------------------------------------------------------------
```

```
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  FGLNWQLQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRR
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
```

```
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  AIQEIPPGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHTQYDFVELLDGSR
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  LIAREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIE
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
```

```
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  SVVECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGT
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ECDVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCS
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
```

```
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  PGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMG
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  GTCRCEEGWTGPTCNQRACHPRCAEHGTCKDGKCECSHGWNGEHCTIEGCPGLCNSNGRC
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------
```

```
NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  TLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDPDCCLQSSCQNQPYCR
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  GLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESPFNKSLASVIRGQVLT
NOV40x  ------------------------------------------------------------
```

```
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFERSPFLTQYHTVWIPW
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
```

```
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  NVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSSPEDSPIIPETQVLHE
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMVAVVGRLFQKWFPASP
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
```

```
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  NLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRTAILQGYELDASNMGG
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  WTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSISCPSCNGQADGNKLL
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
```

```
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  APVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELRNKDFRHSSNPAHRYYLATDPVTG
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  DLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEARCGDGGKAVEATLMSP
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------
```

```
NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  KGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTCDTSMHISQVRLEWPT
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  DLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEYPVGKHAVQTTLESAT
NOV40x  ------------------------------------------------------------
```

```
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  AIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCKNDANCDCYQSGDGYA
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
```

```
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  KDAKLSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFYEVASPTDQELYIFDI
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  NGTHQYTVSLVTGDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPNRMPVRVVSPDNQVIW
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
```

```
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  LTIGTNGCLKGMTAQGLELVLFTYHGNSGLLATKSDETGWTTFFDYDSEGRLTNVTFPTG
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ---------MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFD
NOV40v  ------------------------------------------------------------
NOV40w  VVTNLHGDMDKAITVDIESSSREEDVSITSNLSSIDSFYTMVQDQLRNSYQIGYDGSLRI
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
```

```
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  HDSSRLLYGNRVKDLVHREADEFTRQGQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYS
NOV40v  ------------------------------------------------------------
NOV40w  IYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGENGQNLVEWRFRKEQAQGKVNVFGRK
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  ISAGSDADTENEAVMSPEHAMRLWGRGVKSGRSSCLSSRSNSALTLTDTEHENKSDSENE
NOV40v  ------------------------------------------------------------
NOV40w  LRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLWLPSSKLMAVNVTYSST
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------
```

```
NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  QPASNQGQSTLQPLPPSHKQHSAQHHPSITSLNRNSLTNRRNQSPAPPAALPAELQTTPE
NOV40v  ------------------------------------------------------------
NOV40w  GQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVLLLHSQRQYIFEYDMW
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ----------------------MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQK
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  SVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYTMASGSVYSPPTRPLPRNTL
NOV40v  ------------------------------------------------------------
NOV40w  DRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEEGLLLQTAFLGTSRRV
NOV40x  ------------------------------------------------------------
```

```
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  SYSSSETLKAFDHDSSRLLYGNRVKDLVHREADEFTRQGQNFTLRQLGVCEPATRRGLAF
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------MDVKER
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  SRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQLQQTENDTFENG
NOV40v  ------------------------------------------------------------
NOV40w  LFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLIDRQIFR
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  CAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWGRGVKSGRSSCLSSRSNSALTLTD
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  RPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDL
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
```

```
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  KVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIPPGIFWRSQLFID
NOV40v  ------------------------------------------------------------
NOV40w  FSEDGMVNARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISGKVEQFGKFGVIYYDI
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  TEHENKSDSENEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLNRNSLTNRRNQSPAPP
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  VHREADEFTRQEQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLNRNSLTNRRNQSPAPP
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  QPQFLKFNISLQKDALIGVYGRKKLPPSHTQSSPQYDFVELLDGSRLIAREQRSLLETER
NOV40v  ------------------------------------------------------------
NOV40w  NQIISTAVMTYTKHFDAHGRIKEIQYEIFRSLMYWITIQYDNMGRVTKREIKIGPFANTT
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  AALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYTMASGSVY
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  AALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTGTGTTPLFSTATPGYTMASGSVY
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
```

```
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  AGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESVVECPRNCHGNGE
NOV40v  ------------------------------------------------------------
NOV40w  KYAYEYDVDGQLQTVYLNEKIMWRYNYDLNGNLHLLNPSNSARLTPLRYDLRDRITRLGD
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  SPPTRPLPRNTLSRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQ
NOV40b  ------------------------------------------------------------
NOV40c  ----------------------------------------------------------GTNWQ
NOV40d  ----------------------------------------------------------GTNWQ
NOV40e  ----------------------------------------------------------GTNWQ
NOV40f  ---------------------------------------------------------TKLNWQ
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  SPPTRPLPRNTLSRSAFKFKKSSKYCSWKCTALCAVGVSVLLAILLSYFIAMHLFGLNWQ
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ----------------------------------------------------------NWQ
NOV40r  ---------------------------------------------------------TSRNWQ
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  CVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTECDVPTTQCIDPQC
NOV40v  ------------------------------------------------------------
NOV40w  VQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYSKGSGWTVIYRYDGLGRRVSSKTSLGQHL
NOV40x  ----------------------------------------------------------NWQ
NOV40y  ----------------------------------------------------------NWQ
NOV40z  ----------------------------------------------------------NWQ
NOV40aa ------------------------------------------MLPGLALLLLAAWTARANWQ
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  LQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
NOV40b  ------------------------------------------------------------
NOV40c  LQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
NOV40d  LQQTENDAFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
NOV40e  LQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
NOV40f  LQQTENDAFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
```

```
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  LQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  LQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
NOV40r  LQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  GGRGICIMGSCACSSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPGWGGSNCEILKTM
NOV40v  ------------------------------------------------------------
NOV40w  QFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEISSGDEFYIASDNTGTPLAVF
NOV40x  LQQTENDAFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
NOV40y  LQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
NOV40z  LQQTENDAFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
NOV40aa LQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGFTQENNTIDSGELDIGRRAIQEIP
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKKLPPSHTQSSPQYDFVELLDGSRLI
NOV40b  ------------------------------------------------------------
NOV40c  PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHT----QYDFVELLDGSRLI
NOV40d  PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHT----QYDFVELLDGSRLI
NOV40e  PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHT----QYDFVELLDGSRLI
NOV40f  PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHT----QYDFVELLDGSRLI
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHT----QYDFVELLDGSRLI
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHT----QYDFVELLDGSRLI
NOV40r  PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHT----QYDFVELLDGSRLI
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  CPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHG----VCMGGTCRCEEGWT
NOV40v  ------------------------------------------------------------
NOV40w  SSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIHFGERDYDILAGRWTT
NOV40x  PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHT----QYDFVELLDGSRLI
NOV40y  PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHT----QYDFVELLDGSRLI
NOV40z  PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHT----QYDFVELLDGSRLI
NOV40aa PGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGLPPSHT----QYDFVELLDGSRLI
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------
```

```
NOV40a  AREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40b  ------------------------------------------------------------
NOV40c  AREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40d  AREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40e  AREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40f  AREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  AREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  AREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40r  AREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40s  ------------------------------------------------------------
NOV40t  ------------------------------------------------------------
NOV40u  GPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVKDKIGYKEGCPGLCN
NOV40v  ------------------------------------------------------------
NOV40w  PDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLVTFGFHLHNAIPGFPV
NOV40x  AREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40y  AREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40z  AREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40aa AREQRSLLETERAGRQARSVSLHEAGSIQYLDSGIWHLAFYNDGKNAEQVSFNTIVIESV
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40b  --CPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40c  VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40d  VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40e  VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40f  VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40g  --TSRTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLG---KMAEVQVSRRRAGGA
NOV40h  ------------------------SMDVKERRPYCSLTKSRREKERRYTNSSADNEEC
NOV40i  VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40j  -----KLDQNG-----------------------------------GHCVCQPGWRGAGC
NOV40k  -----KLDQNG-----------------------------------GHCVCQPGWRGAGC
NOV40l  -----KLDQNG-----------------------------------GHCVCQPGWRGAGC
NOV40m  -----KLDQNG-----------------------------------GHCVCQPGWRGAGC
NOV40n  -----KLDQNG-----------------------------------GHCVCQPGWRGAGC
NOV40o  -----KLDQNG-----------------------------------GHCVCQPGWRGAGC
NOV40p  -----KLDQNG-----------------------------------GHCVCQPGWRGAGC
NOV40q  VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40r  VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40s  ------------------------MDVKERRPYCSLTKSRREKERRYTNSSADNEEC
NOV40t  --CPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40u  SNGRCTLDQNG-----------------------------------GHCVCQPGWRGAGC
NOV40v  -------DQNG-----------------------------------GHCVCQPGWRGAGC
NOV40w  PKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLG---KMAEVQVSRRRAGGA
NOV40x  VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
```

```
NOV40y  VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40z  VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40aa VECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGNGQYSKGRCLCFSGWKGTEC
NOV40ab ---------------------------MDVKERRPYCSLTKSRREKERRYTNSSADNEEC
NOV40ac ---------------------------MDVKERRPYCSLTKSRREKERRYTNSSADNEEC

NOV40a  DVPTTQCIDPQCGGRGICIMGSCACSSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40b  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40c  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40d  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40e  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40f  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40g  QSWLWFATVKSLIGKGVMLAVSQG-RVQTNVLNIANEDCIKVAAVLNNAFYLENLHFTIE
NOV40h  RVPTQKSYSSSE----------------TLKAFDHDSSRLLYGNRVKDLVHREADEFTRQ
NOV40i  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGKSCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40j  DVAMETLCTDSK-----------------DNEGDGLIDCMDPDCCLQSSCQN-QPYCRGL
NOV40k  DVAMETLCTDSK-----------------DNEGDGLIDCMDPDCCLQSSCQN-QPYCRGL
NOV40l  DVAMETLCTDSK-----------------DNEGDGLIDCMDPDCCLQSSCQN-QPYCRGL
NOV40m  DVAMETLCTDSK-----------------DNEGDGLIDCMDPDCCLQSSCQN-QPYCRGL
NOV40n  DVAMETLCTDSK-----------------DNEGDGLIDCMDPDCCLQSSCQN-QPYCRGL
NOV40o  DVAMETLCTDSK-----------------DNEGDGLIDCMDPDCCLQSSCQN-QPYCRGL
NOV40p  DVAMETLCTDSK-----------------DNEGDGLIDCMDPDCCLQSSCQN-QSYCRGL
NOV40q  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40r  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40s  RVPTQKSYSSSE----------------TLKAFDHDSSRLLYGNRVKDLVHREADEFTRQ
NOV40t  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40u  DVAMETLCTDSK-----------------DNEGDGLIDCMDPDCCLQSSCQN-QPYCRGL
NOV40v  DVAMETLCTDSK-----------------DNEGDGLIDCMDPDCCLQSSCQN-QPYCRGL
NOV40w  QSWLWFATVKSLIGKGVMLAVSQG-RVQTNVLNIANEDCIKVAAVLNNAFYLENLHFTIE
NOV40x  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40y  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40z  DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40aa DVPTTQCIDPQCGGRGICIMGSCACNSGYKGESCEEADCIDPGCSNHGVCIHGECHCSPG
NOV40ab RVPTHNSYSSSE----------------TLKAFDHDSSRLLYGNRVKDLVHREADEFTRQ
NOV40ac RVPTQKSYSSSE----------------TLKAFDHDSSRLLYGNRVKDLVHREADEFTRQ

NOV40a  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40b  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40c  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40d  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40e  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40f  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40g  -GKDTHYFIKTTTPESDLGTLRLTSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFG
NOV40h  EQPASNQGQSTLQPLPPSHKQHSAQHHPSITSLNRNSLTNRRNQSPAPPAALPAELQTTP
NOV40i  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40j  PDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGES-----PFNKSLASVIRGQ
NOV40k  PDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGES-----PFNKSLASVIRGQ
NOV40l  PDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGES-----PFNKSLASVIRGQ
NOV40m  PDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGES-----PFNKSLASVIRGQ
NOV40n  PDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGES-----PFNKSLASVIRGQ
NOV40o  PDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGES-----PFNKSLASVIRGQ
NOV40p  PDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGES-----PFNKSLASVIRGQ
NOV40q  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40r  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
```

```
NOV40s  GQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWGR
NOV40t  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40u  PDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGES-----PFNKSLASVIRGQ
NOV40v  PDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGES-----PFNKSLASVIRGQ
NOV40w  -GKDTHYFIKTTTPESDLGTLRLTSGRKALENGINVTVSQSTTVVNGRTRRFADVEMQFG
NOV40x  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40y  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40z  WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40aa WGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDCSNEICSVDCGSHGVCMGGT
NOV40ab GQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWGR
NOV40ac GQNFTLRQLGVCEPATRRGLAFCAEMGLPHRGYSISAGSDADTENEAVMSPEHAMRLWGR

NOV40a  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVKDKIGYK
NOV40b  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTI---------------
NOV40c  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTI---------------
NOV40d  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVKDKIGYK
NOV40e  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVK------
NOV40f  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVKDKIGYK
NOV40g  ALALHVRYGMTLDEEKARILEQA------------QQRVRDG------------------
NOV40h  ESVQLQDSWVLG-SNVPLESRHFLFKTGTGTTP--------------------------
NOV40i  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVKDKIGYK
NOV40j  VLTADGTPLIGVNVSFFHYPEYGYTITRQ------------------------------
NOV40k  VLTADGTPLIGVNVSFFHYPEYGYTITRQ------------------------------
NOV40l  VLTADGTPLIGVNVSFFHYPEYGYTITRQ------------------------------
NOV40m  VLTADGTPLIGVNVSFFHYPEYGYTITRQ------------------------------
NOV40n  VLTADGTPLIGVNVSFFHYPEYGYTITRQ------------------------------
NOV40o  VLTADGTPLIGVNVSFFHYPEYGYTITRQ------------------------------
NOV40p  VLTADGTPLIGVNVSFFHYPEYGYTITRQ------------------------------
NOV40q  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVK------
NOV40r  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVK------
NOV40s  GVKSGRSSCLSSRSNSALTLTDTEHENKSDSENEQPASNQGQSTLQPLPPSHKQHSAQHH
NOV40t  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVK------
NOV40u  VLTADGTPLIGVNVSFFHYPEYGYTITRQ------------------------------
NOV40v  VLTADGTPLIGVNVSFFHYPEYGYTITRQ------------------------------
NOV40w  ALALHVRYGMTLDEEKARILEQARQRALARAWAREQQRVRDGEEGARLWTEGEKRQL---
NOV40x  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVKDKIGYK
NOV40y  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVK------
NOV40z  CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVKDKIGYK
NOV40aa CRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGEHCTIAHYLDKIVK------
NOV40ab GFKSGRSSCLSSRSNSALTLTDTEHENKSDSEN--------------------------
NOV40ac GVKSGRSSCLSSRSNSALTLTDTEHENKSDSEN--------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
```

```
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  PSITSLNRNSLTNRRNQSPAPPAALPAELQTTPESVQLQDSWVLGSNVPLESRHFLFKTG
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  --EGCPGLCNSNGRCTLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40b  --EGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCD------------------------
NOV40c  --EGCPGLSNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40d  --EGCPGLCNSNGRCTLDQNGWHCACQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40e  --EGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40f  --EGCPGLCNSNGRCTLDQNGWHCACQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40g  ------KVQGYDGYYVLSVEQYPELADSANNIQFLRQSEIGRRVDG--------------
NOV40h  -----LFSTATPGYTMASGSVYSPPTRPLPRNTLSRSAFKFKKS-SKYCSWKCTALCAVG
NOV40i  --EGCPGLCNSNGRCTLDQNGGHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40j  -----------DGMFDLVANGGASLTLVFERSP-------------------------
NOV40k  -----------DGMFDLVASGGASLTLVFERSP-------------------------
NOV40l  -----------DGMFDLVANGGASLTLVFERSP-------------------------
NOV40m  -----------DGMFDLVANGGASLTLVFERSP-------------------------
NOV40n  -----------DGMFDLVANGGASLTLVFERSP-------------------------
NOV40o  -----------DGMFDLVANGGASLTLVFERSP-------------------------
NOV40p  -----------DGMFDLVANGGASLTLVFERSP-------------------------
NOV40q  --EGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40r  --EGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40s  TGTTPLFSTATPGYTMASGSVYSPPTRPLPRNTLSRSAFKFKKS-SKYCSWKCTALCAVG
NOV40t  --EGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCD------------------------
NOV40u  -----------DGMFDLVANGGASLTLVFERSP-------------------------
NOV40v  -----------DGMFDLVANGGASLTLVFERSP-------------------------
NOV40w  --LSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQSEIGRR-----------------
NOV40x  --EGCPGLCNSNGRCTLDQNGWHCACQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40y  --EGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40z  --EGCPGLCNSNGRCTLDQNGWHCACQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40aa --EGCPGLCNSNGRCTLDQNGWHCVCQPGWRGAGCDVAMETLCTDSKDNEGDGLIDCMDP
NOV40ab -----GGSSSWFGFHWNFYVSKASCLLRLPRIFLSHNYNVNKEMREKLC-----------
NOV40ac -----GGSSSWFGFHWNFYVGKASCLLRLPRIFLSHNYNVNKEMREKLC-----------

NOV40a  DCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
NOV40b  ------------------------------------------------------------
NOV40c  DCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
NOV40d  DCCLQSSCQNQPYCRGLPDPQGIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
NOV40e  DCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
NOV40f  DCCLQSSCQNQPYCRGLPDPQGIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
```

```
NOV40g  ------------------------------------------------------------
NOV40h  VSVLLAILLSYFIAMHLFGLNWQLQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGF
NOV40i  DCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  DCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
NOV40r  DCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
NOV40s  VSVLLAILLSYFIAMHLFGLNWQLQQTENDTFENGKVNSDTMPTNTVSLPSGDNGKLGGF
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  DCCLQSSCQNQPYCRGLPDPQGIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
NOV40y  DCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
NOV40z  DCCLQSSCQNQPYCRGLPDPQGIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
NOV40aa DCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRISFLIGSDSTHVIPGESP
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40b  ------------------------------------------------------------
NOV40c  FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40d  FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40e  FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40f  FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40g  ------------------------------------------------------------
NOV40h  TQENNTIDSGELDIGRRAIQEIPPGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGL
NOV40i  FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40r  FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40s  TQENNTIDSGELDIGRRAIQEIPPGIFWRSQLFIDQPQFLKFNISLQKDALIGVYGRKGL
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40y  FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40z  FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40aa FNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDGMFDLVANGGASLTLVFE
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------
```

```
NOV40a  RSPFLTQYHTVWIPWNVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40b  ------------------------------------------------------------
NOV40c  RSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40d  RSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40e  RSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40f  RSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40g  ------------------------------------------------------------
NOV40h  PPSHTQYDFVELLDGSRLIALEG-------------------------------------
NOV40i  RSPFLTQYHTVWIPWNVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40j  ---FLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40k  ---FLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40l  ---FLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40m  ---FLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40n  ---FLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40o  ---FLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40p  ---FLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40q  RSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40r  RSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40s  PPSHTQYDFVELLDGSRLIAREQRSLLETERAGRQARSVSLHEAGFIQYLDSGIWHLAFY
NOV40t  ------------------------------------------------------------
NOV40u  ---FLTQYHTVWIPWNVFYVMDTLVMEKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40v  ---FLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40w  ------------------------------------------------------------
NOV40x  RSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40y  RSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40z  RSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40aa RSPFLTQYHTVWIPWNVFYVMDTLVMKKEENDIPSCDLSGFVRPNPIIVSSPLSTFFRSS
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40b  ------------------------------------------------------------
NOV40c  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40d  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40e  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40f  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40j  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40k  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40l  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40m  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40n  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40o  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40p  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40q  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40r  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40s  NDGKNAEQVSFNTIVIESVVECPRNCHGNGECVSGTCHCFPGFLGPDCSRAACPVLCSGN
NOV40t  ------------------------------------------------------------
NOV40u  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40v  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40w  ------------------------------------------------------------
NOV40x  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
```

```
NOV40y  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40z  PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40aa PEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKITMTQSIIPFNLMKVHLMV
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40b  ------------------------------------------------------------
NOV40c  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40d  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40e  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40f  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40j  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40k  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40l  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40m  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40n  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSAGYEYESCLDLTLWEKRT
NOV40o  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40p  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40q  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40r  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40s  GQYSKGRCLCFSGWKGTECDVPTTQCIDPQCGGRGICIMGSCACSSGYKGESCEEADCID
NOV40t  ------------------------------------------------------------
NOV40u  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40v  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40w  ------------------------------------------------------------
NOV40x  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40y  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40z  AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40aa AVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSVGYEYESCLDLTLWEKRT
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSI
NOV40b  ------------------------------------------------------------
NOV40c  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSVD--------
NOV40d  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSVDPAASSRE-
NOV40e  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSVDATHDWRLL
NOV40f  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSLEG-------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSI
NOV40j  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSI
NOV40k  AILQGYELDASNMGGWTLDKHHALDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSI
NOV40l  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSI
NOV40m  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSI
NOV40n  AILQGYELDASNMGGWTLDKHHVLDVRNGILYKGNGENQFISQQPPVVSSIMGNGRRRSI
NOV40o  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSI
NOV40p  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSI
NOV40q  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSS----------
NOV40r  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSLEG-------
```

```
NOV40s  PGCSNHGVCIHGECHCSPGWGGSNCEILKTMCPDQCSGHGTYLQESGSCTCDPNWTGPDC
NOV40t  ------------------------------------------------------------
NOV40u  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSI
NOV40v  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSSIMGNGRRRSI
NOV40w  ------------------------------------------------------------
NOV40x  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSS----------
NOV40y  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSS----------
NOV40z  AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSS----------
NOV40aa AILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQQPPVVSS----------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  SCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELRNKDFRHSS
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  VIFSSAVAASWY--------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  SCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELRNKDFRHSS
NOV40j  SCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELRNKDFRHSS
NOV40k  SCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELRNKDFRHSS
NOV40l  SCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELS-------S
NOV40m  SCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELS-------S
NOV40n  SCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELS-------S
NOV40o  SCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELS-------S
NOV40p  SCPSCNGQADGNKLLAPVALACGIDGSLHVGDFNYVRRIFPSGNVTSVLELS-------S
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  SNEICSVDCGSHGVCMGGTCRCEEGWTGPACNQRACHPRCAEHGTCKDGKCECSQGWNGE
NOV40t  ------------------------------------------------------------
NOV40u  SCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELRNKDFRHSS
NOV40v  SCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSGNVTSVLELS-------S
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  NPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEAR
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  NPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEAR
NOV40j  NPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEAR
NOV40k  NPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDSTKNAEVVAGTGEQCLPFDEAR
NOV40l  NPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEAR
```

```
NOV40m  NPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEAR
NOV40n  NPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEAR
NOV40o  NPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEAR
NOV40p  NPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEAR
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  HCTIAHYLDKIVKDKIGYKEGCPGLCNSNGRCTLDQNGGHCVCQPGWRGAGCDVAMETLC
NOV40t  ------------------------------------------------------------
NOV40u  NPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEAR
NOV40v  NPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEAR
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  CGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTC
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  CGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTC
NOV40j  CGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTC
NOV40k  CGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTC
NOV40l  CGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTC
NOV40m  CGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTC
NOV40n  CGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTC
NOV40o  CGDGGKAAEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTC
NOV40p  CGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTC
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  TDSKDNEGDGLIDCMDPDCCLQSSCQNQPYCRGLPDPQDIISQSLQSPSQQAAKSFYDRI
NOV40t  ------------------------------------------------------------
NOV40u  CGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTC
NOV40v  CGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGIISTLLGSNDLTSARPLTC
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  DTSMHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEY
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
```

```
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  DTSMHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEY
NOV40j  DTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD--------------------------
NOV40k  DTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD--------------------------
NOV40l  DTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD--------------------------
NOV40m  DTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD--------------------------
NOV40n  DTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD--------------------------
NOV40o  DTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD--------------------------
NOV40p  DTSMHISQVRLEWPTDLAINPMDNSIYVLDNVD--------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  SFLIGSDSTHVIPGESPFNKSLASVIRGQVLTADGTPLIGVNVSFFHYPEYGYTITRQDG
NOV40t  ------------------------------------------------------------
NOV40u  DTSMHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQVRIAAGRPMHCQVPGVEY
NOV40v  DTSMHISQVRLEWPTDLAINPMDNSIYVLDN----------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  PVGKHAVQTTLESATAIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCK
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  PVGKHAVQTTLESATAIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCK
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  MFDLVANGGASLTLVFERSPFLTQYHTVWIPWNVFYVMDTLVMEKEENDIPSCDLSGFVR
NOV40t  ------------------------------------------------------------
NOV40u  PVGKHAVQTTLESATAIAVSYSGVLYITETDEKKINRIRQVTTDGEISLVAGIPSECDCK
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------
```

```
NOV40a  NDANCDCYQSGDGYAKDAKLSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFY
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  NDANCDCYQSGDGYAKDAKLSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFY
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  PNPIIVSSPLSTFFRSSPEDSPIIPETQVLHEETTIPGTDLKLSYLSSRAAGYKSVLKIT
NOV40t  ------------------------------------------------------------
NOV40u  NDANCDCYQSGDGYAKDAKLSAPSSLAASPDGTLYIADLGNIRIRAVSKNKPLLNSMNFY
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  EVASPTDQELYIFDINGTHQYTVSLVTGDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPN
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  EVASPTDQELYIFDINGTHQYTVSLVTGDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPN
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  MTQSIIPFNLMKVHLMVAVVGRLFQKWFPASPNLAYTFIWDKTDAYNQKVYGLSEAVVSV
NOV40t  ------------------------------------------------------------
NOV40u  EVASPTDQELYIFDINGTHQYTVSLVTGDYLYNFSYSNDNDITAVTDSNGNTLRIRRDPN
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
```

```
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  RMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELVLFTYHGNSGLLATKSDETGWTTFFD
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  RMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELVLFTYHGNSGLLATKSDETGWTTFFD
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  GYEYESCLDLTLWEKRTAILQGYELDASNMGGWTLDKHHVLDVQNGILYKGNGENQFISQ
NOV40t  ------------------------------------------------------------
NOV40u  RMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELVLFTYHGNSGLLATKSDETGWTTFFD
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  YDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSITSNLSSIDSFYTMVQDQ
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  YDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSITSNLSSIDSFYTMVQDQ
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
```

```
NOV40s  QPPVVSSIMGNGRRRSISCPSCNGQADGNKLLAPVALACGIDGSLYVGDFNYVRRIFPSG
NOV40t  ------------------------------------------------------------
NOV40u  YDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSITSNLSSIDSFYTMVQDQ
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  LRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGENGQNLVEWRF
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  LRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGENGQNLVEWRF
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  NVTSVLELRNKDFRHSSNPAHRYYLATDPVTGDLYVSDTNTRRIYRPKSLTGAKDLTKNA
NOV40t  ------------------------------------------------------------
NOV40u  LRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKRNMTLPGENGQNLVEWRF
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  RKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLWL
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  RKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLWL
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
```

```
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  EVVAGTGEQCLPFDEARCGDGGKAVEATLMSPKGMAVDKNGLIYFVDGTMIRKVDQNGII
NOV40t  ------------------------------------------------------------
NOV40u  RKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRKFLLRIAYDTSGHPTLWL
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  PSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVL
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  PSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVL
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  STLLGSNDLTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIYVLDNNVVLQITENRQV
NOV40t  ------------------------------------------------------------
NOV40u  PSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVFADGKTWSYTYLEKSMVL
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  LLHSQRQYIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEE
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
```

```
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  LLHSQRQYIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEE
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  RIAAGRPMHCQVPGVEYPVGKHAVQTTLESATAIAVSYSGVLYITETDEKKINRIRQVTT
NOV40t  ------------------------------------------------------------
NOV40u  LLHSQRQYIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNIYNPPESNASIITDYNEE
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  GLLLQTAFLGTSRRVLFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTI
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  GLLLQTAFLGTSRRVLFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTI
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  DGEISLVAGIPSECDCKNDANCDCYQSGDGYAKDAKLSAPSSLAASPDGTLYIADLGNIR
NOV40t  ------------------------------------------------------------
NOV40u  GLLLQTAFLGTSRRVLFKYRRQTRLSEILYDSTRVSFTYDETAGVLKTVNLQSDGFICTI
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------
```

```
NOV40a  RYRQIGPLIDRQIFRFSEDGMVNARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISG
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  RYRQIGPLIDRQIFRFSEDGMVNARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISG
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  IRAVSKNKPLLNSMNFYEVASPTDQELYIFDINGTHQYTVSLVTGDYLYNFSYSNDNDIT
NOV40t  ------------------------------------------------------------
NOV40u  RYRQIGPLIDRQIFRFSEDGMVNARFDYSYDNSFRVTSMQGVINETPLPIDLYQFDDISG
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  KVEQFGKFGVIYYDINQIISTAVMTYTKHFDAHGRIKEIQYEIFRSLMYWITIQYDNMGR
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  KVEQFGKFGVIYYDINQIISTAVMTYTKHFDAHGRIKEIQYEIFRSLMYWITIQYDNMGR
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  AVTDSNGNTLRIRRDPNRMPVRVVSPDNQVIWLTIGTNGCLKGMTAQGLELVLFTYHGNS
NOV40t  ------------------------------------------------------------
NOV40u  KVEQFGKFGVIYYDINQIISTAVMTYTKHFDAHGRIKEIQYEIFRSLMYWITIQYDNMGR
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
```

```
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  VTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWRYNYDLNGNLHLLNPSNSARLT
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  VTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWRYNYDLNGNLHLLNPSNSARLT
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  GLLATKSDETGWTTFFDYDSEGRLTNVTFPTGVVTNLHGDMDKAITVDIESSSREEDVSI
NOV40t  ------------------------------------------------------------
NOV40u  VTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWRYNYDLNGNLHLLNPSNSARLT
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  PLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYSKGSGWTVIYRYDG
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  PLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYSKGSGWTVIYRYDG
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
```

```
NOV40s  TSNLSSIDSFYTMVQDQLRNSYQIGYDGSLRIIYASGLDSHYQTEPHVLAGTANPTVAKR
NOV40t  ------------------------------------------------------------
NOV40u  PLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLTRVYSKGSGWTVIYRYDG
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  LGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEISSGDE
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  LGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEISSGDE
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  NMTLPGENGQNLVEWRFRKEQAQGKVNVFGRKLRVNGRNLLSVDFDRTTKTEKIYDDHRK
NOV40t  ------------------------------------------------------------
NOV40u  LGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLYYDLQGHLFAMEISSGDE
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  FYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIH
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  FYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIH
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
```

```
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  FLLRIAYDTSGHPTLWLPSSKLMAVNVTYSSTGQIASIQRGTTSEKVDYDGQGRIVSRVF
NOV40t  ------------------------------------------------------------
NOV40u  FYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQLVIGFHGGLYDPLTKLIH
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  FGERDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLV
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  FGERDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLV
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  ADGKTWSYTYLEKSMVLLLHSQRQYIFEYDMWDRLSAITMPSVARHTMQTIRSIGYYRNI
NOV40t  ------------------------------------------------------------
NOV40u  FGERDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPASKIHDVKDYITDVNSWLV
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  TFGFHLHNAIPGFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMA
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
```

```
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  TFGFHLHNAIPGFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMA
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  YNPPESNASIITDYNEEGLLLQTAFLGTSRRVLFKYRRQTRLSEILYDSTRVSFTYDETA
NOV40t  ------------------------------------------------------------
NOV40u  TFGFHLHNAIPGFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQQQVARQAKAFLSLGKMA
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  EVQVSRRRAGGAQSWLWFATVKSLIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNA
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  EVQVSRRRAGGAQSWLWFATVKSLIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNA
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  GVLKTVNLQSDGFICTIRYRQIGPLIDRQIFRFSEDGMVNARFDYSYDNSFRVTSMQGVI
NOV40t  ------------------------------------------------------------
NOV40u  EVQVSRRRAGGAQSWLWFATVKSLIGKGVMLAVSQGRVQTNVLNIANEDCIKVAAVLNNA
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------
```

```
NOV40a  FYLENLHFTIEGKDTHYFIKTTTPESDLGTLRLTSGRKALENGINVTVSQSTTVVNGRTR
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  FYLENLHFTIEGKDTHYFIKTTTPESDLGTLRLTSGRKALENGINVTVSQSTTVVNGRTR
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  NETPLPIDLYQFDDISGKVEQFGKFGVIYYDINQIISTAVMTYTKHFDAHGRIKEIQYEI
NOV40t  ------------------------------------------------------------
NOV40u  FYLENLHFTIEGKDTHYFIKTTTPESDLGTLRLTSGRKALENGINVTVSQSTTVVNGRTR
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  RFADVEMQFGALALHVRYGMTLDEEKARILEQARQRALARAWAREQQRVRDGEEGARLWT
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  RFADVEMQFGALALHVRYGMTLDEEKARILEQARQRALARAWAREQQRVRDGEEGARLWT
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  FRSLMYWITIQYDNMGRVTKREIKIGPFANTTKYAYEYDVDGQLQTVYLNEKIMWRYNYD
NOV40t  ------------------------------------------------------------
NOV40u  RFADVEMQFGALALHVRYGMTLDEEKARILEQARQRALARAWAREQQRVRDGEEGARLWT
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
```

```
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  EGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQSEIGRR------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  EGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQSEIGRR------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  LNGNLHLLNPSNSARLTPLRYDLRDRITRLGDVQYRLDEDGFLRQRGTEIFEYSSKGLLT
NOV40t  ------------------------------------------------------------
NOV40u  EGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQSEIGRR------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
```

```
NOV40s  RVYSKGSGWTVIYRYDGLGRRVSSKTSLGQHLQFFYADLTYPTRITHVYNHSSSEITSLY
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  YDLQGHLFAMEISSGDEFYIASDNTGTPLAVFSSNGLMLKQIQYTAYGEIYFDSNIDFQL
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
```

```
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  VIGFHGGLYDPLTKLIHFGERDYDILAGRWTTPDIEIWKRIGKDPAPFNLYMFRNNNPAS
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  KIHDVKDYITDVNSWLVTFGFHLHNAIPGFPVPKFDLTEPSYELVKSQQWDDIPPIFGVQ
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
```

```
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  QQVARQAKAFLSLGKMAEVQVSRRRAGGAQSWLWFATVKSLIGKGVMLAVSQGRVQTNVL
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  NIANEDCIKVAAVLNNAFYLENLHFTIEGKDTHYFIKTTTPESDLGTLRLTSGRKALENG
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------
```

```
NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  INVTVSQSTTVVNGRTRRFADVEMQFGALALHVRYGMTLDEEKARILEQARQRALARAWA
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  ------------------------------------------------------------
NOV40b  ------------------------------------------------------------
NOV40c  ------------------------------------------------------------
NOV40d  ------------------------------------------------------------
NOV40e  ------------------------------------------------------------
NOV40f  ------------------------------------------------------------
NOV40g  ------------------------------------------------------------
NOV40h  ------------------------------------------------------------
NOV40i  ------------------------------------------------------------
NOV40j  ------------------------------------------------------------
NOV40k  ------------------------------------------------------------
NOV40l  ------------------------------------------------------------
NOV40m  ------------------------------------------------------------
NOV40n  ------------------------------------------------------------
NOV40o  ------------------------------------------------------------
NOV40p  ------------------------------------------------------------
NOV40q  ------------------------------------------------------------
NOV40r  ------------------------------------------------------------
NOV40s  REQQRVRDGEEGARLWTEGEKRQLLSAGKVQGYDGYYVLSVEQYPELADSANNIQFLRQS
NOV40t  ------------------------------------------------------------
NOV40u  ------------------------------------------------------------
NOV40v  ------------------------------------------------------------
NOV40w  ------------------------------------------------------------
NOV40x  ------------------------------------------------------------
```

```
NOV40y  ------------------------------------------------------------
NOV40z  ------------------------------------------------------------
NOV40aa ------------------------------------------------------------
NOV40ab ------------------------------------------------------------
NOV40ac ------------------------------------------------------------

NOV40a  -----
NOV40b  -----
NOV40c  -----
NOV40d  -----
NOV40e  -----
NOV40f  -----
NOV40g  -----
NOV40h  -----
NOV40i  -----
NOV40j  -----
NOV40k  -----
NOV40l  -----
NOV40m  -----
NOV40n  -----
NOV40o  -----
NOV40p  -----
NOV40q  -----
NOV40r  -----
NOV40s  EIGRR
NOV40t  -----
NOV40u  -----
NOV40v  -----
NOV40w  -----
NOV40x  -----
NOV40y  -----
NOV40z  -----
NOV40aa -----
NOV40ab -----
NOV40ac -----

NOV40a  (SEQ ID NO:  486)
NOV40b  (SEQ ID NO:  488)
NOV40c  (SEQ ID NO:  490)
NOV40d  (SEQ ID NO:  492)
NOV40e  (SEQ ID NO:  494)
NOV40f  (SEQ ID NO:  496)
NOV40g  (SEQ ID NO:  498)
NOV40h  (SEQ ID NO:  500)
NOV40i  (SEQ ID NO:  502)
NOV40j  (SEQ ID NO:  504)
NOV40k  (SEQ ID NO:  506)
NOV40l  (SEQ ID NO:  508)
NOV40m  (SEQ ID NO:  510)
NOV40n  (SEQ ID NO:  512)
NOV40o  (SEQ ID NO:  514)
NOV40p  (SEQ ID NO:  516)
NOV40q  (SEQ ID NO:  518)
NOV40r  (SEQ ID NO:  520)
```

NOV40s (SEQ ID NO: 522) NOV40t (SEQ ID NO: 524) NOV40u (SEQ ID NO: 526) NOV40v (SEQ ID NO: 528) NOV40w (SEQ ID NO: 530) NOV40x (SEQ ID NO: 532) NOV40y (SEQ ID NO: 534) NOV40z (SEQ ID NO: 536) NOV40aa (SEQ ID NO: 538) NOV40ab (SEQ ID NO: 540) NOV40ac (SEQ ID NO: 542)

Further analysis of the NOV40a protein yielded the following properties shown in Table 40C.

Table 40C. Protein Sequence Properties NOV40a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 7;  pos.chg 3;  neg.chg 2
      H-region:  length 6;  peak value  -1.30
      PSG score:  -5.70

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1): -13.53
      possible cleavage site: between 60 and 61

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   3
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -9.39   Transmembrane  309 - 325
      PERIPHERAL  Likelihood =  2.54 (at 2515)
      ALOM score:  -9.39  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 316
      Charge difference: -6.5   C(-1.5) - N( 5.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 2   (cytoplasmic tail 1 to 309)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  3.03
      Hyd Moment(95):  4.83    G content:       0
      D/E content:     2       S/T content:     0
      Score: -7.62

Gavel: prediction of cleavage sites for mitochondrial preseq
```

```
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: RRPYCSLTKSRREKERR at    6
      bipartite: RKEQAQGKVNVFGRKLR at 1778
      content of basic residues:   9.9%
      NLS Score:  0.51

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: too long tail

Dileucine motif in the tail: found
      LL at 57

checking 63 PROSITE DNA binding motifs:
      Leucine zipper pattern (PS00029):  *** found ***
         LTPLRYDLRDRITRLGDVQYRL at 2196
 none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 55.5

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

-------------------------
```

```
Final Results (k = 9/23):

       34.8 %: nuclear
       26.1 %: mitochondrial
       13.0 %: Golgi
        8.7 %: cytoplasmic
        4.3 %: plasma membrane
        4.3 %: vesicles of secretory system
        4.3 %: extracellular, including cell wall
        4.3 %: peroxisomal

>> prediction for CG55069-01 is nuc (k=23)
```

A search of the NOV40a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 40D.

Table 40D. Geneseq Results for NOV40a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV40a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABP53586	Human NOV15a protein SEQ ID NO:36 - Homo sapiens, 2725 aa. [WO200262999-A2, 15-AUG-2002]	1..2725 1..2725	2725/2725 (100%) 2725/2725 (100%)	0.0
ABP53587	Human NOV15b protein SEQ ID NO:38 - Homo sapiens, 2721 aa. [WO200262999-A2, 15-AUG-2002]	1..2725 1..2721	2720/2725 (99%) 2720/2725 (99%)	0.0
ABP53588	Human NOV15c protein SEQ ID NO:40 - Homo sapiens, 2628 aa. [WO200262999-A2, 15-AUG-2002]	148..2725 54..2628	2553/2579 (98%) 2558/2579 (98%)	0.0
ABP53589	Human NOV15d protein SEQ ID NO:42 - Homo sapiens, 2613 aa. [WO200262999-A2, 15-AUG-2002]	148..2725 54..2613	2536/2579 (98%) 2540/2579 (98%)	0.0
ABB98401	Human NOV1, a TEN-M4 like protein - Homo sapiens, 2794 aa. [WO200255704-A2, 18-JUL-2002]	1..2725 1..2794	1917/2813 (68%) 2260/2813 (80%)	0.0

In a BLAST search of public sequence databases, the NOV40a protein was found to have homology to the proteins shown in the BLASTP data in Table 40E.

Table 40E. Public BLASTP Results for NOV40a				
Protein Accession Number	**Protein/Organism/Length**	**NOV40a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9WTS6	Ten-m3 - Mus musculus (Mouse), 2715 aa.	1..2725 1..2715	2663/2725 (97%) 2696/2725 (98%)	0.0
Q9JLC1	ODZ3 - Mus musculus (Mouse), 2346 aa (fragment).	354..2725 1..2346	2308/2372 (97%) 2334/2372 (98%)	0.0
Q9W7R4	Ten-m3 - Brachydanio rerio (Zebrafish) (Danio rerio), 2590 aa.	143..2725 27..2590	2116/2598 (81%) 2352/2598 (90%)	0.0
Q9WTS7	Ten-m4 - Mus musculus (Mouse), 2771 aa.	1..2725 1..2771	1867/2790 (66%) 2227/2790 (78%)	0.0
Q9DER5	Teneurin-2 - Gallus gallus (Chicken), 2802 aa (fragment).	173..2725 267..2802	1777/2561 (69%) 2138/2561 (83%)	0.0

PFam analysis indicates that the NOV40a protein contains the domains shown in the Table 40F.

Table 40F. Domain Analysis of NOV40a			
Pfam Domain	**NOV40a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
EGF	522..548	12/47 (26%) 24/47 (51%)	0.98
EGF	586..613	10/47 (21%) 19/47 (40%)	0.43
EGF	618..645	12/47 (26%) 22/47 (47%)	0.093
EGF	652..680	12/47 (26%) 22/47 (47%)	0.87
EGF	716..742	12/47 (26%) 23/47 (49%)	0.024
EGF	762..792	11/47 (23%) 21/47 (45%)	0.66
NHL	1497..1524	10/29 (34%) 22/29 (76%)	0.043

Example 41.

The NOV41 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 41A.

Table 41A. NOV41 Sequence Analysis

NOV41a, CG55343-03 DNA Sequence	SEQ ID NO: 553	943 bp
	ORF Start: ATG at 2	ORF Stop: TAA at 941

TATGAATTGGGTAAATGACAGCATCATACAGGAGTTTATTCTGCTGGGTTTCTCAGATCGACCTTGGC
TGGAGTTTCCACTCCTTGTGGTCTTCTTGATTTCTTACACTGTGACCATCTTTGGCAATCTGACCATT
ATTCTAGTGTCACGCCTGGACACCAAACTTCATACCCCCATGTATTTTTTTCTTACCAATCTATCACT
CCTGGATCTTTGTTACACCACATGTACAGTCCCACAAATGCTAGTAAATTTATGCAGCATCAGGAAAG
TAATCAGTTATCGTGGCTGTGTAGCCCAGCTTTTCATATTTCTGGCCTTGGGGGCTACTGAATATCTT
CTCCTGGCCGTCATGTCCTTTGATTGGTTTGTAGCTATTTGTCGGCCTCTCCATTACTCAGTTATCAT
GCACCAGAGACTCTGCCTCCAGTTGGCAGCTGCATCCTGGGTTACTGGTTTTAGTAACTCAGTGTGGT
TGTCTACCCTGACTCTCCAGCTGCCACTCTGTGACCCCTATGTGATAGATCACTTTCTCTGTGAAGTC
CCTGCACTGCTCAAGTTATCTTGTGTTGAGACAACAGCAAATGAGGCTGAACTATTCCTTGTCAGTGA
GCTCTTCCATCTAATACCCCTGACACTCATCCTTATATCATATGCTTTTATTGTCCGAGCAGTATTGA
GGATACAGTCTGCTGAAGGTCGACAAAAAGCATTTGGGACATGTGGTTCCCATCTAATTGTGGTGTCT
CTTTTTTATAGTACAGCCGTCTCTGTGTACCTGCAACCACCTTCGCCCAGCTCCAAGGACCAAGGAAA
GATGGTTTCTCTCTTCTATGGAATCATTGCACCCATGCTGAATCCCCTTATATATACACTTAGGAACA
AGGAGGTAAAGGAAGGCTTTAAAAGGTTGGTTGCAAGAGTCTTCTTAATCAAGAAA**TAA**

NOV41a, CG55343-03 Protein Sequence	SEQ ID NO: 554	313 aa	MW at 35443.7kD

MNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLHTPMYFFLTNLSL
LDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLAVMSFDWFVAICRPLHYSVIM
HQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDHFLCEVPALLKLSCVETTANEAELFLVSE
LFHLIPLTLILISYAFIVRAVLRIQSAEGRQKAFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGK
MVSLFYGIIAPMLNPLIYTLRNKEVKEGFKRLVARVFLIKK

NOV41b, 260568382 DNA Sequence	SEQ ID NO: 555	946 bp
	ORF Start: at 3	ORF Stop: TAG at 936

ATTGGGTAAATGACAGCATCATACAGGAGTTTATTCTGCTGGGTTTCTCAGATCGACCTTGGCTGGAG
TTTCCACTCCTTGTGGTCTTCTTGATTTCTTACACTGTGACCATCTTTGGCAATCTGACCATTATTCT
AGTGTCACGCCTGGACACCAAACTTCATACCCCCATGTATTTTTTTCTTACCAATCTATCACTCCTGG
ATCTTTGTTACACCACATGTACAGTCCCACAAATGCTAGTAAATTTATGCAGCATCAGGAAAGTAATC
AGTTATCGTGGCTGTGTAGCCCAGCTTTTCATATTTCTGGCCTTGGGGGCTACTGAATATCTTCTCCT
GGCCGTCATGTCCTTTGATAGGTTTGTAGCTATTTGTCGGCCTCTCCATTACTCAGTTATCATGCACC
AGAGACTCTGCCTCCAGTTGGCAGCTGCATCCTGGGTTACTGGTTTTAGTAACTCAGTGTGGTTGTCT
ACCCTGACTCTCCAGCTGCCACTCTGTGACCCCTATGTGATAGATCACTTTCTCTGTGAAGTCCCTGC
ACTGCTCAAGTTATCTTGTGTTGAGACAACAGCAAATGAGGCTGAACTATTCCTTGTCAGTGAGCTCT
TCCATCTAATACCCCTGACACTCATCCTTATATCATATGCTTTTATTGTCCGAGCAGTATTGAGGATA
CAGTCTGCTGAAGGTCGACAAAAAGCATTTGGGACATGTGGTTCCCATCTAATTGTGGTGTCTCTTTT
TTATAGTACAGCCGTCTCTGTGTACCTGCAACCACCTTCGCCCAGCTCCAAGGACCAAGGAAAGATGG
TTTCTCTCTTCTATGGAATCATTGCACCCATGCTGAATCCCCTTATATATACACTTAGGAACAAGGAG
GTAAAGGAAGGCTTTAAAAGGTTGGTTGCAAGAGTCTTCTTAATCAAGAAA**TAG**GGATTCGG

NOV41b, 260568382 Protein Sequence	SEQ ID NO: 556	311 aa	MW at 35168.4kD

WVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLHTPMYFFLTNLSLLD
LCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLAVMSFDRFVAICRPLHYSVIMHQ
RLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDHFLCEVPALLKLSCVETTANEAELFLVSELF
HLIPLTLILISYAFIVRAVLRIQSAEGRQKAFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGKMV

SLFYGIIAPMLNPLIYTLRNKEVKEGFKRLVARVFLIKK		
NOV41c, 314361391 DNA Sequence	SEQ ID NO: 557	960 bp
	ORF Start: at 1	ORF Stop: end of sequence

CGCGGATCCACCATGAATTGGGTAAATGACAGCATCATACAGGAGTTTATTTTGCTGGGTTTCTCAGA
TCGACCTTGGCTGGAGTTTCCACTCCTTGTGGTCTTCTTGATTTCTTACACTGTGACCATCTTTGGCA
ATCTGACCATTATTCTAGTGTCACGCCTGGACACCAAACTTCATACCCCCATGTATTTTTTTCTTACC
AATCTATCACTCCTGGATCTTTGTTACACCACATGTACAGTCCCACAAATGCTAGTAAATTTATGCAG
CATCAGGAAAGTAATCAGTTATCGTGGCTGTGTAGCCCAGCTTTTCATATTTCTGGCCTTGGGGGCTA
CTGAATATCTTCTCCTGGCCGTCATGTCCTTTGATAGGTTTGTAGCTATTTGTCGGCCTCTCCATTAC
TCAGTTATCATGCACCAGAGACTCTGCCTCCAGTTGGCAGCTGCATCCTGGGTTACTGGTTTTAGTAA
CTCAGTGTGGTTGTCTACCCTGACTCTCCAGCTGCCACTCTGTGACCCCTATGTGATAGATCACTTTC
TCTGTGAAGTCCCTGCACTGCTCAAGTTATCTTGTGTTGAGACAACAGCAAATGAGGCTGAACTATTC
CTTGTCAGTGAGCTCTTCCATCTAATACCCCTGACACTCATCCTTATATCATATGCTTTTATTGTCCG
AGCAGTATTGAGGATACAGTCTGCTGAAGGTCGACAAAAAGCATTTGGGACATGTGGTTCCCATCTAA
TTGTGGTGTCTCTTTTTTATAGTACAGCCGTCTCTGTGTACCTGCAACCACCTTCGCCCAGCTCCAAG
GACCAAGGAAAGATGGTTTCTCTCTTCTATGGAATCATTGCACCCATGCTGAATCCCCTTATATATAC
ACTTAGGAACAAGGAGGTAAAGGAAGGCTTTAAAAGGTTGGTTGCAAGAGTCTTCTTAATCAAGAAAC
TCGAGGCG

NOV41c, 314361391 Protein Sequence	SEQ ID NO: 558	320 aa	MW at 36128.5kD

RGSTMNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLHTPMYFFLT
NLSLLDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLAVMSFDRFVAICRPLHY
SVIMHQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDHFLCEVPALLKLSCVETTANEAELF
LVSELFHLIPLTLILISYAFIVRAVLRIQSAEGRQKAFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSK
DQGKMVSLFYGIIAPMLNPLIYTLRNKEVKEGFKRLVARVFLIKKLEA

NOV41d, 317286137 DNA Sequence	SEQ ID NO: 559	321 bp	
	ORF Start: at 1	ORF Stop: end of sequence	

CGCGGATCCACCATGAATTGGGTAAATGACAGCATCATACAGGAGTTTATTCTGCTGGGTTTCTCAGA
TCGACCTTGGCTGGAGTTTCCACTCCTTGTGGTCTTCTTGATTTCTTACACTGTGACCATCTTTGGCA
ATCTGACCATTATTCTAGTGTCACGCCTGGACACCAAACTTCATACCCCCATGTATTTTTTTCTTACC
AATCTATCACTCCTGGATCTTTGTTACACCACATGTACAGTCCCACAAATGCTAGTAAATTTATGCAG
CATCAGGAAAGTAATCAGTTATCGTGGCTGTGTAGCCCAGCTCGAGGCG

NOV41d, 317286137 Protein Sequence	SEQ ID NO: 560	107 aa	MW at 12270.4kD

RGSTMNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLHTPMYFFLT
NLSLLDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLEA

NOV41e, CG55343-01 DNA Sequence	SEQ ID NO: 561	943 bp
	ORF Start: ATG at 2	ORF Stop: TAA at 941

TATGAATTGGGTAAATGACAGCATCATACAGGAGTTTATTCTGCTGGGTTTCTCAGATCGACCTTGGC
TGGAGTTTCCACTCCTTGTGGTCTTCTTGATTTCTTACACTGTGACCATCTTTGGCAATCTGACCATT
ATTCTAGTGTCACGCCTGGACACCAAACTTCATACCCCCATGTATTTTTTTCTTACCAATCTATCACT
CCTGGATCTTTGTTACACCACATGTACAGTCCCACAAATGCTAGTAAATTTATGCAGCATCAGGAAAG
TAATCAGTTATCGTGGCTGTGTAGCCCAGCTTTTCATATTTCTGGCCTTGGGGGCTACTGAATATCTT
CTCCTGGCCGTCATGTCCTTTGATTGGTTTGTAGCTATTTGTCGGCCTCTCCATTACTCAGTTATCAT
GCACCAGAGACTCTGCCTCCAGTTGGCAGCTGCATCCTGGGTTACTGGTTTTAGTAACTCAGTGTGGT
TGTCTACCCTGACTCTCCAGCTGCCACTCTGTGACCCCTATGTGATAGATCACTTTCTCTGTGAAGTC
CCTGCACTGCTCAAGTTATCTTGTGTTGAGACAACAGCAAATGAGGCTGAACTATTCCTTGTCAGTGA
GCTCTTCCATCTAATACCCCTGACACTCATCCTTATATCATATGCTTTTATTGTCCGAGCAGTATTGA
GGATACAGTCTGCTGAAGGTCGACAAAAAGCATTTGGGACATGTGGTTCCCATCTAATTGTGGTGTCT
CTTTTTTATAGTACAGCCGTCTCTGTGTACCTGCAACCACCTTCGCCCAGCTCCAAGGACCAAGGAAA
GATGGTTTCTCTCTTCTATGGAATCATTGCACCCATGCTGAATCCCCTTATATATACACTTAGGAACA
AGGAGGTAAAGGAAGGCTTTAAAAGGTTGGTTGCAAGAGTCTTCTTAATCAAGAAATAA

NOV41e, CG55343-01 Protein Sequence	SEQ ID NO: 562	313 aa	MW at 35443.7kD

MNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLHTPMYFFLTNLSL
LDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLAVMSFDWFVAICRPLHYSVIM
HQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDHFLCEVPALLKLSCVETTANEAELFLVSE
LFHLIPLTLILISYAFIVRAVLRIQSAEGRQKAFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGK
MVSLFYGIIAPMLNPLIYTLRNKEVKEGFKRLVARVFLIKK

NOV41f, CG55343-02 DNA Sequence	SEQ ID NO: 563	993 bp
	ORF Start: ATG at 15	ORF Stop: TAA at 954

TTGAAACATTAATCATGAATTGGGTAAATGACAGCATCATACAGGAGTTTATTCTGCTGGGTTTCTCA
GATCGACCTTGGCTGGAGTTTCCACTCCTTGTGGTCTTCTTGATTTCTTACACTGTGACCATCTTTGG
CAATCTGACCATTATTCTAGTGTCACGCCTGGACACCAAACTTCATACCCCCATGTATTTTTTTCTTA
CCAATCTATCACTCCTGGATCTTTGTTACACCACATGTACAGTCCCACAAATGCTAGTAAATTTATGC
AGCATCAGGAAAGTAATCAGTTATCGTGGCTGTGTAGCCCAGCTTTTCATATTTCTGGCCTTGGGGGC
TACTGAATATCTTCTCCTGGCCGTCATGTCCTTTGATAGGTTTGTAGCTATTTGTCGGCCTCTCCATT
ACTCAGTTATCATGCACCAGAGACTCTGCCTCCAGTTGGCAGCTGCATCCTGGGTTACTGGTTTTAGT
AACTCAGTGTGGTTGTCTACCCTGACTCTCCAGCTGCCACTCTGTGACCCCTATGTGATAGATCACTT
TCTCTGTGAAGTCCCTGCACTGCTCAAGTTATCTTGTGTTGAGACAACAGCAAATGAGGCTGAACTAT
TCCTTGTCAGTGAGCTCTTCCATCTAATACCCCTGACACTCATCCTTATATCATATGCTTTTATTGTC
CGAGCAGTATTGAGGATACAGTCTGCTGAAGGTCGACAAAAAGCATTTGGGACATGTGGTTCCCATCT
AATTGTGGTGTCTCTTTTTTATAGTACAGCCGTCTCTGTGTACCTGCAACCACCTTCGCCCAGCTCCA
AGGACCAAGGAAAGATGGTTTCTCTCTTCTATGGAATCATTGCACCCATGCTGAATCCCCTTATATAT
ACACTTAGGAACAAGGAGGTAAAGGAAGGCTTTAAAAGGTTGGTTGCAAGAGTCTTCTTAATCAAGAA
ATAAGAAATATGCAAATGATAAGCTTTGCTAAAGACAAAAT

NOV41f, CG55343-02 Protein Sequence	SEQ ID NO: 564	313 aa	MW at 35413.7kD

MNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLHTPMYFFLTNLSL
LDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLAVMSFDRFVAICRPLHYSVIM
HQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDHFLCEVPALLKLSCVETTANEAELFLVSE
LFHLIPLTLILISYAFIVRAVLRIQSAEGRQKAFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGK
MVSLFYGIIAPMLNPLIYTLRNKEVKEGFKRLVARVFLIKK

NOV41g, CG55343-04 DNA Sequence	SEQ ID NO: 565	60 bp
	ORF Start: at 1	ORF Stop: end of sequence

GACAGCATCATACAGGAGTTTATTCTGCTGGGTTTCTCAGATCGACCTTGGCTGGAGTTT

NOV41g, CG55343-04 Protein Sequence	SEQ ID NO: 566	20 aa	MW at 2425.7kD

DSIIQEFILLGFSDRPWLEF

NOV41h, CG55343-05 DNA Sequence	SEQ ID NO: 567	60 bp
	ORF Start: at 1	ORF Stop: end of sequence

GAAGTCCCTGCACTGCTCAAGTTATCTTGTGTTGAGACAACAGCAAATGAGGCTGAACTA

NOV41h, CG55343-05 Protein Sequence	SEQ ID NO: 568	20 aa	MW at 2130.4kD

EVPALLKLSCVETTANEAEL

NOV41i, CG55343-06 DNA Sequence	SEQ ID NO: 569	946 bp
	ORF Start: at 3	ORF Stop: TAG at 936

ATTGGGTAAATGACAGCATCATACAGGAGTTTATTCTGCTGGGTTTCTCAGATCGACCTTGGCTGGAG
TTTCCACTCCTTGTGGTCTTCTTGATTTCTTACACTGTGACCATCTTTGGCAATCTGACCATTATTCT
AGTGTCACGCCTGGACACCAAACTTCATACCCCCATGTATTTTTTTCTTACCAATCTATCACTCCTGG
ATCTTTGTTACACCACATGTACAGTCCCACAAATGCTAGTAAATTTATGCAGCATCAGGAAAGTAATC
AGTTATCGTGGCTGTGTAGCCCAGCTTTTCATATTTCTGGCCTTGGGGGCTACTGAATATCTTCTCCT

GGCCGTCATGTCCTTTGATAGGTTTGTAGCTATTTGTCGGCCTCTCCATTACTCAGTTATCATGCACC
AGAGACTCTGCCTCCAGTTGGCAGCTGCATCCTGGGTTACTGGTTTTAGTAACTCAGTGTGGTTGTCT
ACCCTGACTCTCCAGCTGCCACTCTGTGACCCCTATGTGATAGATCACTTTCTCTGTGAAGTCCCTGC
ACTGCTCAAGTTATCTTGTGTTGAGACAACAGCAAATGAGGCTGAACTATTCCTTGTCAGTGAGCTCT
TCCATCTAATACCCCTGACACTCATCCTTATATCATATGCTTTTATTGTCCGAGCAGTATTGAGGATA
CAGTCTGCTGAAGGTCGACAAAAAGCATTTGGGACATGTGGTTCCCATCTAATTGTGGTGTCTCTTTT
TTATAGTACAGCCGTCTCTGTGTACCTGCAACCACCTTCGCCCAGCTCCAAGGACCAAGGAAAGATGG
TTTCTCTCTTCTATGGAATCATTGCACCCATGCTGAATCCCCTTATATATACACTTAGGAACAAGGAG
GTAAAGGAAGGCTTTAAAAGGTTGGTTGCAAGAGTCTTCTTAATCAAGAAATAGGGATTCGG

NOV41i, CG55343-06 Protein Sequence	SEQ ID NO: 570	311 aa	MW at 35168.4kD

WVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLHTPMYFFLTNLSLLD
LCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLAVMSFDRFVAICRPLHYSVIMHQ
RLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDHFLCEVPALLKLSCVETTANEAELFLVSELF
HLIPLTLILISYAFIVRAVLRIQSAEGRQKAFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGKMV
SLFYGIIAPMLNPLIYTLRNKEVKEGFKRLVARVFLIKK

NOV41j, SNP13373740 of CG55343-03, DNA Sequence	SEQ ID NO: 571	943 bp
	ORF Start: ATG at 2	ORF Stop: TAA at 941
	SNP Pos: 568	SNP Change: T to C

TATGAATTGGGTAAATGACAGCATCATACAGGAGTTTATTCTGCTGGGTTTCTCAGATCGACCTTGGC
TGGAGTTTCCACTCCTTGTGGTCTTCTTGATTTCTTACACTGTGACCATCTTTGGCAATCTGACCATT
ATTCTAGTGTCACGCCTGGACACCAAACTTCATACCCCCATGTATTTTTTTCTTACCAATCTATCACT
CCTGGATCTTTGTTACACCACATGTACAGTCCCACAAATGCTAGTAAATTTATGCAGCATCAGGAAAG
TAATCAGTTATCGTGGCTGTGTAGCCCAGCTTTTCATATTTCTGGCCTTGGGGGCTACTGAATATCTT
CTCCTGGCCGTCATGTCCTTTGATTGGTTTGTAGCTATTTGTCGGCCTCTCCATTACTCAGTTATCAT
GCACCAGAGACTCTGCCTCCAGTTGGCAGCTGCATCCTGGGTTACTGGTTTTAGTAACTCAGTGTGGT
TGTCTACCCTGACTCTCCAGCTGCCACTCTGTGACCCCTATGTGATAGATCACTTTCTCTGTGAAGTC
CCTGCACTGCTCAAGTTATCTTGCGTTGAGACAACAGCAAATGAGGCTGAACTATTCCTTGTCAGTGA
GCTCTTCCATCTAATACCCCTGACACTCATCCTTATATCATATGCTTTTATTGTCCGAGCAGTATTGA
GGATACAGTCTGCTGAAGGTCGACAAAAAGCATTTGGGACATGTGGTTCCCATCTAATTGTGGTGTCT
CTTTTTTATAGTACAGCCGTCTCTGTGTACCTGCAACCACCTTCGCCCAGCTCCAAGGACCAAGGAAA
GATGGTTTCTCTCTTCTATGGAATCATTGCACCCATGCTGAATCCCCTTATATATACACTTAGGAACA
AGGAGGTAAAGGAAGGCTTTAAAAGGTTGGTTGCAAGAGTCTTCTTAATCAAGAAATAA

NOV41j, SNP13373740 of CG55343-03, Protein Sequence	SEQ ID NO: 572	313 aa	MW at 35443.7kD
	SNP Pos: 189		SNP Change: Cys to Cys

MNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLHTPMYFFLTNLSL
LDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLAVMSFDWFVAICRPLHYSVIM
HQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDHFLCEVPALLKLSCVETTANEAELFLVSE
LFHLIPLTLILISYAFIVRAVLRIQSAEGRQKAFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGK
MVSLFYGIIAPMLNPLIYTLRNKEVKEGFKRLVARVFLIKK

NOV41k, SNP13376425 of CG55343-03, DNA Sequence	SEQ ID NO: 573	943 bp
	ORF Start: ATG at 2	ORF Stop: TAA at 941
	SNP Pos: 758	SNP Change: A to G

TATGAATTGGGTAAATGACAGCATCATACAGGAGTTTATTCTGCTGGGTTTCTCAGATCGACCTTGGC
TGGAGTTTCCACTCCTTGTGGTCTTCTTGATTTCTTACACTGTGACCATCTTTGGCAATCTGACCATT
ATTCTAGTGTCACGCCTGGACACCAAACTTCATACCCCCATGTATTTTTTTCTTACCAATCTATCACT
CCTGGATCTTTGTTACACCACATGTACAGTCCCACAAATGCTAGTAAATTTATGCAGCATCAGGAAAG
TAATCAGTTATCGTGGCTGTGTAGCCCAGCTTTTCATATTTCTGGCCTTGGGGGCTACTGAATATCTT
CTCCTGGCCGTCATGTCCTTTGATTGGTTTGTAGCTATTTGTCGGCCTCTCCATTACTCAGTTATCAT
GCACCAGAGACTCTGCCTCCAGTTGGCAGCTGCATCCTGGGTTACTGGTTTTAGTAACTCAGTGTGGT
TGTCTACCCTGACTCTCCAGCTGCCACTCTGTGACCCCTATGTGATAGATCACTTTCTCTGTGAAGTC
CCTGCACTGCTCAAGTTATCTTGTGTTGAGACAACAGCAAATGAGGCTGAACTATTCCTTGTCAGTGA
GCTCTTCCATCTAATACCCCTGACACTCATCCTTATATCATATGCTTTTATTGTCCGAGCAGTATTGA

GGATACAGTCTGCTGAAGGTCGACAAAAAGCATTTGGGACATGTGGTTCCCATCTAATTGTGGTGTCT
CTTTTTTATGGTACAGCCGTCTCTGTGTACCTGCAACCACCTTCGCCCAGCTCCAAGGACCAAGGAAA
GATGGTTTCTCTCTTCTATGGAATCATTGCACCCATGCTGAATCCCCTTATATATACACTTAGGAACA
AGGAGGTAAAGGAAGGCTTTAAAAGGTTGGTTGCAAGAGTCTTCTTAATCAAGAAA**TAA**

NOV41k, SNP13376425 of CG55343-03, Protein Sequence	SEQ ID NO: 574	313 aa	MW at 35413.7kD
	SNP Pos: 253		SNP Change: Ser to Gly

MNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLHTPMYFFLTNLSL
LDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLAVMSFDWFVAICRPLHYSVIM
HQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDHFLCEVPALLKLSCVETTANEAELFLVSE
LFHLIPLTLILISYAFIVRAVLRIQSAEGRQKAFGTCGSHLIVVSLFYGTAVSVYLQPPSPSSKDQGK
MVSLFYGIIAPMLNPLIYTLRNKEVKEGFKRLVARVFLIKK

NOV41l, SNP13376424 of CG55343-03, DNA Sequence	SEQ ID NO: 575	943 bp
	ORF Start: ATG at 2	ORF Stop: TAA at 941
	SNP Pos: 810	SNP Change: A to G

TATGAATTGGGTAAATGACAGCATCATACAGGAGTTTATTCTGCTGGGTTTCTCAGATCGACCTTGGC
TGGAGTTTCCACTCCTTGTGGTCTTCTTGATTTCTTACACTGTGACCATCTTTGGCAATCTGACCATT
ATTCTAGTGTCACGCCTGGACACCAAACTTCATACCCCCATGTATTTTTTTCTTACCAATCTATCACT
CCTGGATCTTTGTTACACCACATGTACAGTCCCACAAATGCTAGTAAATTTATGCAGCATCAGGAAAG
TAATCAGTTATCGTGGCTGTGTAGCCCAGCTTTTCATATTTCTGGCCTTGGGGGCTACTGAATATCTT
CTCCTGGCCGTCATGTCCTTTGATTGGTTTGTAGCTATTTGTCGGCCTCTCCATTACTCAGTTATCAT
GCACCAGAGACTCTGCCTCCAGTTGGCAGCTGCATCCTGGGTTACTGGTTTTAGTAACTCAGTGTGGT
TGTCTACCCTGACTCTCCAGCTGCCACTCTGTGACCCCTATGTGATAGATCACTTTCTCTGTGAAGTC
CCTGCACTGCTCAAGTTATCTTGTGTTGAGACAACAGCAAATGAGGCTGAACTATTCCTTGTCAGTGA
GCTCTTCCATCTAATACCCCTGACACTCATCCTTATATCATATGCTTTTATTGTCCGAGCAGTATTGA
GGATACAGTCTGCTGAAGGTCGACAAAAAGCATTTGGGACATGTGGTTCCCATCTAATTGTGGTGTCT
CTTTTTTATAGTACAGCCGTCTCTGTGTACCTGCAACCACCTTCGCCCAGCTCCAAGGACCGAGGAAA
GATGGTTTCTCTCTTCTATGGAATCATTGCACCCATGCTGAATCCCCTTATATATACACTTAGGAACA
AGGAGGTAAAGGAAGGCTTTAAAAGGTTGGTTGCAAGAGTCTTCTTAATCAAGAAA**TAA**

NOV41l, SNP13376424 of CG55343-03, Protein Sequence	SEQ ID NO: 576	313 aa	MW at 35471.8kD
	SNP Pos: 270		SNP Change: Gln to Arg

MNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLHTPMYFFLTNLSL
LDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLAVMSFDWFVAICRPLHYSVIM
HQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDHFLCEVPALLKLSCVETTANEAELFLVSE
LFHLIPLTLILISYAFIVRAVLRIQSAEGRQKAFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDRGK
MVSLFYGIIAPMLNPLIYTLRNKEVKEGFKRLVARVFLIKK

NOV41m, SNP13376423 of CG55343-03, DNA Sequence	SEQ ID NO: 577	943 bp
	ORF Start: ATG at 2	ORF Stop: TAA at 941
	SNP Pos: 908	SNP Change: A to G

TATGAATTGGGTAAATGACAGCATCATACAGGAGTTTATTCTGCTGGGTTTCTCAGATCGACCTTGGC
TGGAGTTTCCACTCCTTGTGGTCTTCTTGATTTCTTACACTGTGACCATCTTTGGCAATCTGACCATT
ATTCTAGTGTCACGCCTGGACACCAAACTTCATACCCCCATGTATTTTTTTCTTACCAATCTATCACT
CCTGGATCTTTGTTACACCACATGTACAGTCCCACAAATGCTAGTAAATTTATGCAGCATCAGGAAAG
TAATCAGTTATCGTGGCTGTGTAGCCCAGCTTTTCATATTTCTGGCCTTGGGGGCTACTGAATATCTT
CTCCTGGCCGTCATGTCCTTTGATTGGTTTGTAGCTATTTGTCGGCCTCTCCATTACTCAGTTATCAT
GCACCAGAGACTCTGCCTCCAGTTGGCAGCTGCATCCTGGGTTACTGGTTTTAGTAACTCAGTGTGGT
TGTCTACCCTGACTCTCCAGCTGCCACTCTGTGACCCCTATGTGATAGATCACTTTCTCTGTGAAGTC
CCTGCACTGCTCAAGTTATCTTGTGTTGAGACAACAGCAAATGAGGCTGAACTATTCCTTGTCAGTGA
GCTCTTCCATCTAATACCCCTGACACTCATCCTTATATCATATGCTTTTATTGTCCGAGCAGTATTGA
GGATACAGTCTGCTGAAGGTCGACAAAAAGCATTTGGGACATGTGGTTCCCATCTAATTGTGGTGTCT
CTTTTTTATAGTACAGCCGTCTCTGTGTACCTGCAACCACCTTCGCCCAGCTCCAAGGACCAAGGAAA
GATGGTTTCTCTCTTCTATGGAATCATTGCACCCATGCTGAATCCCCTTATATATACACTTAGGAACA
AGGAGGTAAAGGAAGGCTTTAAAGGGTTGGTTGCAAGAGTCTTCTTAATCAAGAAA**TAA**

NOV41m, SNP13376423 of CG55343-03, Protein Sequence	SEQ ID NO: 578	313 aa	MW at 35344.6kD
	SNP Pos: 303		SNP Change: Arg to Gly

MNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLHTPMYFFLTNLSL
LDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLAVMSFDWFVAICRPLHYSVIM
HQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDHFLCEVPALLKLSCVETTANEAELFLVSE
LFHLIPLTLILISYAFIVRAVLRIQSAEGRQKAFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGK
MVSLFYGIIAPMLNPLIYTLRNKEVKEGFKGLVARVFLIKK

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 41B.

Table 41B. Comparison of the NOV41 protein sequences.

```
NOV41a  ----MNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLH
NOV41b  ------WVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLH
NOV41c  RGSTMNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLH
NOV41d  RGSTMNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLH
NOV41e  ----MNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLH
NOV41f  ----MNWVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLH
NOV41g  ---------DSIIQEFILLGFSDRPWLEF----------------------------
NOV41h  ------------------------------------------------------------
NOV41i  ------WVNDSIIQEFILLGFSDRPWLEFPLLVVFLISYTVTIFGNLTIILVSRLDTKLH

NOV41a  TPMYFFLTNLSLLDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLA
NOV41b  TPMYFFLTNLSLLDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLA
NOV41c  TPMYFFLTNLSLLDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLA
NOV41d  TPMYFFLTNLSLLDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLEA-------------
NOV41e  TPMYFFLTNLSLLDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLA
NOV41f  TPMYFFLTNLSLLDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLA
NOV41g  ------------------------------------------------------------
NOV41h  ------------------------------------------------------------
NOV41i  TPMYFFLTNLSLLDLCYTTCTVPQMLVNLCSIRKVISYRGCVAQLFIFLALGATEYLLLA

NOV41a  VMSFDWFVAICRPLHYSVIMHQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDH
NOV41b  VMSFDRFVAICRPLHYSVIMHQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDH
NOV41c  VMSFDRFVAICRPLHYSVIMHQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDH
NOV41d  ------------------------------------------------------------
NOV41e  VMSFDWFVAICRPLHYSVIMHQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDH
NOV41f  VMSFDRFVAICRPLHYSVIMHQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDH
NOV41g  ------------------------------------------------------------
NOV41h  ------------------------------------------------------------
NOV41i  VMSFDRFVAICRPLHYSVIMHQRLCLQLAAASWVTGFSNSVWLSTLTLQLPLCDPYVIDH

NOV41a  FLCEVPALLKLSCVETTANEAELFLVSELFHLIPLTLILISYAFIVRAVLRIQSAEGRQK
NOV41b  FLCEVPALLKLSCVETTANEAELFLVSELFHLIPLTLILISYAFIVRAVLRIQSAEGRQK
NOV41c  FLCEVPALLKLSCVETTANEAELFLVSELFHLIPLTLILISYAFIVRAVLRIQSAEGRQK
NOV41d  ------------------------------------------------------------
NOV41e  FLCEVPALLKLSCVETTANEAELFLVSELFHLIPLTLILISYAFIVRAVLRIQSAEGRQK
NOV41f  FLCEVPALLKLSCVETTANEAELFLVSELFHLIPLTLILISYAFIVRAVLRIQSAEGRQK
NOV41g  ------------------------------------------------------------
NOV41h  ---EVPALLKLSCVETTANEAEL--------------------------------------
NOV41i  FLCEVPALLKLSCVETTANEAELFLVSELFHLIPLTLILISYAFIVRAVLRIQSAEGRQK
```

```
NOV41a  AFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGKMVSLFYGIIAPMLNPLIYTLRNKE
NOV41b  AFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGKMVSLFYGIIAPMLNPLIYTLRNKE
NOV41c  AFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGKMVSLFYGIIAPMLNPLIYTLRNKE
NOV41d  ------------------------------------------------------------
NOV41e  AFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGKMVSLFYGIIAPMLNPLIYTLRNKE
NOV41f  AFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGKMVSLFYGIIAPMLNPLIYTLRNKE
NOV41g  ------------------------------------------------------------
NOV41h  ------------------------------------------------------------
NOV41i  AFGTCGSHLIVVSLFYSTAVSVYLQPPSPSSKDQGKMVSLFYGIIAPMLNPLIYTLRNKE

NOV41a  VKEGFKRLVARVFLIKK---
NOV41b  VKEGFKRLVARVFLIKK---
NOV41c  VKEGFKRLVARVFLIKKLEA
NOV41d  --------------------
NOV41e  VKEGFKRLVARVFLIKK---
NOV41f  VKEGFKRLVARVFLIKK---
NOV41g  --------------------
NOV41h  --------------------
NOV41i  VKEGFKRLVARVFLIKK---

NOV41a  (SEQ ID NO:  554)
NOV41b  (SEQ ID NO:  556)
NOV41c  (SEQ ID NO:  558)
NOV41d  (SEQ ID NO:  560)
NOV41e  (SEQ ID NO:  562)
NOV41f  (SEQ ID NO:  564)
NOV41g  (SEQ ID NO:  566)
NOV41h  (SEQ ID NO:  568)
NOV41i  (SEQ ID NO:  570)
```

Further analysis of the NOV41a protein yielded the following properties shown in Table 41C.

Table 41C. Protein Sequence Properties NOV41a

SignalP analysis:	Cleavage site between residues 39 and 40

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 0;  neg.chg 2
      H-region:  length 7;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -4.39
      possible cleavage site: between 38 and 39

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   5
      INTEGRAL    Likelihood = -4.78   Transmembrane   25 -  41
```

```
      INTEGRAL    Likelihood = -5.15   Transmembrane  101 - 117
      INTEGRAL    Likelihood = -5.47   Transmembrane  206 - 222
      INTEGRAL    Likelihood = -0.80   Transmembrane  244 - 260
      INTEGRAL    Likelihood = -0.75   Transmembrane  273 - 289
      PERIPHERAL  Likelihood =  3.02 (at 173)
      ALOM score:  -5.47  (number of TMSs: 5)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 32
      Charge difference:  3.5   C( 1.5) - N(-2.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  8.26
      Hyd Moment(95): 10.83    G content:       0
      D/E content:     2       S/T content:     1
      Score: -5.89

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.3%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: FLIK

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none
```

```
Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       22.2 %: vacuolar
       11.1 %: Golgi
       11.1 %: mitochondrial

>> prediction for CG55343-03 is end (k=9)
```

A search of the NOV41a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 41D.

Table 41D. Geneseq Results for NOV41a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV41a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAE09961	G-protein coupled-receptor (GPCR) NOV1b protein - Unidentified, 313 aa. [WO200166746-A2, 13-SEP-2001]	1..313 1..313	313/313 (100%) 313/313 (100%)	e-180
AAU95518	Human olfactory and pheromone G protein-coupled receptor #5 - Homo sapiens, 313 aa. [WO200224726-A2, 28-MAR-2002]	1..313 1..313	312/313 (99%) 312/313 (99%)	e-178
ABP95746	Human GPCR polypeptide SEQ ID NO 302 - Homo sapiens, 313 aa. [WO200216548-A2, 28-FEB-2002]	1..313 1..313	312/313 (99%) 312/313 (99%)	e-178
AAE09960	G-protein coupled-receptor (GPCR) NOV1a protein - Unidentified, 313 aa. [WO200166746-A2, 13-SEP-2001]	1..313 1..313	312/313 (99%) 312/313 (99%)	e-178
AAG71842	Human olfactory receptor polypeptide, SEQ ID NO: 1523 - Homo sapiens, 313 aa. [WO200127158-A2, 19-APR-2001]	1..313 1..313	312/313 (99%) 312/313 (99%)	e-178

In a BLAST search of public sequence databases, the NOV41a protein was found to have homology to the proteins shown in the BLASTP data in Table 41E.

Table 41E. Public BLASTP Results for NOV41a				
Protein Accession Number	**Protein/Organism/Length**	**NOV41a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
P58173	Olfactory receptor 2B6 (Hs6M1-32) (Olfactory receptor 6-31) (OR6-31) - Homo sapiens (Human), 313 aa.	1..313 1..313	312/313 (99%) 312/313 (99%)	e-178
Q8VFH1	Olfactory receptor MOR256-11 - Mus musculus (Mouse), 313 aa.	1..312 1..312	254/312 (81%) 273/312 (87%)	e-145
Q9GZK3	Olfactory receptor 2B2 (Olfactory receptor 6-1) (OR6-1) (Hs6M1-10) - Homo sapiens (Human), 357 aa.	1..310 1..310	252/310 (81%) 277/310 (89%)	e-143
Q8VFH2	Olfactory receptor MOR256-10 - Mus musculus (Mouse), 313 aa.	1..313 1..313	247/313 (78%) 275/313 (86%)	e-140
Q63394	OL1 receptor - Rattus norvegicus (Rat), 313 aa.	1..313 1..313	244/313 (77%) 273/313 (86%)	e-139

PFam analysis indicates that the NOV41a protein contains the domains shown in the Table 41F.

Table 41F. Domain Analysis of NOV41a			
Pfam Domain	**NOV41a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	41..290	56/276 (20%) 176/276 (64%)	1.3e-33
TAS2R	99..308	51/309 (17%) 126/309 (41%)	0.27

Example 42.

The NOV42 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 42A.

Table 42A. NOV42 Sequence Analysis

NOV42a, CG55358-04 DNA Sequence	SEQ ID NO: 579	947 bp
	ORF Start: ATG at 2	ORF Stop: TAG at 938

TATGGAAAACGATAATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACC
TAGAGCTGATCGTCTTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATT
GTCTTGCTTTCAGCTCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATT
CCTGGACATGTGTTTCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGA
CCATCAGCTATGTGGGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTC
CTCCTGGCTGTCATGGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCAT
GCACCCACGTCTCTGTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAA
TGGCACCCCAGACATTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATG
CCAGCACTAATTGGTATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAAT
CTTTATCATCCTGGCACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTA
GGATCAAGTCAGCTGCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCT
CTCTTCTATGGTACAATCATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAA
GTTTCTTACCCTTTTCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACA
AAGATGTTAAAGAGGCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATATAGTAAGGGA

NOV42a, CG55358-04 Protein Sequence	SEQ ID NO: 580	312 aa	MW at 34640.9kD

MENDNTSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSF
LDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIM
HPRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAI
FIVLAPLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGK
FLTLFYTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42b, CG55358-03 DNA Sequence	SEQ ID NO: 581		932 bp
	ORF Start: at 3		ORF Stop: TAG at 924

ATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGAGCTGATCGTC
TTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCTTGCTTTCAGC
TCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGGACATGTGTT
TCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCATCAGCTATGTG
GGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGTCAT
GGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTCTCT
GTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGACA
TTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGG
TATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGG
CACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCAAGTCAGCT
GCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCTTCTATGGTAC
AATTATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTTCTTACCCTTT
TCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGATGTTAAAGAG
GCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATATAGTAAGGG

NOV42b, CG55358-03 Protein Sequence	SEQ ID NO: 582	307 aa	MW at 34051.3kD

TSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSFLDMCF
TTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMHPRLC
GQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFIILA
PLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLF
YTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42c, 317863291 DNA Sequence	SEQ ID NO: 583	693 bp
	ORF Start: at 1	ORF Stop: end of sequence

CGCGGATCCACCATGTACAAGTTCTTTCGAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCT
AGAGCTGATCGTCTTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTG
TCTTGCTTTCAGCTCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTC

CTGGACATGTGTTTCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGAC
CATCAGCTATGTGGGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCC
TCCTGGCTGTCATGGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATG
CACCCACGTCTCTGTGGACAGCTGGTTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAAT
GGCACCCCAGACATTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGC
CAGCACTAATTGGTATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATC
TTTATCATCCTGGCACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAG
GATCCTCGAGGCG

NOV42c, 317863291 Protein Sequence	SEQ ID NO: 584	231 aa	MW at 25750.8kD

RGSTMYKFFRGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSF
LDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIM
HPRLCGQLVSVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAI
FIILAPLILILISYGYVGGTVLRILEA

NOV42d, 317863328 DNA Sequence	SEQ ID NO: 585	609 bp
	ORF Start: at 1	ORF Stop: end of sequence

CGCGGATCCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCTTGCTTTCAGCTCT
GGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGGACATGTGTTTCA
CCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCATCAGCTATGTGGGT
TGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGTCATGGC
ATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTCTCTGTG
GACAGCTGGTTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGACATTG
ATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGGTAT
GGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGGCAC
CACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCCTCGAGGCG

NOV42d, 317863328 Protein Sequence	SEQ ID NO: 586	203 aa	MW at 22437.8kD

RGSLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSFLDMCFTTGSIPQMLYNLWGPDKTISYVG
CAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMHPRLCGQLVSVAWLSGFGNSLIMAPQTL
MLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFIILAPLILILISYGYVGGTVLRILEA

NOV42e, 317863350 DNA Sequence	SEQ ID NO: 587	433 bp
	ORF Start: at 1	ORF Stop: at 433

CGCGGATCCGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGT
CATGGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTC
TCTGTGGACAGCTGGTTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAG
ACATTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAAT
TGGTATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCC
TGGCACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCCTCGAG
GCGAAGGGCGAATTCCAGCACACTG

NOV42e, 317863350 Protein Sequence	SEQ ID NO: 588	144 aa	MW at 15783.9kD

RGSAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMHPRLCGQLVSVAWLSGFGNSLIMAPQ
TLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFIILAPLILILISYGYVGGTVLRILE
AKGEFQHT

NOV42f, 271624076 DNA Sequence	SEQ ID NO: 589	953 bp
	ORF Start: at 1	ORF Stop: TAG at 934

GAAAACGATAATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGA
GCTGATCGTCTTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCT
TGCTTTCAGCTCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTG
GACATGTGTTTCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCAT
CAGCTATGTGGGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCC
TGGCTGTCATGGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCAC

CCACGTCTCTGTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGC
ACCCCAGACATTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAG
CACTAATTGGTATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTT
ATCATCCTGGCACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGAT
CAAGTCAGCTGCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCT
TCTATGGTACAATCATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTT
CTTACCCTTTTCTACACAATCGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGA
TGTTAAAGAGGCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATA**TAG**GAATTCGGAAGGGCGA
A

NOV42f, 271624076 Protein Sequence	SEQ ID NO: 590	311 aa	MW at 34523.7kD

ENDNTSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSFL
DMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMH
PRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIF
IILAPLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKF
LTLFYTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42g, CG55358-01 DNA Sequence	SEQ ID NO: 591	945 bp
	ORF Start: ATG at 1	ORF Stop: TAG at 937

ATGGAAAACGATAATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCT
AGAGCTGATCGTCTTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTG
TCTTGCTTTCAGCTCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTC
CTGGACATGTGTTTCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGAC
CATCAGCTATGTGGGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCC
TCCTGGCTGTCATGGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATG
CACCCACGTCTCTGTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAAT
GGCACCCCAGACATTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGC
CAGCACTAATTGGTATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATC
TTTATCATCCTGGCACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAG
GATCAAGTCAGCTGCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTC
TCTTCTATGGTACAATCATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAG
TTTCTTACCCTTTTCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAA
AGATGTTAAAGAGGCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATA**TAG**TAAGGG

NOV42g, CG55358-01 Protein Sequence	SEQ ID NO: 592	312 aa	MW at 34654.9kD

MENDNTSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSF
LDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIM
HPRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAI
FIILAPLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGK
FLTLFYTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42h, CG55358-02 DNA Sequence	SEQ ID NO: 593	947 bp
	ORF Start: ATG at 3	ORF Stop: TAG at 939

GT**ATG**GAAAACGATAATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCAC
CTAGAGCTGATCGTCTTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCAT
TGTCTTGCTTTCAGCTCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCAT
TCCTGGACATGTGTTTCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAG
ACCATCAGCTATGTGGGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGT
CCTCCTGGCTGTCATGGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCA
TGCACCCACGTCTCTGTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATA
ATGGCACCCCAGACATTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGAT
GCCAGCACTAATTGGTATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAA
TCTTTATCATCCTGGCACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTT
AGGATCAAGTCAGCTGCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTC
TCTCTTCTATGGTACAATCATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCA
AGTTTCTTACCCTTTTCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAAC

AAAGATGTTAAAGAGGCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATATAGTAAGGG

NOV42h, CG55358-02 Protein Sequence	SEQ ID NO: 594	312 aa	MW at 34654.9kD

MENDNTSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSF
LDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIM
HPRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAI
FIILAPLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGK
FLTLFYTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42i, CG55358-05 DNA Sequence	SEQ ID NO: 595	63 bp
	ORF Start: ATG at 1	ORF Stop: end of sequence

ATGGAAAACGATAATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCC

NOV42i, CG55358-05 Protein Sequence	SEQ ID NO: 596	21 aa	MW at 2362.5kD

MENDNTSSFEGFILVGFSDRP

NOV42j, CG55358-06 DNA Sequence	SEQ ID NO: 597	60 bp
	ORF Start: at 1	ORF Stop: end of sequence

ACATTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCA

NOV42j, CG55358-06 Protein Sequence	SEQ ID NO: 598	20 aa	MW at 2411.9kD

TLMLPRCGHRRVDHFLCEMP

NOV42k, CG55358-07 DNA Sequence	SEQ ID NO: 599	953 bp
	ORF Start: at 1	ORF Stop: TAG at 934

GAAAACGATAATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGA
GCTGATCGTCTTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCT
TGCTTTCAGCTCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTG
GACATGTGTTTCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCAT
CAGCTATGTGGGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCC
TGGCTGTCATGGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCAC
CCACGTCTCTGTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGC
ACCCCAGACATTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAG
CACTAATTGGTATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTT
ATCATCCTGGCACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGAT
CAAGTCAGCTGCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCT
TCTATGGTACAATCATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTT
CTTACCCTTTTCTACACAATCGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGA
TGTTAAAGAGGCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATATAGGAATTCGGAAGGGCGA
A

NOV42k, CG55358-07 Protein Sequence	SEQ ID NO: 600	311 aa	MW at 34523.7kD

ENDNTSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSFL
DMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMH
PRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIF
IILAPLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKF
LTLFYTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42l, SNP13375122 of CG55358-03, DNA Sequence	SEQ ID NO: 601	932 bp
	ORF Start: at 3	ORF Stop: TAG at 924
	SNP Pos: 64	SNP Change: T to C

ATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGAGCTGACCGTC
TTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCTTGCTTTCAGC

TCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGGACATGTGTT
TCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCATCAGCTATGTG
GGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGTCAT
GGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTCTCT
GTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGACA
TTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGG
TATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGG
CACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCAAGTCAGCT
GCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCTTCTATGGTAC
AATTATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTTCTTACCCTTT
TCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGATGTTAAAGAG
GCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATA**TAG**TAAGGG

NOV42l, SNP13375122 of CG55358-03, Protein Sequence	SEQ ID NO: 602	307 aa	MW at 34039.2kD
	SNP Pos: 21		SNP Change: Ile to Thr

TSSFEGFILVGFSDRPHLELTVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSFLDMCF
TTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMHPRLC
GQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFIILA
PLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLF
YTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42m, SNP13375123 of CG55358-03, DNA Sequence	SEQ ID NO: 603	932 bp
	ORF Start: at 3	ORF Stop: TAG at 924
	SNP Pos: 124	SNP Change: T to C

ATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGAGCTGATCGTC
TTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGCCTTGCTTTCAGC
TCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGGACATGTGTT
TCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCATCAGCTATGTG
GGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGTCAT
GGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTCTCT
GTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGACA
TTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGG
TATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGG
CACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCAAGTCAGCT
GCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCTTCTATGGTAC
AATTATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTTCTTACCCTTT
TCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGATGTTAAAGAG
GCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATA**TAG**TAAGGG

NOV42m, SNP13375123 of CG55358-03, Protein Sequence	SEQ ID NO: 604	307 aa	MW at 34023.2kD
	SNP Pos: 41		SNP Change: Val to Ala

TSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIALLSALDSRLHTPMYFFLANLSFLDMCF
TTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMHPRLC
GQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFIILA
PLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLF
YTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42n, SNP13382494 of CG55358-03, DNA Sequence	SEQ ID NO: 605	932 bp
	ORF Start: at 3	ORF Stop: TAG at 924
	SNP Pos: 254	SNP Change: T to C

ATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGAGCTGATCGTC
TTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCTTGCTTTCAGC
TCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGGACATGTGTT
TCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGACAAGACCATCAGCTATGTG
GGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGTCAT
GGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTCTCT
GTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGACA

TTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGG
TATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGG
CACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCAAGTCAGCT
GCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCTTCTATGGTAC
AATTATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTTCTTACCCTTT
TCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGATGTTAAAGAG
GCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATA**TAG**TAAGGG

NOV42n, SNP13382494 of CG55358-03, Protein Sequence	SEQ ID NO: 606	307 aa	MW at 34051.3kD
	SNP Pos: 84		SNP Change: Asp to Asp

TSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSFLDMCF
TTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMHPRLC
GQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFIILA
PLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLF
YTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42o, SNP13375124 of CG55358-03, DNA Sequence	SEQ ID NO: 607	932 bp
	ORF Start: at 3	ORF Stop: TAG at 924
	SNP Pos: 283	SNP Change: T to C

ATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGAGCTGATCGTC
TTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCTTGCTTTCAGC
TCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGGACATGTGTT
TCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCATCAGCTATGTG
GGTTGTGCCACCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGTCAT
GGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTCTCT
GTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGACA
TTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGG
TATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGG
CACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCAAGTCAGCT
GCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCTTCTATGGTAC
AATTATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTTCTTACCCTTT
TCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGATGTTAAAGAG
GCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATA**TAG**TAAGGG

NOV42o, SNP13375124 of CG55358-03, Protein Sequence	SEQ ID NO: 608	307 aa	MW at 34039.2kD
	SNP Pos: 94		SNP Change: Ile to Thr

TSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSFLDMCF
TTGSIPQMLYNLWGPDKTISYVGCATQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMHPRLC
GQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFIILA
PLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLF
YTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42p, SNP13375125 of CG55358-03, DNA Sequence	SEQ ID NO: 609	932 bp
	ORF Start: at 3	ORF Stop: TAG at 924
	SNP Pos: 336	SNP Change: G to A

ATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGAGCTGATCGTC
TTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCTTGCTTTCAGC
TCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGGACATGTGTT
TCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCATCAGCTATGTG
GGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTATCAT
GGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTCTCT
GTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGACA
TTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGG
TATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGG
CACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCAAGTCAGCT
GCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCTTCTATGGTAC
AATTATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTTCTTACCCTTT

TCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGATGTTAAAGAG
GCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATATAGTAAGGG

NOV42p, SNP13375125 of CG55358-03, Protein Sequence	SEQ ID NO: 610	307 aa	MW at 34065.3kD
	SNP Pos: 112		SNP Change: Val to Ile

TSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSFLDMCF
TTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAIMAYDRYAAVCKPLHYTIIMHPRLC
GQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFIILA
PLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLF
YTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42q, SNP13376426 of CG55358-03, DNA Sequence	SEQ ID NO: 611	932 bp
	ORF Start: at 3	ORF Stop: TAG at 924
	SNP Pos: 415	SNP Change: A to G

ATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGAGCTGATCGTC
TTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCTTGCTTTCAGC
TCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGGACATGTGTT
TCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCATCAGCTATGTG
GGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGTCAT
GGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTCTCT
GTGGACGGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGACA
TTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGG
TATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGG
CACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCAAGTCAGCT
GCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCTTCTATGGTAC
AATTATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTTCTTACCCTTT
TCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGATGTTAAAGAG
GCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATATAGTAAGGG

NOV42q, SNP13376426 of CG55358-03, Protein Sequence	SEQ ID NO: 612	307 aa	MW at 34079.3kD
	SNP Pos: 138		SNP Change: Gln to Arg

TSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSFLDMCF
TTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMHPRLC
GRLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFIILA
PLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLF
YTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42r, SNP13376427 of CG55358-03, DNA Sequence	SEQ ID NO: 613	932 bp
	ORF Start: at 3	ORF Stop: TAG at 924
	SNP Pos: 421	SNP Change: C to T

ATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGAGCTGATCGTC
TTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCTTGCTTTCAGC
TCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGGACATGTGTT
TCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCATCAGCTATGTG
GGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGTCAT
GGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTCTCT
GTGGACAGCTGGTTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGACA
TTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGG
TATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGG
CACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCAAGTCAGCT
GCTGGGCGAAAGAAAGCCTTCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCTTCTATGGTAC
AATTATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTTCTTACCCTTT
TCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGATGTTAAAGAG
GCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATATAGTAAGGG

5

NOV42r, SNP13376427 of CG55358-03, Protein Sequence	SEQ ID NO: 614	307 aa	MW at 34079.3kD
	SNP Pos: 140		SNP Change: Ala to Val

TSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSFLDMCF
TTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMHPRLC
GQLVSVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFIILA
PLILILISYGYVGGTVLRIKSAAGRKKAFNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLF
YTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

NOV42s, SNP13375126 of CG55358-03, DNA Sequence	SEQ ID NO: 615	932 bp
	ORF Start: at 3	ORF Stop: TAG at 924
	SNP Pos: 700	SNP Change: T to C

ATACAAGTTCTTTCGAAGGCTTCATCCTGGTGGGCTTCTCTGATCGTCCCCACCTAGAGCTGATCGTC
TTTGTGGTTGTCCTCATCTTTTATCTGCTGACTCTTCTTGGCAACATGACCATTGTCTTGCTTTCAGC
TCTGGATTCCCGGCTGCACACACCAATGTATTTCTTTTTGGCAAACCTCTCATTCCTGGACATGTGTT
TCACCACAGGTTCCATCCCTCAGATGCTCTACAACCTTTGGGGTCCAGATAAGACCATCAGCTATGTG
GGTTGTGCCATCCAGCTGTACTTTGTCCTGGCCCTGGGAGGGGTGGAGTGTGTCCTCCTGGCTGTCAT
GGCATATGACCGCTATGCTGCAGTCTGCAAACCCCTGCACTACACCATCATCATGCACCCACGTCTCT
GTGGACAGCTGGCTTCAGTGGCATGGCTGAGTGGCTTTGGCAATTCTCTCATAATGGCACCCCAGACA
TTGATGCTACCCCGCTGTGGGCACAGACGAGTTGACCACTTTCTCTGTGAGATGCCAGCACTAATTGG
TATGGCCTGTGTAGACACCATGATGCTTGAGGCACTGGCTTTTGCCCTGGCAATCTTTATCATCCTGG
CACCACTCATCCTCATTCTCATTTCTTATGGTTACGTTGGAGGAACAGTGCTTAGGATCAAGTCAGCT
GCTGGGCGAAAGAAAGCCTCCAACACTTGCAGCTCGCATCTAATTGTTGTCTCTCTCTTCTATGGTAC
AATTATATACATGTACCTCCAGCCAGCAAATACTTATTCCCAGGACCAGGGCAAGTTTCTTACCCTTT
TCTACACAATTGTCACTCCCAGTGTTAACCCCCTGATCTATACACTAAGAAACAAAGATGTTAAAGAG
GCCATGAAGAAGGTGCTAGGGAAGGGGAGTGCAGAAATATAGTAAGGG

NOV42s, SNP13375126 of CG55358-03, Protein Sequence	SEQ ID NO: 616	307 aa	MW at 33991.2kD
	SNP Pos: 233		SNP Change: Phe to Ser

TSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMYFFLANLSFLDMCF
TTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAYDRYAAVCKPLHYTIIMHPRLC
GQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCEMPALIGMACVDTMMLEALAFALAIFIILA
PLILILISYGYVGGTVLRIKSAAGRKKASNTCSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLF
YTIVTPSVNPLIYTLRNKDVKEAMKKVLGKGSAEI

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 42B.

Table 42B. Comparison of the NOV42 protein sequences.

```
NOV42a  MENDNTSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMY
NOV42b  -----TSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMY
NOV42c  RGSTMYKFFRGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMY
NOV42d  ----------------------------RGSLIFYLLTLLGNMTIVLLSALDSRLHTPMY
NOV42e  ------------------------------------------------------------
NOV42f  -ENDNTSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMY
NOV42g  MENDNTSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMY
NOV42h  MENDNTSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMY
NOV42i  ------------------------------------------------------------
NOV42j  ------------------------------------------------------------
NOV42k  -ENDNTSSFEGFILVGFSDRPHLELIVFVVVLIFYLLTLLGNMTIVLLSALDSRLHTPMY

NOV42a  FFLANLSFLDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAY
NOV42b  FFLANLSFLDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAY
```

```
NOV42c  FFLANLSFLDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAY
NOV42d  FFLANLSFLDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAY
NOV42e  ---------------------------------RGSAIQLYFVLALGGVECVLLAVMAY
NOV42f  FFLANLSFLDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAY
NOV42g  FFLANLSFLDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAY
NOV42h  FFLANLSFLDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAY
NOV42i  ------------------------------------------------------------
NOV42j  ------------------------------------------------------------
NOV42k  FFLANLSFLDMCFTTGSIPQMLYNLWGPDKTISYVGCAIQLYFVLALGGVECVLLAVMAY

NOV42a  DRYAAVCKPLHYTIIMHPRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCE
NOV42b  DRYAAVCKPLHYTIIMHPRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCE
NOV42c  DRYAAVCKPLHYTIIMHPRLCGQLVSVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCE
NOV42d  DRYAAVCKPLHYTIIMHPRLCGQLVSVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCE
NOV42e  DRYAAVCKPLHYTIIMHPRLCGQLVSVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCE
NOV42f  DRYAAVCKPLHYTIIMHPRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCE
NOV42g  DRYAAVCKPLHYTIIMHPRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCE
NOV42h  DRYAAVCKPLHYTIIMHPRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCE
NOV42i  -------------------------------MENDNTSSFEGFILVGFSDRP-------
NOV42j  ----------------------------------------TLMLPRCGHRRVDHFLCE
NOV42k  DRYAAVCKPLHYTIIMHPRLCGQLASVAWLSGFGNSLIMAPQTLMLPRCGHRRVDHFLCE

NOV42a  MPALIGMACVDTMMLEALAFALAIFIVLAPLILILISYGYVGGTVLRIKSAAGRKKAFNT
NOV42b  MPALIGMACVDTMMLEALAFALAIFIILAPLILILISYGYVGGTVLRIKSAAGRKKAFNT
NOV42c  MPALIGMACVDTMMLEALAFALAIFIILAPLILILISYGYVGGTVLRILEA---------
NOV42d  MPALIGMACVDTMMLEALAFALAIFIILAPLILILISYGYVGGTVLRILEA---------
NOV42e  MPALIGMACVDTMMLEALAFALAIFIILAPLILILISYGYVGGTVLRILEAKGEFQHT--
NOV42f  MPALIGMACVDTMMLEALAFALAIFIILAPLILILISYGYVGGTVLRIKSAAGRKKAFNT
NOV42g  MPALIGMACVDTMMLEALAFALAIFIILAPLILILISYGYVGGTVLRIKSAAGRKKAFNT
NOV42h  MPALIGMACVDTMMLEALAFALAIFIILAPLILILISYGYVGGTVLRIKSAAGRKKAFNT
NOV42i  ------------------------------------------------------------
NOV42j  MP----------------------------------------------------------
NOV42k  MPALIGMACVDTMMLEALAFALAIFIILAPLILILISYGYVGGTVLRIKSAAGRKKAFNT

NOV42a  CSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLFYTIVTPSVNPLIYTLRNKDVKEA
NOV42b  CSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLFYTIVTPSVNPLIYTLRNKDVKEA
NOV42c  ------------------------------------------------------------
NOV42d  ------------------------------------------------------------
NOV42e  ------------------------------------------------------------
NOV42f  CSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLFYTIVTPSVNPLIYTLRNKDVKEA
NOV42g  CSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLFYTIVTPSVNPLIYTLRNKDVKEA
NOV42h  CSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLFYTIVTPSVNPLIYTLRNKDVKEA
NOV42i  ------------------------------------------------------------
NOV42j  ------------------------------------------------------------
NOV42k  CSSHLIVVSLFYGTIIYMYLQPANTYSQDQGKFLTLFYTIVTPSVNPLIYTLRNKDVKEA

NOV42a  MKKVLGKGSAEI
NOV42b  MKKVLGKGSAEI
NOV42c  ------------
NOV42d  ------------
NOV42e  ------------
NOV42f  MKKVLGKGSAEI
NOV42g  MKKVLGKGSAEI
NOV42h  MKKVLGKGSAEI
```

```
NOV42i  ------------
NOV42j  ------------
NOV42k  MKKVLGKGSAEI

NOV42a  (SEQ ID NO:  580)
NOV42b  (SEQ ID NO:  582)
NOV42c  (SEQ ID NO:  584)
NOV42d  (SEQ ID NO:  586)
NOV42e  (SEQ ID NO:  588)
NOV42f  (SEQ ID NO:  590)
NOV42g  (SEQ ID NO:  592)
NOV42h  (SEQ ID NO:  594)
NOV42i  (SEQ ID NO:  596)
NOV42j  (SEQ ID NO:  598)
NOV42k  (SEQ ID NO:  600)
```

Further analysis of the NOV42a protein yielded the following properties shown in Table 42C.

Table 42C. Protein Sequence Properties NOV42a

SignalP analysis:	Cleavage site between residues 42 and 43

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 10;  pos.chg 0;  neg.chg 3
      H-region:  length 8;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -2.80
      possible cleavage site: between 41 and 42

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   5
      INTEGRAL    Likelihood =-13.21   Transmembrane   25 -  41
      INTEGRAL    Likelihood = -7.11   Transmembrane  103 - 119
      INTEGRAL    Likelihood = -0.27   Transmembrane  183 - 199
      INTEGRAL    Likelihood =-14.97   Transmembrane  200 - 216
      INTEGRAL    Likelihood = -1.12   Transmembrane  244 - 260
      PERIPHERAL  Likelihood =  1.80 (at 273)
      ALOM score: -14.97  (number of TMSs: 5)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 32
      Charge difference:  2.0   C( 0.5) - N(-1.5)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b
```

```
MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  6.56
      Hyd Moment(95):  3.47     G content:       0
      D/E content:     2        S/T content:     0
      Score: -7.29

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   6.7%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues
```

```
--------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       22.2 %: vacuolar
       11.1 %: vesicles of secretory system
       11.1 %: Golgi

>> prediction for CG55358-04 is end (k=9)
```

A search of the NOV42a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 42D.

Table 42D. Geneseq Results for NOV42a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE10685	G-protein coupled receptor 4a (GPCR4a) - Unidentified, 312 aa. [WO200164879-A2, 07-SEP-2001]	1..312 1..312	311/312 (99%) 312/312 (99%)	e-180
AAG72062	Human olfactory receptor polypeptide, SEQ ID NO: 1743 - Homo sapiens, 334 aa. [WO200127158-A2, 19-APR-2001]	1..312 3..314	310/312 (99%) 312/312 (99%)	e-179
AAE10686	G-protein coupled receptor 4b (GPCR4b) - Unidentified, 307 aa. [WO200164879-A2, 07-SEP-2001]	6..312 1..307	306/307 (99%) 307/307 (99%)	e-177
ABP95745	Human GPCR polypeptide SEQ ID NO 300 - Homo sapiens, 270 aa. [WO200216548-A2, 28-FEB-2002]	43..312 1..270	269/270 (99%) 270/270 (99%)	e-155
AAG71998	Human olfactory receptor polypeptide, SEQ ID NO: 1679 - Homo sapiens, 270 aa. [WO200127158-A2, 19-APR-2001]	43..312 1..270	269/270 (99%) 270/270 (99%)	e-155

In a BLAST search of public sequence databases, the NOV42a protein was found to have homology to the proteins shown in the BLASTP data in Table 42E.

Table 42E. Public BLASTP Results for NOV42a

Protein Accession Number	Protein/Organism/Length	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC88325	Sequence 11 from Patent WO0164879 - Homo sapiens (Human), 312 aa.	1..312 1..312	311/312 (99%) 312/312 (99%)	e-179
Q8NHA6	Seven transmembrane helix receptor - Homo sapiens (Human), 318 aa.	12..312 18..318	300/301 (99%) 301/301 (99%)	e-173
Q8VFH0	Olfactory receptor MOR256-12 - Mus musculus (Mouse), 317 aa.	1..312 6..317	295/312 (94%) 304/312 (96%)	e-172
Q8VG14	Olfactory receptor MOR256-8 - Mus musculus (Mouse), 308 aa.	1..307 1..307	212/307 (69%) 251/307 (81%)	e-125
Q8VFG9	Olfactory receptor MOR256-13 - Mus musculus (Mouse), 314 aa.	5..305 11..311	205/301 (68%) 247/301 (81%)	e-120

PFam analysis indicates that the NOV42a protein contains the domains shown in the Table 42F.

Table 42F. Domain Analysis of NOV42a

Pfam Domain	NOV42a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1	41..290	57/276 (21%) 177/276 (64%)	3.2e-41

Example 43.

The NOV43 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 43A.

Table 43A. NOV43 Sequence Analysis

NOV43a, CG55604-04 DNA Sequence	SEQ ID NO: 617	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43a, CG55604-04 Protein Sequence	SEQ ID NO: 618	314 aa	MW at 35728.1kD

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43b, CG55604-01 DNA Sequence	SEQ ID NO: 619	840 bp
	ORF Start: ATG at 42	ORF Stop: TGA at 810

TCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATCTCTCC
TTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAA
AACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAGTGTG
CGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTACCATC
ATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTTTAGT
AGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGTGAAC
CTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGC
GTGGTAATCCTCCTGGCCCCTGTCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTAT
CCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTG
TCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTGGAT
AAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGAGGAA
CAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGACCTC
TGAGTCTGACTTTTAGAGCTATGG

NOV43b, CG55604-01 Protein Sequence	SEQ ID NO: 620	256 aa	MW at 29023.1kD

MYFFLRNLSFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAV
CKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYS
TEMAIFSMGVVILLAPVSLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRP
NSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43c, CG55604-02 DNA Sequence	SEQ ID NO: 621	994 bp
	ORF Start: ATG at 31	ORF Stop: TGA at 973

TGCCAAACAGGTAAACAGGCAAAAATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTAT
CTTCCTGGGTCTTTCACAGGACTTGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATC
TGCTGACCGTGCTTGGAAACCAGCTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCC
ATGTATTTTTTTCTTAGAAATATCTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGT

GTTGGTTCACTTCTTGGTAAAGAGGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCT
TTCTTCTGGTTGGGTGTACAGAGTGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTC
TGCAAGCCCCTGTACTACTCTACCATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTG
GGCCAGTGGGGCACTAGTGTCTTTAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGA
ATATAATCAATCACTACTTTTGTGAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGC
ACAGAAATGGCCATCTTTTCAATGGGCGTGGTAATCCTCCTGGCCCCTGTCTCCCTGATTCTTGGTTC
TTATTGGAATATTATCTCCACTGTTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCA
CCTGTGGCTCCCATCTTATTGTTGTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCA
AACTCCAAGACTACAAAAGAACTGGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTT
GAACCCCATAATTTATAGCTTGAGGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAA
AGTGCTTCTCTCATAGGCAGTGACCTCTGAGTCTGACTTTTA

NOV43c, CG55604-02 Protein Sequence	SEQ ID NO: 622	314 aa	MW at 35714.1kD

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPVSLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43d, CG55604-03 DNA Sequence	SEQ ID NO: 623	960 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 943

ATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACTTGCAGAC
CCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAGCTCATCA
TCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATATCTCCTTT
GCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAAAAC
CATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAGTGTGCGC
TGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTACCATCATG
ACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTTTAGTAGA
TACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGTGAACCTC
CTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGCGTG
GTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCCA
GATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTCC
TCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTGGATAAA
ATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGAGGAACAA
AGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGACCTCTGA
GTCTGACT

NOV43d, CG55604-03 Protein Sequence	SEQ ID NO: 624	314 aa	MW at 35728.1kD

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43e, CG55604-05 DNA Sequence	SEQ ID NO: 625	995 bp
	ORF Start: ATG at 31	ORF Stop: TGA at 973

TGCCAAACAGGTAAACAGGCAAAAATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTAT
CTTCCTGGGTCTTTCACAGGACTTGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATC
TGCTGACCGTGCTTGGAAACCAGCTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCC
ATGTATTTTTTTCTTAGAAATCTCTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGT
GTTGGTTCACTTCTTGGTAAAGAGGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCT
TTCTTCTGGTTGGGTGTACAGAGTGTGCGCTGCTGGCCGTGATGTCCTATGACCGGTATGTGGCTGTC
TGCAAGCCCCTGTACTACTCTACCATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTG
GGCCAGTGGGGCACTAGTGTCTTTAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGA
ATATAATCAATCACTACTTTTGTGAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGC
ACAGAAATGGCCATCTTTTCAATGGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTC

TTATTGGAATATTATCTCCACTGTTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCA
CCTGTGGCTCCCATCTTATTGTTGTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCA
AACTCCAAGACTACAAAAGAACTGGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTT
GAACCCCATAATTTATAGCTTGAGGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAA
AGTGCTTCTCTCATAGGCAGTGACCTCTGAGTCTGACTTTTAA

NOV43e, CG55604-05 Protein Sequence	SEQ ID NO: 626	314 aa	MW at 35728.1kD

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNLSF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43f, CG55604-06 DNA Sequence	SEQ ID NO: 627	946 bp
	ORF Start: ATG at 2	ORF Stop: TGA at 944

TATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACTTGCAGA
CCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAGCTCATC
ATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATATCTCCTT
TGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAAAA
CCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAGTGTGCG
CTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTACCATCAT
GACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTCGGCCAGTGGGGCACTAGTGTCTTTAGTAG
ATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGTGAACCT
CCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGCGT
GGTAATCCTCCTGGCCCCTGTCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCC
AGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTC
CTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTGGATAA
AATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGAGGAACA
AAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA

NOV43f, CG55604-06 Protein Sequence	SEQ ID NO: 628	314 aa	MW at 35614.9kD

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSSASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPVSLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43g, CG55604-07 DNA Sequence	SEQ ID NO: 629	962 bp
	ORF Start: ATG at 2	ORF Stop: TGA at 944

TATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACTTGCAGA
CCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAGCTCATC
ATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATATCTCCTT
TGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAAAA
CCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAGTGTGCG
CTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTACCATCAT
GACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTTTAGTAG
ATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGTGAACCT
CCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGCGT
GGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCC
AGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTC
CTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTGGATAA
AATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGAGGAACA
AAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGACCTCTG
AGTCTGACTA

NOV43g, CG55604-07 Protein Sequence	SEQ ID NO: 630	314 aa	MW at 35728.1kD

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43h, SNP13019742 of CG55604-04, DNA Sequence	SEQ ID NO: 631	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 203	SNP Change: A to C

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATCT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43h, SNP13019742 of CG55604-04, Protein Sequence	SEQ ID NO: 632	314 aa	MW at 35728.1kD
	SNP Pos: 66		SNP Change: Ile to Leu

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNLSF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43i, SNP13373776 of CG55604-04, DNA Sequence	SEQ ID NO: 633	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 355	SNP Change: A to C

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCCGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43i, SNP13373776 of CG55604-04, Protein Sequence	SEQ ID NO: 634	314 aa	MW at 35728.1kD
	SNP Pos: 116		SNP Change: Ala to Ala

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43j, SNP13373777 of CG55604-04, DNA Sequence	SEQ ID NO: 635	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 372	SNP Change: G to A

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCAGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43j, SNP13373777 of CG55604-04, Protein Sequence	SEQ ID NO: 636	314 aa	MW at 35700.0kD
	SNP Pos: 122		SNP Change: Arg to Gln

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDQYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43k, SNP13373778 of CG55604-04, DNA Sequence	SEQ ID NO: 637	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 422	SNP Change: C to T

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAATGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43k, SNP13373778 of CG55604-04, Protein Sequence	SEQ ID NO: 638	314 aa	MW at 35758.1kD
	SNP Pos: 139		SNP Change: Arg to Trp

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQWVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43l, SNP13373838 of CG55604-04, DNA Sequence	SEQ ID NO: 639	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 453	SNP Change: G to C

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTCGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43l, SNP13373838 of CG55604-04, Protein Sequence	SEQ ID NO: 640	314 aa	MW at 35628.9kD
	SNP Pos: 149		SNP Change: Trp to Ser

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSSASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43m, SNP13373837 of CG55604-04, DNA Sequence	SEQ ID NO: 641	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 461	SNP Change: G to A

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTAGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43m, SNP13373837 of CG55604-04, Protein Sequence	SEQ ID NO: 642	314 aa	MW at 35827.2kD
	SNP Pos: 152		SNP Change: Gly to Arg

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASRALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43n, SNP13373836 of CG55604-04, DNA Sequence	SEQ ID NO: 643	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 587	SNP Change: A to G

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACGGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43n, SNP13373836 of CG55604-04, Protein Sequence	SEQ ID NO: 644	314 aa	MW at 35698.1kD
	SNP Pos: 194		SNP Change: Ser to Gly

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYGTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43o, SNP13373780 of CG55604-04, DNA Sequence	SEQ ID NO: 645	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 615	SNP Change: G to A

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGACGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43o, SNP13373780 of CG55604-04, Protein Sequence	SEQ ID NO: 646	314 aa	MW at 35786.1kD
	SNP Pos: 203		SNP Change: Gly to Asp

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMDV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43p, SNP13373781 of CG55604-04, DNA Sequence	SEQ ID NO: 647	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 638	SNP Change: A to G

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTGTCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43p, SNP13373781 of CG55604-04, Protein Sequence	SEQ ID NO: 648	314 aa	MW at 35714.1kD
	SNP Pos: 211		SNP Change: Ile to Val

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPVSLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43q, SNP13373782 of CG55604-04, DNA Sequence	SEQ ID NO: 649	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 679	SNP Change: T to C

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACCG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

5

NOV43q, SNP13373782 of CG55604-04, Protein Sequence	SEQ ID NO: 650	314 aa	MW at 35728.1kD
	SNP Pos: 224		SNP Change: Thr to Thr

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43r, SNP13373783 of CG55604-04, DNA Sequence	SEQ ID NO: 651	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 844	SNP Change: A to G

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACGGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43r, SNP13373783 of CG55604-04, Protein Sequence	SEQ ID NO: 652	314 aa	MW at 35728.1kD
	SNP Pos: 279		SNP Change: Thr to Thr

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43s, SNP13373833 of CG55604-04, DNA Sequence	SEQ ID NO: 653	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 850	SNP Change: G to A

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTAACTCCAATGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

5

NOV43s, SNP13373833 of CG55604-04, Protein Sequence	SEQ ID NO: 654	314 aa	MW at 35728.1kD
	SNP Pos: 281		SNP Change: Val to Val

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43t, SNP13373784 of CG55604-04, DNA Sequence	SEQ ID NO: 655	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 857	SNP Change: A to G

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAGTGTTGAACCCCATAATTTATAGCTTGA
GGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

NOV43t, SNP13373784 of CG55604-04, Protein Sequence	SEQ ID NO: 656	314 aa	MW at 35696.0kD
	SNP Pos: 284		SNP Change: Met to Val

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPVLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ

NOV43u, SNP13373832 of CG55604-04, DNA Sequence	SEQ ID NO: 657	968 bp
	ORF Start: ATG at 8	ORF Stop: TGA at 950
	SNP Pos: 885	SNP Change: G to A

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACT
TGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAG
CTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATAT
CTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGA
GGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAG
TGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTAC
CATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTT
TAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGT
GAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAAT
GGGCGTGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTG
TTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTT
GTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACT
GGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGA
AGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGA
CCTCTGAGTCTGACTA

5

NOV43u, SNP13373832 of CG55604-04, Protein Sequence	SEQ ID NO: 658	314 aa	MW at 35700.1kD
	SNP Pos: 293		SNP Change: Arg to Lys

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISF
ADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIM
TQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGV
VILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDK
MISVFYTAVTPMLNPIIYSLKNKDVKGALRKLVGRKCFSHRQ

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 43B.

Table 43B. Comparison of the NOV43 protein sequences.

```
NOV43a  MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY
NOV43b  ----------------------------------------------------------MY
NOV43c  MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY
NOV43d  MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY
NOV43e  MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY
NOV43f  MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY
NOV43g  MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY

NOV43a  FFLRNISFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY
NOV43b  FFLRNLSFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY
NOV43c  FFLRNISFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY
NOV43d  FFLRNISFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY
NOV43e  FFLRNLSFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY
NOV43f  FFLRNISFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY
NOV43g  FFLRNISFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY

NOV43a  DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE
NOV43b  DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE
NOV43c  DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE
NOV43d  DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE
NOV43e  DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE
NOV43f  DRYVAVCKPLYYSTIMTQRVCLWLSFRSSASGALVSLVDTSFTFHLPYWGQNIINHYFCE
NOV43g  DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE

NOV43a  PPALLKLASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGRLKAFST
NOV43b  PPALLKLASIDTYSTEMAIFSMGVVILLAPVSLILGSYWNIISTVIQMQSGEGRLKAFST
NOV43c  PPALLKLASIDTYSTEMAIFSMGVVILLAPVSLILGSYWNIISTVIQMQSGEGRLKAFST
NOV43d  PPALLKLASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGRLKAFST
NOV43e  PPALLKLASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGRLKAFST
NOV43f  PPALLKLASIDTYSTEMAIFSMGVVILLAPVSLILGSYWNIISTVIQMQSGEGRLKAFST
NOV43g  PPALLKLASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGRLKAFST

NOV43a  CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA
NOV43b  CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA
NOV43c  CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA
NOV43d  CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA
NOV43e  CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA
NOV43f  CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA
NOV43g  CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA
```

```
NOV43a  LRKLVGRKCFSHRQ
NOV43b  LRKLVGRKCFSHRQ
NOV43c  LRKLVGRKCFSHRQ
NOV43d  LRKLVGRKCFSHRQ
NOV43e  LRKLVGRKCFSHRQ
NOV43f  LRKLVGRKCFSHRQ
NOV43g  LRKLVGRKCFSHRQ

NOV43a  (SEQ ID NO:  618)
NOV43b  (SEQ ID NO:  620)
NOV43c  (SEQ ID NO:  622)
NOV43d  (SEQ ID NO:  624)
NOV43e  (SEQ ID NO:  626)
NOV43f  (SEQ ID NO:  628)
NOV43g  (SEQ ID NO:  630)
```

Further analysis of the NOV43a protein yielded the following properties shown in Table 43C.

Table 43C. Protein Sequence Properties NOV43a

SignalP analysis:	Cleavage site between residues 42 and 43

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 1;  neg.chg 2
      H-region:  length 8;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   0.24
      possible cleavage site: between 41 and 42

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   5
      INTEGRAL    Likelihood =-13.32   Transmembrane   25 -  41
      INTEGRAL    Likelihood = -1.75   Transmembrane   71 -  87
      INTEGRAL    Likelihood = -8.86   Transmembrane  101 - 117
      INTEGRAL    Likelihood = -9.34   Transmembrane  199 - 215
      INTEGRAL    Likelihood = -1.65   Transmembrane  241 - 257
      PERIPHERAL  Likelihood =  1.59 (at 273)
      ALOM score: -13.32  (number of TMSs: 5)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 32
      Charge difference:  1.5   C( 1.5) - N( 0.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b
```

```
MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  8.77
      Hyd Moment(95):  7.26     G content:       1
      D/E content:     2        S/T content:     0
      Score: -7.02

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   8.0%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:
      Leucine zipper pattern (PS00029):  *** found ***
         LFLIIYLLTVLGNQLIIILIFL at 30
 none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
```

```
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       44.4 %: endoplasmic reticulum
       22.2 %: vacuolar
       11.1 %: Golgi
       11.1 %: vesicles of secretory system
       11.1 %: mitochondrial

>> prediction for CG55604-04 is end (k=9)
```

A search of the NOV43a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 43D.

Table 43D. Geneseq Results for NOV43a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV43a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAG66338	Human NOV 18 protein sequence - Homo sapiens, 314 aa. [WO200155179-A2, 02-AUG-2001]	1..314 1..314	314/314 (100%) 314/314 (100%)	0.0
AAG66331	Human NOV 11 protein sequence - Homo sapiens, 314 aa. [WO200155179-A2, 02-AUG-2001]	1..314 1..314	314/314 (100%) 314/314 (100%)	0.0
AAG66337	Human NOV 17 protein sequence - Homo sapiens, 314 aa. [WO200155179-A2, 02-AUG-2001]	1..314 1..314	313/314 (99%) 314/314 (99%)	0.0
AAG66333	Human NOV 13 protein sequence - Homo sapiens, 314 aa. [WO200155179-A2, 02-AUG-2001]	1..314 1..314	313/314 (99%) 314/314 (99%)	0.0
AAG66332	Human NOV 12 protein sequence - Homo sapiens, 314 aa. [WO200155179-A2, 02-AUG-2001]	1..314 1..314	313/314 (99%) 314/314 (99%)	0.0

In a BLAST search of public sequence databases, the NOV43a protein was found to have homology to the proteins shown in the BLASTP data in Table 43E.

Table 43E. Public BLASTP Results for NOV43a				
Protein Accession Number	**Protein/Organism/Length**	**NOV43a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
CAC69632	Sequence 21 from Patent WO0155179 - Homo sapiens (Human), 314 aa.	1..314 1..314	314/314 (100%) 314/314 (100%)	e-180
CAC69634	Sequence 25 from Patent WO0155179 - Homo sapiens (Human), 314 aa.	1..314 1..314	313/314 (99%) 314/314 (99%)	e-180
CAC69637	Sequence 31 from Patent WO0155179 - Homo sapiens (Human), 314 aa.	1..314 1..314	313/314 (99%) 314/314 (99%)	e-180
Q8NGH3	Seven transmembrane helix receptor - Homo sapiens (Human), 314 aa.	1..314 1..314	312/314 (99%) 314/314 (99%)	e-180
Q9EP67	B5 olfactory receptor - Mus musculus (Mouse), 307 aa.	1..304 1..304	253/304 (83%) 275/304 (90%)	e-144

PFam analysis indicates that the NOV43a protein contains the domains shown in the Table 43F.

Table 43F. Domain Analysis of NOV43a			
Pfam Domain	**NOV43a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	41..290	58/277 (21%) 179/277 (65%)	8.4e-36

Example 44.

The NOV44 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 44A.

Table 44A. NOV44 Sequence Analysis

NOV44a, CG55752-07 DNA Sequence	SEQ ID NO: 659	1998 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1969

ATGGGCATTTATGGTTCAGTACCTTATCTCCTGGCCCACAAACTGGGCAGAACTATAGGTATTTTCTG
GCTGAATGCCTCGGAAACACTGGTGGAGATCAATACAGAGCCTGCAGTAGAGTACACACTGACCCAGA
TGGGCCCAGTTGCTGCTAAACAAAAGGTCAGATCTCGCACTCATGTGCACTGGATGTCAGAGAGTGGC
ATCATTGATGTTTTTCTGCTGACAGGACCTACACCTTCTGATGTCTTCAAACAGTACTCACACCTTAC
AGGTACACAAGCCATGCCCCCTCTTTTCTCTTTGGGATACCACCAGTGCCGCTGGAACTATGAAGATG
AGCAGGATGTAAAAGCAGTGGATGCAGGGTTTGATGAGCATGACATTCCTTATGATGCCATGTGGCTG
GACATAGAGCACACTGAGGGCAAGAGGTACTTCACCTGGGACAAAAACAGATTCCCAAACCCCAAGAG
GATGCAAGAGCTGCTCAGGAGCAAAAAGCGTAAGCTTGTGGTCATCAGTGATCCCCACATCAAGATTG
ATCCTGACTACTCAGTATATGTGAAGGCCAAAGATCAGGGCTTCTTTGTGAAGAATCAGGAAGGGGAA
GACTTTGAAGGGGTGTGTTGGCCAGGTCTCTCCTCTTACCTGGATTTCACCAATCCCAAGGTCAGAGA
GTGGTATTCAAGTCTTTTTGCTTTCCCTGTTTATCAGGGATCTACGGACATCCTCTTCCTTTGGAATG
ACATGAATGAGCCTTCTGTCTTTAGAGGGCCAGAGCAAACCATGCAGAAGAATGCCATTCATCATGGC
AATTGGGAGCACAGAGAGCTCCACAACATCTACGGTTTTTATCATCAAATGGCTACTGCAGAAGGACT
GATAAAACGATCTAAAGGGAAGGAGAGACCCTTTGTTCTTACACGTTCTTTCTTTGCTGGATCACAAA
AGTATGGTGGTGCCGTGTGGACAGGCGACAACACAGCAGAATGGAGCAACTTGAAAATTTCTATCCCA
ATGTTACTCACTCTCAGCATTACTGGGATCTCTTTTTGCGGAGCTGACATAGGCGGGTTCATTGGGAA
TCCAGAGACAGAGCTGCTAGTGCGTTGGTACCAGGCTGGAGCCTACCAGCCCTTCTTCCGTGGCCATG
CCACCATGAACACCAAGCGACGAGAGCCCTGGCTCTTTGGGGAGGAACACACCCGACTCATCCGAGAA
GCCATCAGAGAGCGCTATGGCCTCCTGCCATATTGGTATTCTCTGTTCTACCATGCACACGTGGCTTC
CCAACCTGTCATGAGGCCTCTGTGGGTAGAGTTCCCTGATGAACTAAAGACTTTTGATATGGAAGATG
AATACATGCTGGGTAGTGCATTATTGGTTCATCCAGTCACAGAACCAAAAGCCACCACAGTTGATGTG
TTTCTTCCAGGATCAAATGAGGTATGGTATGACTATAAGACATTTGCTCATTGGGAAGGAGGGTGTAC
TGTAAAGATCCCAGTAGCCTTGGACACTATTCCAGTGTTTCAGCGAGGTGGAAGTGTGATACCAATAA
AGACAACTGTAGGAAAATCCACAGGCTGGATGACTGAATCCTCCTATGGACTCCGGGTTGCTCTAAGC
ACTAAGGGTTCTTCAGTGGGTGAGTTATATCTTGATGATGGCCATTCATTCCAATACCTCCACCAGAA
GCAATTTTTGCACAGGAAGTTTTCATTCTGTTCCAGTGTTCTGATCAATTCCAGTTTTGCTGACCAGA
GGGGTCATTATCCCAGCAAGTGTGTGGTGGAGAAGATCTTGGTCTTAGGCTTCAGGAAGGAGCCATCT
TCTGTGACTACCCACTCATCTGGTGATGGTAAAGATCAGCCTGTGGCTTTTACGTATTGTGCCAAAAC
ATCCATCCTGAGCCTGGAGAAGCTCTCACTCAACATTGCCACTGACTGGGAGGTCCGCATCATA**TGA**C
AAAGAACTGCCCCTGGTGATGTGAGC

NOV44a, CG55752-07 Protein Sequence	SEQ ID NO: 660	656 aa	MW at 74971.6kD

MGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAVEYTLTQMGPVAAKQKVRSRTHVHWMSESG
IIDVFLLTGPTPSDVFKQYSHLTGTQAMPPLFSLGYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWL
DIEHTEGKRYFTWDKNRFPNPKRMQELLRSKKRKLVVISDPHIKIDPDYSVYVKAKDQGFFVKNQEGE
DFEGVCWPGLSSYLDFTNPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHG
NWEHRELHNIYGFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYGGAVWTGDNTAEWSNLKISIP
MLLTLSITGISFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEEHTRLIRE
AIRERYGLLPYWYSLFYHAHVASQPVMRPLWVEFPDELKTFDMEDEYMLGSALLVHPVTEPKATTVDV
FLPGSNEVWYDYKTFAHWEGGCTVKIPVALDTIPVFQRGGSVIPIKTTVGKSTGWMTESSYGLRVALS
TKGSSVGELYLDDGHSFQYLHQKQFLHRKFSFCSSVLINSSFADQRGHYPSKCVVEKILVLGFRKEPS
SVTTHSSGDGKDQPVAFTYCAKTSILSLEKLSLNIATDWEVRII

NOV44b, CG55752-06 DNA Sequence	SEQ ID NO: 661	2001 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1972

ATGGGCATTTATGGTTCAGTACCTTATCTCCTGGCCCACAAACTGGGCAGAACTATAGGTATTTTCTG
GCTGAATGCCTCGGAAACACTGGTGGAGATCAATACAGAGCCTGCAGTAGAGTACACACTGACCCAGA
TGGGCCCAGTTGCTGCTAAACAAAAGGTCAGATCTCGCACTCATGTGCACTGGATGTCAGAGAGTGGC
ATCATTGATGTTTTTCTGCTGACAGGACCTACACCTTCTGATGTCTTCAAACAGTACTCACACCTTAC
AGGTATTCAAGCCATGCCCCCTCTTTTCTCTTTGGGATACCACCAGTGCCGCTGGAACTATGAAGATG
AGCAGGATGTAAAAGCAGTGGATGCAGGGTTTGATGAGCATGACATTCCTTATGATGCCATGTGGCTG
GACATAGAGCACACTGAGGGCAAGAGGTACTTCACCTGGGACAAAAACAGATTCCCAAACCCCAAGAG

GATGCAAGAGCTGCTCAGGAGCAAAAAGCGTAAGGTACTTGTGGTCATCAGTGATCCCCACATCAAGA
TTGATCCTGACTACTCAGTATATGTGAAGGCCAAAGATCAGGGCTTCTTTGTGAAGAATCAGGAAGGG
GAAGACTTTGAAGGGGTGTGTTGGCCAGGTCTCTCCTCTTACCTGGATTTCACCAATCCCAAGGTCAG
AGAGTGGTATTCAAGTCTTTTTGCTTTCCCTGTTTATCAGGGATCTACGGACATCCTCTTCCTTTGGA
ATGACATGAATGAGCCTTCTGTCTTTAGAGGGCCAGAGCAAACCATGCAGAAGAATGCCATTCATCAT
GGCAATTGGGAGCACAGAGAGCTCCACAACATCTACGGTTTTTATCATCAAATGGCTACTGCAGAAGG
ACTGATAAAACGATCTAAAGGGAAGGAGAGACCCTTTGTTCTTACACGTTCTTTCTTTGCTGGATCAC
AAAAGTATGGTGCCGTGTGGACAGGCGACAACACAGCAGAATGGAGCAACTTGAAAATTTCTATCCCA
ATGTTACTCACTCTCAGCATTACTGGGATCTCTTTTTGCGGAGCTGACATAGGCGGGTTCATTGGGAA
TCCAGAGACAGAGCTGCTAGTGCGTTGGTACCAGGCTGGAGCCTACCAGCCCTTCTTCCGTGGCCATG
CCACCATGAACACCAAGCGACGAGAGCCCTGGCTCTTTGGGGAGGAACACACCCGACTCATCCGAGAA
GCCATCAGAGAGCGCTATGGCCTCCTGCCATATTGGTATTCTCTGTTCTACCATGCACACGTGGCTTC
CCAACCTGTCATGAGGAGGCCTCTGTGGGTAGAGTTCCCTGATGAACTAAAGACTTTTGATATGGAAG
ATGAATACATGCTGGGTTTAGGGAGTGCATTATTGGTTCATCCAGTCACAGAACCAAAAGCCACCACA
GTTGATGTGTTTCTTCCAGGATCAAATGAGGTATGGTATGACTATAAGACATTTGCTCATTGGGAAGG
AGGGTGTACTGTAAAGATCCCAGTAGCCTTGGACACTATTCCAGTGTTTCAGCGAGGTGGAAGTGTGA
TACCAATAAAGACAACTGTAGGAAAATCCACAGGCTGGATGACTGAATCCTCCTATGGACTCCGGGTT
GCTCTAAGCACTGGTTCTTCAGTGGGTGAGTTATATCTTGATGATGGCCATTCATTCCAATACCTCCA
CCAGAAGCAATTTTTGCACAGGAAGTTTTCATTCTGTTCCAGTGTTCTGATCAATTCCAGTTTTGCTG
ACCAGAGGGGTCATTATCCCAGCAAGTGTGTGGTGGAGAAGATCTTGGTCTTAGGCTTCAGGAAGGAG
CCATCTTCTGTGACTACCCACTCATCTGATGGTAAAGATCAGCCTGTGGCTTTTACGTATTGTGCCAA
AACATCCATCCTGAGCCTGGAGAAGCTCTCACTCAACATTGCCACTGACTGGGAGGTCCGCATCATAT
GACAAAGAACTGCCCCTGGTGATGTGAGC

NOV44b, CG55752-06 Protein Sequence	SEQ ID NO: 662	657 aa	MW at 75166.9kD

MGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAVEYTLTQMGPVAAKQKVRSRTHVHWMSESG
IIDVFLLTGPTPSDVFKQYSHLTGIQAMPPLFSLGYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWL
DIEHTEGKRYFTWDKNRFPNPKRMQELLRSKKRKVLVVISDPHIKIDPDYSVYVKAKDQGFFVKNQEG
EDFEGVCWPGLSSYLDFTNPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHH
GNWEHRELHNIYGFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYGAVWTGDNTAEWSNLKISIP
MLLTLSITGISFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEEHTRLIRE
AIRERYGLLPYWYSLFYHAHVASQPVMRRPLWVEFPDELKTFDMEDEYMLGLGSALLVHPVTEPKATT
VDVFLPGSNEVWYDYKTFAHWEGGCTVKIPVALDTIPVFQRGGSVIPIKTTVGKSTGWMTESSYGLRV
ALSTGSSVGELYLDDGHSFQYLHQKQFLHRKFSFCSSVLINSSFADQRGHYPSKCVVEKILVLGFRKE
PSSVTTHSSDGKDQPVAFTYCAKTSILSLEKLSLNIATDWEVRII

NOV44c, CG55752-01 DNA Sequence	SEQ ID NO: 663	3025 bp
	ORF Start: ATG at 28	ORF Stop: TGA at 2929

ACAGGTGCCTGGGGGTCAGGCTTCCGCATGCGGGCTGCAGTTGCTGGCATTGCCTTCCGCAGGAGGCG
TCAGAAACAGTGGCTTTCCAAGAAGTCCACCTATCAGGCATTATTGGATTCAGTCACAACAGATGAAG
ACAGCACCAGGTTCCAAATCATCAATGAAGCAAGTAAGGTGAGGCGTCAGAAACAGTGGCTTTCCAAG
AAGTCCACCTATCAGGCATTATTGGATTCAGTCACAACAGATGAAGACAGCACCAGGTTCCAAATCAT
CAATGAAGCAAGTAAGGTGCCTCTCCTGGCTGAAATTTATGGTATAGAAGGAAACATTTTCAGGCTTA
AAATTAATGAAGAGACTCCTCTAAAACCCAGATTTGAAGTTCCGGATGTCCTCACAAGCAAGCCAAGC
ACTGTAAGGATTTCATGCTCTGGGGACACAGGCAGTCTGATATTGGCAGATGGAAAAGGAGACCTGAA
GTGCCATATCACAGCAAACCCATTCAAGGTAGACTTGGTGTCTGAAGAAGAGGTTGTGATTAGCATAA
ATTCCCTGGGCCAATTATACTTTGAGCATGGCAGGGCCCCTAGGGTCTCTTTCTCGGATAAGGTTAAT
CTCACGCTTGGTAGCATATGGGATAAGATCAAGAACCTTTTCTCTAGGCAAGGATCAAAAGACCCAGC
TGAGGGCGATGGGGCCCAGCCTGAGGAAACACCCAGGGATGGCGACAAGCCAGAGGAGACTCAGGGGA
AGGCAGAGAAAGATGAGCCAGGAGCCTGGGAGGAGACATTCAAAACTCACTCTGACAGCAAGCCGTAT
GGCCCTTCTTCTATTGGTTTGGATTTCTCCTTGCATGGATTTGAGCATCTTTATGGGATCCCACAACA
TGCAGAATCACACCAACTTAAAAATACTGGGGATGGAGATGCTTACCGTCTTTATAACCTGGATGTCT
ATGGATACCAAATATATGATAAAATGGGCATTTATGGTTCAGTACCTTATCTCCTGGCCCACAAACTG
GGCAGAACTATAGGTATTTTCTGGCTGAATGCCTCGGAAACACTGGTGGAGATCAATACAGAGCCTGC
AGGGATAGTCATCTTTGGTCCTGTCTCTTTGATTTATCAAAGCCAGGGAGATACACCTCTAACAACTC
ATGTGCACTGGATGTCAGAGAGTGGCATCATTGATGTTTTTCTGCTGACAGGACCTACACCTTCTGAT
GTCTTCAAACAGTACTCACACCTTACAGGTACACAAGCCATGCCCCCTCTTTTCTCTTTGGGATACCA
CCAGTGCCGCTGGAACTATGAAGATGAGCAGGATGTAAAAGCAGTGGATGCAGGGTTTGATGAGCATG

ACATTCCTTATGATGCCATGTGGCTGGACATAGAGCACACTGAGGGCAAGAGGTACTTCACCTGGGAC
AAAAACAGATTCCCAAACCCCAAGAGGATGCAAGAGCTGCTCAGGAGCAAAAAGCGTAAGCTTGTGGT
CATCAGTGATCCCCACATCAAGATTGAACCTGACTACTCAGTATATGTGAAGGCCAAAGATCAGGGCT
TCTTTGTGAAGAATCAGGAAGGGGAAGACTTTGAAGGGGTGTGTTGGCCAGGTATGAAATCATACCTG
GATTTCACCAATCCCAAGGTCAGAGAGTGGTATTCAAGTATGTTCAGTTCCAATTGTGATGGATCTAC
GGACATCCTCTTCCTTTGGAATGACATGAATGAGCCTTCTGTCTTTAGAGGGCCAGAGCAAACCATGC
AGAAGAATGCCATTCATCATGGCAATTGGGAGCACAGAGAGCTCCACAACATCTACGGTTTTTATATG
GCTACTGCAGAAGGACTGATAAAACGATCTAAAGGGAAGGAGAGACCCTTTGTTCTTACACGTTCTTT
CTTTGCTGGATCACAAAAGTATGGTGCCGTGTGGACAGGCGACAACACAGCAGAATGGAGCAACTTGA
AAATTTCTATCCCAATGTTACTCACTCTCAGCATTACTGGGATCTCTTTTTGCGGAGCTGACATAGGC
GGGTTCATTGGGAATCCAGAGACAGAGCTGCTAGTGCGTTGGTACCAGGCTGGAGCCTACCAGCCCTT
CTTCCGTGGCCATGCCACCATGAACACCAAGCGACGAGAGCCCTGGCTCTTTGGGGAGGAACACACCC
GACTCATCCGAGAAGCCATCAGAGAGCGCTATGGCCTCCTGCCATATTGGTATTCTCTGTTCTACCAT
GCACACGTGGCTTCCCAACCTGTCATGAGGCCTCTGTGGGTAGAGTTCCCTGATGAACTAAAGACTTT
TGATATGGAAGATGAATACATGTTAGGGAGTGCATTATTGGTTCATCCAGTCACAGAACCAAAAGCCA
CCACAGTTGATGTGTTTCTTCCAGGATCAAATGAGGTAGTCTGGTATGACTATAAGACATTTGCTCAT
TGGGAAGGAGGGTGTACTGTAAAGATCCCAGTACTGTTACAGATTCCAGTGTTTCAGCGAGGTGGAAG
TGTGATACCAATAAAGACAACTGTAGGAAAATCCACAGGCTGGATGACTGAATCCTCCTATGGACTCC
GGGTTGCTCTAAGCACTCTCCAGGGTTCTTCAGTGGGTGAGTTATATCTTGATGATGGCCATTCATTC
CAATACCTCCACCAGAAGCAATTTTTGCACAGGAAGTTTTCATTCTGTTCCAGTGTTCTGGTGGCCTC
CTCTCCAGTATCTCAAGGACACTTACATACCCCACTCAGCATGACAAAAGCCCTGCTTTTCACTGTAT
CGTCTCCAGCCAGCGTGAAAATGCGGCTTCACTACAGCCCAGAGAAAAGGGCCAGGTTTAGTCATTGT
GCCAAAACATCCATCCTGAGCCTGGAGAAGCTCTCACTCAACATTGCCACTGACTGGGAGGTCCGCAT
CATATGACAAAGAACTGCCCCTGGTGATGTGAGCAGGGACCTGCCTGCCCCTTTCAACCTTTCCCCTC
ACCTTTTTTGAGATTTTTGCTGCAATCTGTTTG

NOV44c, CG55752-01 Protein Sequence	SEQ ID NO: 664	967 aa	MW at 109801.3kD

MRAAVAGIAFRRRRQKQWLSKKSTYQALLDSVTTDEDSTRFQIINEASKVRRQKQWLSKKSTYQALLD
SVTTDEDSTRFQIINEASKVPLLAEIYGIEGNIFRLKINEETPLKPRFEVPDVLTSKPSTVRISCSGD
TGSLILADGKGDLKCHITANPFKVDLVSEEEVVISINSLGQLYFEHGRAPRVSFSDKVNLTLGSIWDK
IKNLFSRQGSKDPAEGDGAQPEETPRDGDKPEETQGKAEKDEPGAWEETFKTHSDSKPYGPSSIGLDF
SLHGFEHLYGIPQHAESHQLKNTGDGDAYRLYNLDVYGYQIYDKMGIYGSVPYLLAHKLGRTIGIFWL
NASETLVEINTEPAGIVIFGPVSLIYQSQGDTPLTTHVHWMSESGIIDVFLLTGPTPSDVFKQYSHLT
GTQAMPPLFSLGYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYFTWDKNRFPNPKR
MQELLRSKKRKLVVISDPHIKIEPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGMKSYLDFTNPKVRE
WYSSMFSSNCDGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHNIYGFYMATAEGLIKR
SKGKERPFVLTRSFFAGSQKYGAVWTGDNTAEWSNLKISIPMLLTLSITGISFCGADIGGFIGNPETE
LLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEEHTRLIREAIRERYGLLPYWYSLFYHAHVASQPVM
RPLWVEFPDELKTFDMEDEYMLGSALLVHPVTEPKATTVDVFLPGSNEVVWYDYKTFAHWEGGCTVKI
PVLLQIPVFQRGGSVIPIKTTVGKSTGWMTESSYGLRVALSTLQGSSVGELYLDDGHSFQYLHQKQFL
HRKFSFCSSVLVASSPVSQGHLHTPLSMTKALLFTVSSPASVKMRLHYSPEKRARFSHCAKTSILSLE
KLSLNIATDWEVRII

NOV44d, CG55752-02 DNA Sequence	SEQ ID NO: 665	4483 bp
	ORF Start: ATG at 204	ORF Stop: TGA at 2946

AACGCTAGTTTGGGCCTGAAAAATTCCAGGAGCAAGAGTCAAGATTTGTCACTCCATGAGAATCTGGA
GGGGACTCCCTTCCCAGAAACTTGACGATGAAGTACTGGTTGTAATTTTAGAAAGACACCCAATCGGC
TTTTTTAAAAGATCGCCCAGGGCCCTTGTCCTGAGAGCTGGGAGCTGGTCGGAGTGACAGAGAAGCCA
TGGAAGCAGCAGTGAAAGAGGAAATAAGTGTTGAAGATGAAGCTGTAGATAAAAACATTTTCAGAGAC
TGTAACAAGATCGCATTTTACAGGCGTCAGAAACAGTGGCTTTCCAAGAAGTCCACCTATCAGGCATT
ATTGGATTCAGTCACAACAGATGAAGACAGCACCAGGTTCCAAATCATCAATGAAGCAAGTAAGGTTC
CTCTCCTGGCTGAAATTTATGGTATAGAAGGAAACATTTTCAGGCTTAAAATTAACGAAGAGACTCCT
CTAAAACCCAGATTTGAAGTTCCGGATGTCCTCACAAGCAAGCCAAGCACTGTAAGGCTGATTTCATG
CTCTGGGGACACAGGCAGTCTGATATTGGCAGATGGAAAAGGAGACCTGAAGTGCCATATCACAGCAA
ACCCATTCAAGGTAGACTTGGTGTCTGAAGAAGAGGTTGTGATTAGCATAAATTCCCTGGGCCAATTA
TACTTTGAGCATCTACAGATTCTTCACAAACAAAGAGCTGCTAAAGAAAATGAGGAGGAGACATCAGT
GGACACCTCTCAGGAAAATCAAGAAGATCTGGGCCTGTGGGAAGAGAAATTTGGAAAATTTGTGGATA
TCAAAGCTAATGGCCCTTCTTCTATTGGTTTGGATTTCTCCTTGCATGGATTTGAGCATCTTTATGGG

ATCCCACAACATGCAGAATCACACCAACTTAAAAATACTGGTGATGGAGATGCTTACCGTCTTTATAA
CCTGGATGTCTATGGATACCAAATATATGATAAAATGGGCATTTATGGTTCAGTACCTTATCTCCTGG
CCCACAAACTGGGCAGAACTATAGGTATTTTCTGGCTGAATGCCTCGGAAACACTGGTGGAGATCAAT
ACAGAGCCTGCAGTAGAGTACACACTGACCCAGATGGGCCCAGTTGCTGCTAAACAAAAGGTCAGATC
TCGCACTCATGTGCACTGGATGTCAGAGAGTGGCATCATTGATGTTTTTCTGCTGACAGGACCTACAC
CTTCTGATGTCTTCAAACAGTACTCACACCTTACAGGCACACAAGCCATGCCCCCTCTTTTCTCTTTG
GGATACCACCAGTGCCGCTGGAACTATGAAGATGAGCAGGATGTAAAAGCAGTGGATGCAGGGTTTGA
TGAGCATGACATTCCTTATGATGCCATGTGGCTGGACATAGAGCACACTGAGGGCAAGAGGTACTTCA
CCTGGGACAAAAACAGATTCCCAAACCCCAAGAGGATGCAAGAGCTGCTCAGGAGCAAAAAGCGTAAG
CTTGTGGTCATCAGTGATCCCCACATCAAGATTGATCCTGACTACTCAGTATATGTGAAGGCCAAAGA
TCAGGGCTTCTTTGTGAAGAATCAGGAAGGGGAAGACTTTGAAGGGGTGTGTTGGCCAGGTCTCTCCT
CTTACCTGGATTTCACCAATCCCAAGGTCAGAGAGTGGTATTCAAGTCTTTTTGCTTTCCCTGTTTAT
CAGGGATCTACGGACATCCTCTTCCTTTGGAATGACATGAATGAGCCTTCTGTCTTTAGAGGGCCAGA
GCAAACCATGCAGAAGAATGCCATTCATCATGGCAATTGGGAGCACAGAGAGCTCCACAACATCTACG
GTTTTTATCATCAAATGGCTACTGCAGAAGGACTGATAAAACGATCTAAAGGGAAGGAGAGACCCTTT
GTTCTTACACGTTCTTTCTTTGCTGGATCACAAAAGTATGGTGCCGTGTGGACAGGCGACAACACAGC
AGAATGGAGCAACTTGAAAATTTCTATCCCAATGTTACTCACTCTCAGCATTACTGGGATCTCTTTTT
GCGGAGCTGACATAGGCGGGTTCATTGGGAATCCAGAGACAGAGCTGCTAGTGCGTTGGTACCAGGCT
GGAGCCTACCAGCCCTTCTTCCGTGGCCATGCCACCATGAACACCAAGCGACGAGAGCCCTGGCTCTT
TGGGGAGGAACACACCCGACTCATCCGAGAAGCCATCAGAGAGCGCTATGGCCTCCTGCCATATTGGT
ATTCTCTGTTCTACCATGCACACGTGGCTTCCCAACCTGTCATGAGGCCTCTGTGGGTAGAGTTCCCT
GATGAACTAAAGACTTTTGATATGGAAGATGAATACATGCTGGGGAGTGCATTATTGGTTCATCCAGT
CACAGAACCAAAAGCCACCACAGTTGATGTGTTTCTTCCAGGATCAAATGAGGTCTGGTATGACTATA
AGACATTTGCTCATTGGGAAGGAGGGTGTACTGTAAAGATCCCAGTAGCCTTGGACACTATTCCAGTG
TTTCAGCGAGGTGGAAGTGTGATACCAATAAAGACAACTGTAGGAAAATCCACAGGCTGGATGACTGA
ATCCTCCTATGGACTCCGGGTTGCTCTAAGCACTAAGGGTTCTTCAGTGGGTGAGTTATATCTTGATG
ATGGCCATTCATTCCAATACCTCCACCAGAAGCAATTTTTGCACAGGAAGTTTTCATTCTGTTCCAGT
GTTCTGATCAATAGTTTTGCTGACCAGAGGGGTCATTATCCCAGCAAGTGTGTGGTGGAGAAGATCTT
GGTCTTAGGCTTCAGGAAGGAGCCATCTTCTGTGACTACCCACTCATCTGATGGTAAAGATCAGCCTG
TGGCTTTTACGTATTGTGCCAAAACATCCATCCTGAGCCTGGAGAAGCTCTCACTCAACATTGCCACT
GACTGGGAGGTCCGCATCATATGACAAAGAACTGCCCCTGGTGATGTGAGCAGGGACCTGCCTGCCCC
TTTCAACCTTTCCCCTCACCTTTTTTGAGATTTTTGCTGCAATCTGTTTGCCTTCCCTGAATCAAAAT
AATCTTTCATTCGTCACCATTATACTAATGAACAATAGATTTCATGTTTCAAAATTTCAGATTTTACA
TGTTAAGATGTACTAACAATATTCCTTGTATCAAACATCTCCTTTTCTCCCTGATACATAGCCCTGAG
ACATTTATAGCGTTCAGGAGTCTTCTATTGCTTCCATTCCTTCAGCAGGGCTGCGTGGGTCTGTTTTA
ACGTGGGCCAAGCCTACCTGGGCAGCCCATTTGCCAGGGCTTGCCTCAGGCCATGCAGCATTGGCGCT
CTGGCTGCAGCAGCTGAGTTGCTCAAGGCCAGTGTCCAAGTGGACAGCAGCCTCTGGTACTCCCCCCA
GTTATCTTCCACCCACATGGACTGGGCAGAGCAGCCCTCTTCTGTGTGCACTGCATACGCTGCAGCCG
TGGGAGTTATTCTCCCCTAGAGATCGACTTGGCAGCACGAAGGATTCTTTTCTCTTTCATGCTTCTCA
GGCTCAATAGTTTCTAATTAATCTTAAAATCCATGTCTTTTACATTGTTTTTTTAATTAAGTGCTGTT
TACTAACCAAATAATATTTATAACATGAGTAAGCTATAATTAATAACAATGAAATAAATACCCATGTA
CCCACCACTGGACTTCAGAAGTAGAACTCATGACTGGGACTAGGATGAGGCAAGGGAGACCCTGGCCT
TGGGCACAAAATGTAAGGGATGCCAAAAAAATACAGTAATCAAAGTAAGTAATATTTCAATCCAATAT
TTTTAAAAATCAGAATTAATGCAAAAAAAACCATGATGAACAAAATATTAAAATTTAAAATAAAGACA
GGATTAGTATTACTGAGTTTTCCTTTTGTCCCAGGCTTTAATATGGCTTGGCATGGGGCAGAACATTA
CAACATACCAGTCGTGTCATGGTGCCCAAGGCTCCACAGACCTCAGTGGCTCCCTGCTGCCTGCCACA
GCATCTGTTTTAGCAGCCTCGACTCCTCAGCACTCCTCAGCACACACCTCTTCTTATCAGGCTTCCTC
CACTTAGCAACTTGCTAACGGCCACCTCTGTGCCTTCTGATCCCTGGGCGCCAATATCCTCCTGCCCT
TACCATCCTTCCAGGCCCAACTTAAATCCCACTTTCCCATGAAGCCTAACTGCGTGAACACCCCTACC
CCCATACCCATTAGCAGTGATTTTGCCCTTCCCCGTAATGCTGTCCCACTTATAACTGTGCTCTACTT
AGCATTCTCAGGGATCATACCTTAATGTTTTCAGTATGTCTGCGTTCTCCTACTAGATTGTATGTCCC
TCAAGAGCATGTTCTGTTTCTCTTCTGTCTGACAGAGCACTATTATACCTGACTTTCAGTAACTGTTA
GCTGTGATTAGTTAGCTGGTGGATTTAATTGATTAAAAAATTACGATTGAATGTAAAAAAAAA

NOV44d, CG55752-02 Protein Sequence	SEQ ID NO: 666	914 aa	MW at 104319.2kD

MEAAVKEEISVEDEAVDKNIFRDCNKIAFYRRQKQWLSKKSTYQALLDSVTTDEDSTRFQIINEASKV
PLLAEIYGIEGNIFRLKINEETPLKPRFEVPDVLTSKPSTVRLISCSGDTGSLILADGKGDLKCHITA
NPFKVDLVSEEEVVISINSLGQLYFEHLQILHKQRAAKENEEETSVDTSQENQEDLGLWEEKFGKFVD

IKANGPSSIGLDFSLHGFEHLYGIPQHAESHQLKNTGDGDAYRLYNLDVYGYQIYDKMGIYGSVPYLL
AHKLGRTIGIFWLNASETLVEINTEPAVEYTLTQMGPVAAKQKVRSRTHVHWMSESGIIDVFLLTGPT
PSDVFKQYSHLTGTQAMPPLFSLGYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYF
TWDKNRFPNPKRMQELLRSKKRKLVVISDPHIKIDPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGLS
SYLDFTNPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHNIY
GFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYGAVWTGDNTAEWSNLKISIPMLLTLSITGISF
CGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEEHTRLIREAIRERYGLLPYW
YSLFYHAHVASQPVMRPLWVEFPDELKTFDMEDEYMLGSALLVHPVTEPKATTVDVFLPGSNEVWYDY
KTFAHWEGGCTVKIPVALDTIPVFQRGGSVIPIKTTVGKSTGWMTESSYGLRVALSTKGSSVGELYLD
DGHSFQYLHQKQFLHRKFSFCSSVLINSFADQRGHYPSKCVVEKILVLGFRKEPSSVTTHSSDGKDQP
VAFTYCAKTSILSLEKLSLNIATDWEVRII

NOV44e, CG55752-03 DNA Sequence	SEQ ID NO: 667	3015 bp
	ORF Start: ATG at 204	ORF Stop: TGA at 2946

AACGCTAGTTTGGGCCTGAAAAATTCCAGGAGCAAGAGTCAAGATTTGTCACTCCATGAGAATCTGGA
GGGGACTCCCTTCCCAGAAACTTGACGATGAAGTACTGGTTGTAATTTTAGAAAGACACCCAATCGGC
TTTTTTAAAAGATCGCCCAGGGCCCTTGTCCTGAGAGCTGGGAGCTGGTCGGAGTGACAGAGAAGCCA
TGGAAGCAGCAGTGAAAGAGGAAATAAGTGTTGAAGATGAAGCTGTAGATAAAAACATTTTCAGAGAC
TGTAACAAGATCGCATTTTACAGGCGTCAGAAACAGTGGCTTTCCAAGAAGTCCACCTATCAGGCATT
ATTGGATTCAGTCACAACAGATGAAGACAGCACCAGGTTCCAAATCATCAATGAAGCAAGTAAGGTTC
CTCTCCTGGCTGAAATTTATGGTATAGAAGGAAACATTTTCAGGCTTAAAATTAACGAAGAGACTCCT
CTAAAACCCAGATTTGAAGTTCCGGATGTCCTCACAAGCAAGCCAAGCACTGTAAGGCTGATTTCATG
CTCTGGGGACACAGGCAGTCTGATATTGGCAGATGGAAAAGGAGACCTGAAGTGCCATATCACAGCAA
ACCCATTCAAGGTAGACTTGGTGTCTGAAGAAGAGGTTGTGATTAGCATAAATTCCCTGGGCCAATTA
TACTTTGAGCATCTACAGATTCTTCACAAACAAAGAGCTGCTAAAGAAAATGAGGAGGAGACATCAGT
GGACACCTCTCAGGAAAATCAAGAAGATCTGGGCCTGTGGGAAGAGAAATTTGGAAAATTTGTGGATA
TCAAAGCTAATGGCCCTTCTTCTATTGGTTTGGATTTCTCCTTGCATGGATTTGAGCATCTTTATGGG
ATCCCACAACATGCAGAATCACACCAACTTAAAAATACTGGTGATGGAGATGCTTACCGTCTTTATAA
CCTGGATGTCTATGGATACCAAATATATGATAAAATGGGCATTTATGGTTCAGTACCTTATCTCCTGG
CCCACAAACTGGGCAGAACTATAGGTATTTTCTGGCTGAATGCCTCGGAAACACTGGTGGAGATCAAT
ACAGAGCCTGCAGTAGAGTACACACTGACCCAGATGGGCCCAGTTGCTGCTAAACAAAAGGTCGGATC
TCGCACTCATGTGCACTGGATGTCAGAGAGTGGCATCATTGATGTTTTTCTGCTGACAGGACCTACAC
CTTCTGATGTCTTCAAACAGTACTCACACCTTACAGGCACACAAGCCATGCCCCCTCTTTTCTCTTTG
GGATACCACCAGTGCCGCTGGAACTATGAAGATGAGCAGGATGTAAAAGCAGTGGATGCAGGGTTTGA
TGAGCATGACATTCCTTATGATGCCATGTGGCTGGACATAGAGCACACTGAGGGCAAGAGGTACTTCA
CCTGGGACAAAAACAGATTCCCAAACCCCAAGAGGATGCAAGAGCTGCTCAGGAGCAAAAAGCGTAAG
CTTGTGGTCATCAGTGATCCCCACATCAAGATTGATCCTGACTACTCAGTATATGTGAAGGCCAAAGA
TCAGGGCTTCTTTGTGAAGAATCAGGAAGGGGAAGACTTTGAAGGGGTGTGTTGGCCAGGTCTCTCCT
CTTACCTGGATTTCACCAATCCCAAGGTCAGAGAGTGGTATTCAAGTCTTTTTGCTTTCCCTGTTTAT
CAGGGATCTACGGACATCCTCTTCCTTTGGAATGACATGAATGAGCCTTCTGTCTTTAGAGGGCCAGA
GCAAACCATGCAGAAGAATGCCATTCATCATGGCAATTGGGAGCACAGAGAGCTCCACAACATCTACG
GTTTTTATCATCAAATGGCTACTGCAGAAGGACTGATAAAACGATCTAAAGGGAAGGAGAGACCCTTT
GTTCTTACACGTTCTTTCTTTGCTGGATCACAAAAGTATGGTGCCGTGTGGACAGGCGACAACACAGC
AGAATGGAGCAACTTGAAAATTTCTATCCCAATGTTACTCACTCTCAGCATTACTGGGGTCTCTTTTT
GCGGAGCTGACATAGGCGGGTTCATTGGGAATCCAGAGACAGAGCTGCTAGTGCGTTGGTACCAGGCT
GGAGCCTACCAGCCCTTCTTCCGTGGCCATGCCACCATGAACACCAAGCGACGAGAGCCCTGGCTCTT
TGGGGAGGAACACACCCGACTCATCCGAGAAGCCATCAGAGAGCGCTATGGCCTCCTGCCATATTGGT
ATTCTCTGTTCTACCATGCACACGTGGCTTCCCAACCTGTCATGAGGCCTCTGTGGGTAGAGTTCCCT
GATGAACTAAAGACTTTTGATATGGAAGATGAATACATGCTGGGGAGTGCATTATTGGTTCATCCAGT
CACAGAACCAAAAGCCACCACAGTTGATGTGTTTCTTCCAGGATCAAATGAGGTCTGGTATGACTATA
AGACATTTGCTCATTGGGAAGGAGGGTGTACTGTAAAGATCCCAGTAGCCTTGGACACTATTCCAGTG
TTTCAGCGAGGTGGAAGTGTGATACCAATAAAGACAACTGTAGGAAAATCCACAGGCTGGATGACTGA
ATCCTCCTATGGACTCCGGGTTGCTCTAAGCACTAAGGGTTCTTCAGTGGGTGAGTTATATCTTGATG
ATGGCCATTCATTCCAATACCTCCACCAGAAGCAATTTTTGCACAGGAAGTTTTCATTCTGTTCCAGT
GTTCTGATCAATAGTTTTGCTGACCAGAGGGGTCACTATCCCAGCAAGTGTGTGGTGGAGAAGATCTT
GGTCTTAGGCTTCAGGAAGGAGCCATCTTCTGTGACTACCCACTCATCTGATGGTAAAGATCAGCCTG
TGGCTTTTACGTATTGTGCCAAAACATCCATCCTGAGCCTGGAGAAGCTCTCACTCAACATTGCCACT
GACTGGGAGGTCCGCATCATA**TGA**CAAAGAACTGCCCCTGGTGATGTGAGCAGGGACCTGCCTGCCCC
TTTCAACCTTTCCCCTCACCTTT

NOV44e, CG55752-03 Protein Sequence	SEQ ID NO: 668	914 aa	MW at 104206.0kD

MEAAVKEEISVEDEAVDKNIFRDCNKIAFYRRQKQWLSKKSTYQALLDSVTTDEDSTRFQIINEASKV
PLLAEIYGIEGNIFRLKINEETPLKPRFEVPDVLTSKPSTVRLISCSGDTGSLILADGKGDLKCHITA
NPFKVDLVSEEEVVISINSLGQLYFEHLQILHKQRAAKENEEETSVDTSQENQEDLGLWEEKFGKFVD
IKANGPSSIGLDFSLHGFEHLYGIPQHAESHQLKNTGDGDAYRLYNLDVYGYQIYDKMGIYGSVPYLL
AHKLGRTIGIFWLNASETLVEINTEPAVEYTLTQMGPVAAKQKVGSRTHVHWMSESGIIDVFLLTGPT
PSDVFKQYSHLTGTQAMPPLFSLGYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYF
TWDKNRFPNPKRMQELLRSKKRKLVVISDPHIKIDPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGLS
SYLDFTNPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHNIY
GFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYGAVWTGDNTAEWSNLKISIPMLLTLSITGVSF
CGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEEHTRLIREAIRERYGLLPYW
YSLFYHAHVASQPVMRPLWVEFPDELKTFDMEDEYMLGSALLVHPVTEPKATTVDVFLPGSNEVWYDY
KTFAHWEGGCTVKIPVALDTIPVFQRGGSVIPIKTTVGKSTGWMTESSYGLRVALSTKGSSVGELYLD
DGHSFQYLHQKQFLHRKFSFCSSVLINSFADQRGHYPSKCVVEKILVLGFRKEPSSVTTHSSDGKDQP
VAFTYCAKTSILSLEKLSLNIATDWEVRII

NOV44f, CG55752-04 DNA Sequence	SEQ ID NO: 669	3102 bp
	ORF Start: ATG at 103	ORF Stop: TGA at 2839

TACTGGTTGTAATTTTAGAAAGACACCCAATCGGCTTTTTTAAAAGATCGCCCAGGGCCCTTGTCCTG
AGAGCTGGGAGCTGGTCGGAGTGACAGAGAAGCC**ATG**GAAGCAGCAGTGAAAGAGGAAATAAGTGTTG
AAGATGAAGCTGTAGATAAAAACATTTTCAGAGACTGTAACAAGATCGCATTTTACAGGCGTCAGAAA
CAGTGGCTTTCCAAGAAGTCCACCTATCGGGCATTATTGGATTCAGTCACAACAGATGAAGACAGCAC
CAGGTTCCAAATCATCAATGAAGCAAGTAAGGTTCCTCTCCTGGCTGAAATTTATGGTATAGAAGGAA
ACATTTTCAGGCTTAAAATTAACGAAGAGACTCCTCTAAAACCCAGATTTGAAGTTCCGGATGTCCTC
ACAAGCAAGCCAAGCACTGTAAGGCTGATTTCATGCTCTGGGGACACAGGCAGTCTGATATTGGCAGA
TGGAAAAGGAGACCTGAAGTGCCATATCACAGCAAACCCATTCAAGGTAGACTTGGTGTCTGAAGAAG
AGGTTGTGATTAGCATAAATTCCCTGGGCCAATTATACTTTGAGCATCTACAGATTCTTCACAAACAA
AGAGCTGCTAAAGAAAATGAGGAGGAGACATCAGTGGACACCTCTCAGGAAAATCAAGAAGATCTGGG
CCTGTGGGAAGAGAAATTTGGAAAATTTGTGGATATCAAAGCTAATGGCCCTTCTTCTATTGGTTTGG
ATTTCTCCTTGCATGGATTTGAGCATCTTTATGGGATCCCACAACATGCAGAATCACACCAACTTAAA
AATACTGGAGATGCTTACCGTCTTTATAACCTGGATGTCTATGGATACCAAATATATGATAAAATGGG
CATTTATGGTTCAGTACCTTATCTCCTGGCCCACAAACTGGGCAGAACTATAGGTATTTTCTGGCTGA
ATGCCTCGGAAACACTGGTGGAGATCAATACAGAGCCTGCAGTAGAGTACACACTGACCCAGATGGGC
CCAGTTGCTGCTAAACAAAAGGTCAGATCTCGCACTCATGTGCACTGGATGTCAGAGAGTGGCATCAT
TGATGTTTTTCTGCTGACAGGACCTACACCTTCTGATGTCTTCAAACAGTACTCACACCTTACAGGTA
CGCAAGCCATGCCCCCTCTTTTCTCTTTGGGATACCACCAGTGCCGCTGGAACTATGAAGATGAGCAG
GATGTAAAAGCAGTGGATGCAGGGTTTGATGAGCATGACATTCCTTATGATGCCATGTGGCTGGACAT
AGAGCACACTGAGGGCAAGAGGTACTTCACCTGGGACAAAAACAGATTCCCAAACCCCAAGAGGATGC
AAGAGCTGCTCAGGAGCAAAAAGCGTAAGCTTGTGGTCATCAGTGATCCCCACATCAAGATTGAACCT
GACTACTCAGTATATGTGAAGGCCAAAGATCAGGGCTTCTTTGTGAAGAATCAGGAAGGGGAAGACTT
TGAAGGGGTGTGTTGGCCAGGTCTCTCCTCTTACCTGGATTTCACCAATCCCAAGGTCAGAGAGTGGT
ATTCAAGTCTTTTTGCTTTCCCTGTTTATCAGGGATCTACGGACATCCTCTTCCTTTGGAATGACATG
AATGAGCCTTCTGTCTTTAGAGGGCCAGAGCAAACCATGCAGAAGAATGCCATTCATCATGGCAATTG
GGAGCACAGAGAGCTCCACAACATCTACGGTTTTTATCATCAAATGGCTACTGCAGAAGGACTGATAA
AACGATCTAAAGGGAAGGAGAGACCCTTTGTTCTTACACGTTCTTTCTTTGCTGGATCACAAAAGTAT
GGTGCCGTGTGGACAGGCGACAACACAGCAGAATGGAGCAACTTGAAAATTTCTATCCCAATGTTACT
CACTCTCAGCATTACTGGGATCTCTTTTTGCGGAGCTGACATAGGCGGGTTCATTGGGAATCCAGAGA
CAGAGCTGCTAGTGCGTTGGTACCAGGCTGGAGCCTACCAGCCCTTCTTCCGTGGCCATGCCACCATG
AACACCAAGCGACGAGAGCCCTGGCTCTTTGGGGAGGAACACACCCGACTCATCCGAGAAGCCATCAG
AGAGCGCTATGGCCTCCTGCCATATTGGTATTCTCTGTTCTACCATGCACACGTGGCTTCCCAACCTG
TCATGAGGCCTCTGTGGGTAGAGTTCCCTGATGAACTAAAGACTTTTGATATGGAAGATGAATACATG
TTAGGGAGTGCATTATTGGTTCATCCAGTCACAGAACCAAAAGCCACCACAGTTGATGTGTTTCTTCC
AGGATCAAATGAGGTATGGTATGACTATAAGACATTTGCTCATTGGGAAGGAGGGTGTACTGTAAAGA
TCCCAGTAGCCTTGGACACTATTCCAGTGTTTCAGCGAGGTGGAAGTGTGATACCAATAAAGACAACT
GTAGGAAAATCCACAGGCTGGATGACTGAATCCTCCTATGGACTCCGGGTTGCTCTAAGCACTCAGGG
TTCTTCAGTGGGTGAGTTATATCTTGATGATGGCCATTCATTCCAATACCTCCACCAGAAGCAATTTT
TGCACAGGAAGTTTTCATTCTGTTCCAGTGTTCTGATCAATAGTTTTGCTGACCAGAGGGGTCATTAT

CCCAGCAAGTGTGTGGTGGAGAAGATCTTGGTCTTAGGCTTCAGGAAGGAGCCATCTTCTGTGACTAC
CCACTCATCTGATGGTAAAGATCAGCCTGTGGCTTTTACGTATTGTGCCAAAACATCCATCCTGAGCC
TGGAGAAGCTCTCACTCAACATTGCCACTGACTGGGAGGTCCGCATCATATGACAAAGAACTGCCCCT
GGTGATGTGAGCAGGGACCTGCCTGCCCCTTTCAACCTTTCCCCTCACCTTTTTTGAGATTTTTGCTG
CAATCTGTTTGTCTTCCCTGAATCAAAATAATCTTTCATTCGTCACCATTATACTAATGAACAATAGA
TTTCATGTTTCAAAATTTCAGATTTTACATGTTAAGATGTACTAACAATATTCCTTGTATCAAACATC
TCCTTTTCTCCCTGATACATAGCCCTGAGACATTATAGCGTC

NOV44f, CG55752-04 Protein Sequence	SEQ ID NO: 670	912 aa	MW at 104189.1kD

MEAAVKEEISVEDEAVDKNIFRDCNKIAFYRRQKQWLSKKSTYRALLDSVTTDEDSTRFQIINEASKV
PLLAEIYGIEGNIFRLKINEETPLKPRFEVPDVLTSKPSTVRLISCSGDTGSLILADGKGDLKCHITA
NPFKVDLVSEEEVVISINSLGQLYFEHLQILHKQRAAKENEEETSVDTSQENQEDLGLWEEKFGKFVD
IKANGPSSIGLDFSLHGFEHLYGIPQHAESHQLKNTGDAYRLYNLDVYGYQIYDKMGIYGSVPYLLAH
KLGRTIGIFWLNASETLVEINTEPAVEYTLTQMGPVAAKQKVRSRTHVHWMSESGIIDVFLLTGPTPS
DVFKQYSHLTGTQAMPPLFSLGYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYFTW
DKNRFPNPKRMQELLRSKKRKLVVISDPHIKIEPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGLSSY
LDFTNPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHNIYGF
YHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYGAVWTGDNTAEWSNLKISIPMLLTLSITGISFCG
ADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEEHTRLIREAIRERYGLLPYWYS
LFYHAHVASQPVMRPLWVEFPDELKTFDMEDEYMLGSALLVHPVTEPKATTVDVFLPGSNEVWYDYKT
FAHWEGGCTVKIPVALDTIPVFQRGGSVIPIKTTVGKSTGWMTESSYGLRVALSTQGSSVGELYLDDG
HSFQYLHQKQFLHRKFSFCSSVLINSFADQRGHYPSKCVVEKILVLGFRKEPSSVTTHSSDGKDQPVA
FTYCAKTSILSLEKLSLNIATDWEVRII

NOV44g, CG55752-05 DNA Sequence	SEQ ID NO: 671	2951 bp
	ORF Start: ATG at 210	ORF Stop: TGA at 2688

TACTGGTTGTAATTTTAGAAAGACACCCAATCGGCTTTTTTAAAAGATCGCCCAGGGCCCTTGTCCTG
AGAGCTGGGAGCTGGTCGGAGTGACAGAGAAGCCATGGAAGCAGCAGTGAAAGAGGAAATAAGTGTTG
AAGATGAAGCTGTAGATAAAAACATTTTCAGAGACTGTAACAAGATCGCATTTTACAGGAAGAAAAGC
GTATTATGTTAATTCTGAACAGGCGTCAGAAACAGTGGCTTTCCAAGAAGTCCACCTATCGGGCATTA
TTGGATTCAGTCACAACAGATGAAGACAGCACCAGGTTCCAAATCATCAATGAAGCAAGTAAGGTTCC
TCTCCTGGCTGAAATTTATGGTATAGAAGGAAACATTTTCAGGCTTAAAATTAACGAAGAGACTCCTC
TAAAACCCAGATTTGAAGTTCCGGATGTCCTCACAAGCAAGCCAAGCACTGTAAGAGCTGCTAAAGAA
AATGAGGAGGAGACATCAGTGGACACCTCTCAGGAAAATCAAGAAGATCTGGGCCTGTGGGAAGAGAA
ATTTGGAAAATTTGTGGATATCACAGCTAATGGCCCTTCTTCTATTGGTTTGGATTTCTCCTTGCATG
GATTTGAGCATCTTTATGGGATCCCACAACATGCAGAATCACACCAACTTAAAAATACTGGAGATGCT
TACCGTCTTTATAACCTGGATGTCTATGGATACCAAATATATGATAAAATGGGCATTTATGGTTCAGT
ACCTTATCTCCTGGCCCACAAACTGGGCAGAACTATAGGTATTTTCTGGCTGAATGCCTCGGAAACAC
TGGTGGAGATCAATACAGAGCCTGCAGTAGAGTACACACTGACCCAGATGGGCCCAGTTGCTGCTAAA
CAAAAGGTCAGATCTCGCACTCATGTGCACTGGATGTCAGAGAGTGGCATCATTGATGTTTTTCTGCT
GACAGGACCTACACCTTCTGATGTCTTCAAACAGTACTCACACCTTACAGGTACGCAAGCCATGCCCC
CTCTTTTCTCTTTGGGATACCACCAGTGCCGCTGGAACTATGAAGATGAGCAGGATGTAAAAGCAGTG
GATGCAGGGTTTGATGAGCATGACATTCCTTATGATGCCATGTGGCTGGACATAGAGCACACTGAGGG
CAAGAGGTACTTCACCTGGGACAAAAACAGATTCCCAAACCCCAAGAGGATGCAAGAGCTGCTCAGGA
GCAAAAAGCGTAAGCTTGTGGTCATCAGTGATCCCCACATCAAGATTGAACCTGACTACTCAGTATAT
GTGAAGGCCAAAGATCAGGGCTTCTTTGTGAAGAATCAGGAAGGGGAAGACTTTGAAGGGGTGTGTTG
GCCAGGTCTCTCCTCTTACCTGGATTTCACCAATCCCAAGGTCAGAGAGTGGTATTCAAGTCTTTTTG
CTTTCCCTGTTTATCAGGGATCTACGGACATCCTCTTCCTTTGGAATGACATGAATGAGCCTTCTGTC
TTTAGAGGGCCAGAGCAAACCATGCAGAAGAATGCCATTCATCATGGCAATTGGGAGCACAGAGAGCT
CCACAACATCTACGGTTTTTATCATCAAATGGCTACTGCAGAAGGACTGATAAAACGATCTAAAGGGA
AGGAGAGACCCTTTGTTCTTACACGTTCTTTCTTTGCTGGATCACAAAAGTATGGTGCCGTGTGGACA
GGCGACAACACAGCAGAATGGAGCAACTTGAAAATTTCTATCCCAATGTTACTCACTCTCAGCATTAC
TGGGATCTCTTTTTGCGGAGCTGACATAGGCGGGTTCATTGGGAATCCAGAGACAGAGCTGCTAGTGC
GTTGGTACCAGGCTGGAGCCTACCAGCCCTTCTTCCGTGGCCATGCCACCATGAACACCAAGCGACGA
GAGCCCTGGCTCTTTGGGGAGGAACACACCCGACTCATCCGAGAAGCCATCAGAGAGCGCTATGGCCT
CCTGCCATATTGGTATTCTCTGTTCTACCATGCACACGTGGCTTCCCAACCTGTCATGAGGCCTCTGT
GGGTAGAGTTCCCTGATGAACTAAAGACTTTTGATATGGAAGATGAATACATGTTAGGGAGTGCATTA
TTGGTTCATCCAGTCACAGAACCAAAAGCCACCACAGTTGATGTGTTTCTTCCAGGATCAAATGAGGT

ATGGTATGACTATAAGACATTTGCTCATTGGGAAGGAGGGTGTACTGTAAAGATCCCAGTAGCCTTGG
ACACTATTCCAGTGTTTCAGCGAGGTGGAAGTGTGATACCAATAAAGACAACTGTAGGAAAATCCACA
GGCTGGATGACTGAATCCTCCTATGGACTCCGGGTTGCTCTAAGCACTCAGGGTTCTTCAGTGGGTGA
GTTATATCTTGATGATGGCCATTCATTCCAATACCTCCACCAGAAGCAATTTTTGCACAGGAAGTTTT
CATTCTGTTCCAGTGTTCTGATCAATAGTTTTGCTGACCAGAGGGGTCATTATCCCAGCAAGTGTGTG
GTGGAGAAGATCTTGGTCTTAGGCTTCAGGAAGGAGCCATCTTCTGTGACTACCCACTCATCTGATGG
TAAAGATCAGCCTGTGGCTTTTACGTATTGTGCCAAAACATCCATCCTGAGCCTGGAGAAGCTCTCAC
TCAACATTGCCACTGACTGGGAGGTCCGCATCATATGACAAAGAACTGCCCCTGGTGATGTGAGCAGG
GACCTGCCTGCCCCTTTCAACCTTTCCCCTCACCTTTTTTGAGATTTTTGCTGCAATCTGTTTGTCTT
CCCTGAATCAAAATAATCTTTCATTCGTCACCATTATACTAATGAACAATAGATTTCATGTTTCAAAA
TTTCAGATTTTACATGTTAAGATGTACTAACAATATTCCTTGTATCAAACATCTCCTTTTCTCCCTGA
TACATAGCCCTGAGACATTATAGCGTC

NOV44g, CG55752-05 Protein Sequence	SEQ ID NO: 672	826 aa	MW at 94610.4kD

MLILNRRQKQWLSKKSTYRALLDSVTTDEDSTRFQIINEASKVPLLAEIYGIEGNIFRLKINEETPLK
PRFEVPDVLTSKPSTVRAAKENEEETSVDTSQENQEDLGLWEEKFGKFVDITANGPSSIGLDFSLHGF
EHLYGIPQHAESHQLKNTGDAYRLYNLDVYGYQIYDKMGIYGSVPYLLAHKLGRTIGIFWLNASETLV
EINTEPAVEYTLTQMGPVAAKQKVRSRTHVHWMSESGIIDVFLLTGPTPSDVFKQYSHLTGTQAMPPL
FSLGYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYFTWDKNRFPNPKRMQELLRSK
KRKLVVISDPHIKIEPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGLSSYLDFTNPKVREWYSSLFAF
PVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHNIYGFYHQMATAEGLIKRSKGKE
RPFVLTRSFFAGSQKYGAVWTGDNTAEWSNLKISIPMLLTLSITGISFCGADIGGFIGNPETELLVRW
YQAGAYQPFFRGHATMNTKRREPWLFGEEHTRLIREAIRERYGLLPYWYSLFYHAHVASQPVMRPLWV
EFPDELKTFDMEDEYMLGSALLVHPVTEPKATTVDVFLPGSNEVWYDYKTFAHWEGGCTVKIPVALDT
IPVFQRGGSVIPIKTTVGKSTGWMTESSYGLRVALSTQGSSVGELYLDDGHSFQYLHQKQFLHRKFSF
CSSVLINSFADQRGHYPSKCVVEKILVLGFRKEPSSVTTHSSDGKDQPVAFTYCAKTSILSLEKLSLN
IATDWEVRII

NOV44h, SNP13379656 of CG55752-06, DNA Sequence	SEQ ID NO: 673	2001 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1972
	SNP Pos: 617	SNP Change: A to G

ATGGGCATTTATGGTTCAGTACCTTATCTCCTGGCCCACAAACTGGGCAGAACTATAGGTATTTTCTG
GCTGAATGCCTCGGAAACACTGGTGGAGATCAATACAGAGCCTGCAGTAGAGTACACACTGACCCAGA
TGGGCCCAGTTGCTGCTAAACAAAAGGTCAGATCTCGCACTCATGTGCACTGGATGTCAGAGAGTGGC
ATCATTGATGTTTTTCTGCTGACAGGACCTACACCTTCTGATGTCTTCAAACAGTACTCACACCTTAC
AGGTATTCAAGCCATGCCCCCTCTTTTCTCTTTGGGATACCACCAGTGCCGCTGGAACTATGAAGATG
AGCAGGATGTAAAAGCAGTGGATGCAGGGTTTGATGAGCATGACATTCCTTATGATGCCATGTGGCTG
GACATAGAGCACACTGAGGGCAAGAGGTACTTCACCTGGGACAAAAACAGATTCCCAAACCCCAAGAG
GATGCAAGAGCTGCTCAGGAGCAAAAAGCGTAAGGTACTTGTGGTCATCAGTGATCCCCACATCAAGA
TTGATCCTGACTACTCAGTATATGTGAAGGCCAAAGATCAGGGCTTCTTTGTGAAGAATCAGGAAGGG
GAAGGCTTTGAAGGGGTGTGTTGGCCAGGTCTCTCCTCTTACCTGGATTTCACCAATCCCAAGGTCAG
AGAGTGGTATTCAAGTCTTTTTGCTTTCCCTGTTTATCAGGGATCTACGGACATCCTCTTCCTTTGGA
ATGACATGAATGAGCCTTCTGTCTTTAGAGGGCCAGAGCAAACCATGCAGAAGAATGCCATTCATCAT
GGCAATTGGGAGCACAGAGAGCTCCACAACATCTACGGTTTTTATCATCAAATGGCTACTGCAGAAGG
ACTGATAAAACGATCTAAAGGGAAGGAGAGACCCTTTGTTCTTACACGTTCTTTCTTTGCTGGATCAC
AAAAGTATGGTGCCGTGTGGACAGGCGACAACACAGCAGAATGGAGCAACTTGAAAATTTCTATCCCA
ATGTTACTCACTCTCAGCATTACTGGGATCTCTTTTTGCGGAGCTGACATAGGCGGGTTCATTGGGAA
TCCAGAGACAGAGCTGCTAGTGCGTTGGTACCAGGCTGGAGCCTACCAGCCCTTCTTCCGTGGCCATG
CCACCATGAACACCAAGCGACGAGAGCCCTGGCTCTTTGGGGAGGAACACACCCGACTCATCCGAGAA
GCCATCAGAGAGCGCTATGGCCTCCTGCCATATTGGTATTCTCTGTTCTACCATGCACACGTGGCTTC
CCAACCTGTCATGAGGAGGCCTCTGTGGGTAGAGTTCCCTGATGAACTAAAGACTTTTGATATGGAAG
ATGAATACATGCTGGGTTTAGGGAGTGCATTATTGGTTCATCCAGTCACAGAACCAAAAGCCACCACA
GTTGATGTGTTTCTTCCAGGATCAAATGAGGTATGGTATGACTATAAGACATTTGCTCATTGGGAAGG
AGGGTGTACTGTAAAGATCCCAGTAGCCTTGGACACTATTCCAGTGTTTCAGCGAGGTGGAAGTGTGA
TACCAATAAAGACAACTGTAGGAAAATCCACAGGCTGGATGACTGAATCCTCCTATGGACTCCGGGTT
GCTCTAAGCACTGGTTCTTCAGTGGGTGAGTTATATCTTGATGATGGCCATTCATTCCAATACCTCCA
CCAGAAGCAATTTTTGCACAGGAAGTTTTCATTCTGTTCCAGTGTTCTGATCAATTCCAGTTTTGCTG

ACCAGAGGGGTCATTATCCCAGCAAGTGTGTGGTGGAGAAGATCTTGGTCTTAGGCTTCAGGAAGGAG
CCATCTTCTGTGACTACCCACTCATCTGATGGTAAAGATCAGCCTGTGGCTTTTACGTATTGTGCCAA
AACATCCATCCTGAGCCTGGAGAAGCTCTCACTCAACATTGCCACTGACTGGGAGGTCCGCATCATAT
GACAAAGAACTGCCCCTGGTGATGTGAGC

NOV44h, SNP13379656 of CG55752-06, Protein Sequence	SEQ ID NO: 674	657 aa	MW at 75108.9kD
	SNP Pos: 206		SNP Change: Asp to Gly

MGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAVEYTLTQMGPVAAKQKVRSRTHVHWMSESG
IIDVFLLTGPTPSDVFKQYSHLTGIQAMPPLFSLGYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWL
DIEHTEGKRYFTWDKNRFPNPKRMQELLRSKKRKVLVVISDPHIKIDPDYSVYVKAKDQGFFVKNQEG
EGFEGVCWPGLSSYLDFTNPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHH
GNWEHRELHNIYGFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYGAVWTGDNTAEWSNLKISIP
MLLTLSITGISFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEEHTRLIRE
AIRERYGLLPYWYSLFYHAHVASQPVMRRPLWVEFPDELKTFDMEDEYMLGLGSALLVHPVTEPKATT
VDVFLPGSNEVWYDYKTFAHWEGGCTVKIPVALDTIPVFQRGGSVIPIKTTVGKSTGWMTESSYGLRV
ALSTGSSVGELYLDDGHSFQYLHQKQFLHRKFSFCSSVLINSSFADQRGHYPSKCVVEKILVLGFRKE
PSSVTTHSSDGKDQPVAFTYCAKTSILSLEKLSLNIATDWEVRII

NOV44i, SNP13379655 of CG55752-06, DNA Sequence	SEQ ID NO: 675	2001 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1972
	SNP Pos: 1763	SNP Change: T to C

ATGGGCATTTATGGTTCAGTACCTTATCTCCTGGCCCACAAACTGGGCAGAACTATAGGTATTTTCTG
GCTGAATGCCTCGGAAACACTGGTGGAGATCAATACAGAGCCTGCAGTAGAGTACACACTGACCCAGA
TGGGCCCAGTTGCTGCTAAACAAAAGGTCAGATCTCGCACTCATGTGCACTGGATGTCAGAGAGTGGC
ATCATTGATGTTTTTCTGCTGACAGGACCTACACCTTCTGATGTCTTCAAACAGTACTCACACCTTAC
AGGTATTCAAGCCATGCCCCCTCTTTTCTCTTTGGGATACCACCAGTGCCGCTGGAACTATGAAGATG
AGCAGGATGTAAAAGCAGTGGATGCAGGGTTTGATGAGCATGACATTCCTTATGATGCCATGTGGCTG
GACATAGAGCACACTGAGGGCAAGAGGTACTTCACCTGGGACAAAAACAGATTCCCAAACCCCAAGAG
GATGCAAGAGCTGCTCAGGAGCAAAAAGCGTAAGGTACTTGTGGTCATCAGTGATCCCCACATCAAGA
TTGATCCTGACTACTCAGTATATGTGAAGGCCAAAGATCAGGGCTTCTTTGTGAAGAATCAGGAAGGG
GAAGACTTTGAAGGGGTGTGTTGGCCAGGTCTCTCCTCTTACCTGGATTTCACCAATCCCAAGGTCAG
AGAGTGGTATTCAAGTCTTTTTGCTTTCCCTGTTTATCAGGGATCTACGGACATCCTCTTCCTTTGGA
ATGACATGAATGAGCCTTCTGTCTTTAGAGGGCCAGAGCAAACCATGCAGAAGAATGCCATTCATCAT
GGCAATTGGGAGCACAGAGAGCTCCACAACATCTACGGTTTTTATCATCAAATGGCTACTGCAGAAGG
ACTGATAAAACGATCTAAAGGGAAGGAGAGACCCTTTGTTCTTACACGTTCTTTCTTTGCTGGATCAC
AAAAGTATGGTGCCGTGTGGACAGGCGACAACACAGCAGAATGGAGCAACTTGAAAATTTCTATCCCA
ATGTTACTCACTCTCAGCATTACTGGGATCTCTTTTTGCGGAGCTGACATAGGCGGGTTCATTGGGAA
TCCAGAGACAGAGCTGCTAGTGCGTTGGTACCAGGCTGGAGCCTACCAGCCCTTCTTCCGTGGCCATG
CCACCATGAACACCAAGCGACGAGAGCCCTGGCTCTTTGGGGAGGAACACACCCGACTCATCCGAGAA
GCCATCAGAGAGCGCTATGGCCTCCTGCCATATTGGTATTCTCTGTTCTACCATGCACACGTGGCTTC
CCAACCTGTCATGAGGAGGCCTCTGTGGGTAGAGTTCCCTGATGAACTAAAGACTTTTGATATGGAAG
ATGAATACATGCTGGGTTTAGGGAGTGCATTATTGGTTCATCCAGTCACAGAACCAAAAGCCACCACA
GTTGATGTGTTTCTTCCAGGATCAAATGAGGTATGGTATGACTATAAGACATTTGCTCATTGGGAAGG
AGGGTGTACTGTAAAGATCCCAGTAGCCTTGGACACTATTCCAGTGTTTCAGCGAGGTGGAAGTGTGA
TACCAATAAAGACAACTGTAGGAAAATCCACAGGCTGGATGACTGAATCCTCCTATGGACTCCGGGTT
GCTCTAAGCACTGGTTCTTCAGTGGGTGAGTTATATCTTGATGATGGCCATTCATTCCAATACCTCCA
CCAGAAGCAATTTTTGCACAGGAAGTTTTCATTCTGTTCCAGTGTTCTGATCAATTCCAGTTCTGCTG
ACCAGAGGGGTCATTATCCCAGCAAGTGTGTGGTGGAGAAGATCTTGGTCTTAGGCTTCAGGAAGGAG
CCATCTTCTGTGACTACCCACTCATCTGATGGTAAAGATCAGCCTGTGGCTTTTACGTATTGTGCCAA
AACATCCATCCTGAGCCTGGAGAAGCTCTCACTCAACATTGCCACTGACTGGGAGGTCCGCATCATAT
GACAAAGAACTGCCCCTGGTGATGTGAGC

NOV44i, SNP13379655 of CG55752-06, Protein Sequence	SEQ ID NO: 676	657 aa	MW at 75106.8kD
	SNP Pos: 588		SNP Change: Phe to Ser

MGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAVEYTLTQMGPVAAKQKVRSRTHVHWMSESG
IIDVFLLTGPTPSDVFKQYSHLTGIQAMPPLFSLGYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWL
DIEHTEGKRYFTWDKNRFPNPKRMQELLRSKKRKVLVVISDPHIKIDPDYSVYVKAKDQGFFVKNQEG
EDFEGVCWPGLSSYLDFTNPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHH

GNWEHRELHNIYGFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYGAVWTGDNTAEWSNLKISIP
MLLTLSITGISFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEEHTRLIRE
AIRERYGLLPYWYSLFYHAHVASQPVMRRPLWVEFPDELKTFDMEDEYMLGLGSALLVHPVTEPKATT
VDVFLPGSNEVWYDYKTFAHWEGGCTVKIPVALDTIPVFQRGGSVIPIKTTVGKSTGWMTESSYGLRV
ALSTGSSVGELYLDDGHSFQYLHQKQFLHRKFSFCSSVLINSSSADQRGHYPSKCVVEKILVLGFRKE
PSSVTTHSSDGKDQPVAFTYCAKTSILSLEKLSLNIATDWEVRII

NOV44j, SNP13379654 of CG55752-06, DNA Sequence	SEQ ID NO: 677	2001 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 1972
	SNP Pos: 1772	SNP Change: A to G

ATGGGCATTTATGGTTCAGTACCTTATCTCCTGGCCCACAAACTGGGCAGAACTATAGGTATTTTCTG
GCTGAATGCCTCGGAAACACTGGTGGAGATCAATACAGAGCCTGCAGTAGAGTACACACTGACCCAGA
TGGGCCCAGTTGCTGCTAAACAAAAGGTCAGATCTCGCACTCATGTGCACTGGATGTCAGAGAGTGGC
ATCATTGATGTTTTTCTGCTGACAGGACCTACACCTTCTGATGTCTTCAAACAGTACTCACACCTTAC
AGGTATTCAAGCCATGCCCCCTCTTTTCTCTTTGGGATACCACCAGTGCCGCTGGAACTATGAAGATG
AGCAGGATGTAAAAGCAGTGGATGCAGGGTTTGATGAGCATGACATTCCTTATGATGCCATGTGGCTG
GACATAGAGCACACTGAGGGCAAGAGGTACTTCACCTGGGACAAAAACAGATTCCCAAACCCCAAGAG
GATGCAAGAGCTGCTCAGGAGCAAAAAGCGTAAGGTACTTGTGGTCATCAGTGATCCCCACATCAAGA
TTGATCCTGACTACTCAGTATATGTGAAGGCCAAAGATCAGGGCTTCTTTGTGAAGAATCAGGAAGGG
GAAGACTTTGAAGGGGTGTGTTGGCCAGGTCTCTCCTCTTACCTGGATTTCACCAATCCCAAGGTCAG
AGAGTGGTATTCAAGTCTTTTTGCTTTCCCTGTTTATCAGGGATCTACGGACATCCTCTTCCTTTGGA
ATGACATGAATGAGCCTTCTGTCTTTAGAGGGCCAGAGCAAACCATGCAGAAGAATGCCATTCATCAT
GGCAATTGGGAGCACAGAGAGCTCCACAACATCTACGGTTTTTATCATCAAATGGCTACTGCAGAAGG
ACTGATAAAACGATCTAAAGGGAAGGAGAGACCCTTTGTTCTTACACGTTCTTTCTTTGCTGGATCAC
AAAAGTATGGTGCCGTGTGGACAGGCGACAACACAGCAGAATGGAGCAACTTGAAAATTTCTATCCCA
ATGTTACTCACTCTCAGCATTACTGGGATCTCTTTTTGCGGAGCTGACATAGGCGGGTTCATTGGGAA
TCCAGAGACAGAGCTGCTAGTGCGTTGGTACCAGGCTGGAGCCTACCAGCCCTTCTTCCGTGGCCATG
CCACCATGAACACCAAGCGACGAGAGCCCTGGCTCTTTGGGGAGGAACACACCCGACTCATCCGAGAA
GCCATCAGAGAGCGCTATGGCCTCCTGCCATATTGGTATTCTCTGTTCTACCATGCACACGTGGCTTC
CCAACCTGTCATGAGGAGGCCTCTGTGGGTAGAGTTCCCTGATGAACTAAAGACTTTTGATATGGAAG
ATGAATACATGCTGGGTTTAGGGAGTGCATTATTGGTTCATCCAGTCACAGAACCAAAAGCCACCACA
GTTGATGTGTTTCTTCCAGGATCAAATGAGGTATGGTATGACTATAAGACATTTGCTCATTGGGAAGG
AGGGTGTACTGTAAAGATCCCAGTAGCCTTGGACACTATTCCAGTGTTTCAGCGAGGTGGAAGTGTGA
TACCAATAAAGACAACTGTAGGAAAATCCACAGGCTGGATGACTGAATCCTCCTATGGACTCCGGGTT
GCTCTAAGCACTGGTTCTTCAGTGGGTGAGTTATATCTTGATGATGGCCATTCATTCCAATACCTCCA
CCAGAAGCAATTTTTGCACAGGAAGTTTTCATTCTGTTCCAGTGTTCTGATCAATTCCAGTTTTGCTG
ACCGGAGGGGTCATTATCCCAGCAAGTGTGTGGTGGAGAAGATCTTGGTCTTAGGCTTCAGGAAGGAG
CCATCTTCTGTGACTACCCACTCATCTGATGGTAAAGATCAGCCTGTGGCTTTTACGTATTGTGCCAA
AACATCCATCCTGAGCCTGGAGAAGCTCTCACTCAACATTGCCACTGACTGGGAGGTCCGCATCATAT
GACAAAGAACTGCCCCTGGTGATGTGAGC

NOV44j, SNP13379654 of CG55752-06, Protein Sequence	SEQ ID NO: 678	657 aa	MW at 75194.9kD
	SNP Pos: 591		SNP Change: Gln to Arg

MGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAVEYTLTQMGPVAAKQKVRSRTHVHWMSESG
IIDVFLLTGPTPSDVFKQYSHLTGIQAMPPLFSLGYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWL
DIEHTEGKRYFTWDKNRFPNPKRMQELLRSKKRKVLVVISDPHIKIDPDYSVYVKAKDQGFFVKNQEG
EDFEGVCWPGLSSYLDFTNPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHH
GNWEHRELHNIYGFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYGAVWTGDNTAEWSNLKISIP
MLLTLSITGISFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEEHTRLIRE
AIRERYGLLPYWYSLFYHAHVASQPVMRRPLWVEFPDELKTFDMEDEYMLGLGSALLVHPVTEPKATT
VDVFLPGSNEVWYDYKTFAHWEGGCTVKIPVALDTIPVFQRGGSVIPIKTTVGKSTGWMTESSYGLRV
ALSTGSSVGELYLDDGHSFQYLHQKQFLHRKFSFCSSVLINSSFADRRGHYPSKCVVEKILVLGFRKE
PSSVTTHSSDGKDQPVAFTYCAKTSILSLEKLSLNIATDWEVRII

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 44B.

Table 44B. Comparison of the NOV44 protein sequences.

```
NOV44a  ------------------------------------------------------------
NOV44b  ------------------------------------------------------------
NOV44c  MRAAVAGIAFRRRRQKQWLSKKSTYQALLDSVTTDEDSTRFQIINEASKVRRQKQWLSKK
NOV44d  MEAAVKEEIS-----------------VEDEAVDKN--IFRDCNKIAFYRRQKQWLSKK
NOV44e  MEAAVKEEIS-----------------VEDEAVDKN--IFRDCNKIAFYRRQKQWLSKK
NOV44f  MEAAVKEEIS-----------------VEDEAVDKN--IFRDCNKIAFYRRQKQWLSKK
NOV44g  ------------------------------------------MLILNRRQKQWLSKK

NOV44a  ------------------------------------------------------------
NOV44b  ------------------------------------------------------------
NOV44c  STYQALLDSVTTDEDSTRFQIINEASKVPLLAEIYGIEGNIFRLKINEETPLKPRFEVPD
NOV44d  STYQALLDSVTTDEDSTRFQIINEASKVPLLAEIYGIEGNIFRLKINEETPLKPRFEVPD
NOV44e  STYQALLDSVTTDEDSTRFQIINEASKVPLLAEIYGIEGNIFRLKINEETPLKPRFEVPD
NOV44f  STYRALLDSVTTDEDSTRFQIINEASKVPLLAEIYGIEGNIFRLKINEETPLKPRFEVPD
NOV44g  STYRALLDSVTTDEDSTRFQIINEASKVPLLAEIYGIEGNIFRLKINEETPLKPRFEVPD

NOV44a  ------------------------------------------------------------
NOV44b  ------------------------------------------------------------
NOV44c  VLTSKPSTVR-ISCSGDTGSLILADGKGDLKCHITANPFKVDLVSEEEVVISINSLGQLY
NOV44d  VLTSKPSTVRLISCSGDTGSLILADGKGDLKCHITANPFKVDLVSEEEVVISINSLGQLY
NOV44e  VLTSKPSTVRLISCSGDTGSLILADGKGDLKCHITANPFKVDLVSEEEVVISINSLGQLY
NOV44f  VLTSKPSTVRLISCSGDTGSLILADGKGDLKCHITANPFKVDLVSEEEVVISINSLGQLY
NOV44g  VLTSKP------------------------------------------------------

NOV44a  ------------------------------------------------------------
NOV44b  ------------------------------------------------------------
NOV44c  FEHGRAPRVSFSDKVNLTLGSIWDKIKNLFSRQGSKDPAEGDGAQPEETPRDGDKPEETQ
NOV44d  FEH----------------------LQILHKQRAAKENEE-------ETSVDTS--QENQ
NOV44e  FEH----------------------LQILHKQRAAKENEE-------ETSVDTS--QENQ
NOV44f  FEH----------------------LQILHKQRAAKENEE-------ETSVDTS--QENQ
NOV44g  --------------------------S--TVRAAKENEE-------ETSVDTS--QENQ

NOV44a  ------------------------------------------------------------
NOV44b  ------------------------------------------------------------
NOV44c  GKAEKDEPGAWEETFKTHSDSKPYGPSSIGLDFSLHGFEHLYGIPQHAESHQLKNTGDGD
NOV44d  -----EDLGLWEEKFGKFVDIKANGPSSIGLDFSLHGFEHLYGIPQHAESHQLKNTGDGD
NOV44e  -----EDLGLWEEKFGKFVDIKANGPSSIGLDFSLHGFEHLYGIPQHAESHQLKNTGDGD
NOV44f  -----EDLGLWEEKFGKFVDIKANGPSSIGLDFSLHGFEHLYGIPQHAESHQLKNTG--D
NOV44g  -----EDLGLWEEKFGKFVDITANGPSSIGLDFSLHGFEHLYGIPQHAESHQLKNTG--D

NOV44a  -----------------MGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAVEYTL
NOV44b  ----------------MGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAVEYTL
NOV44c  AYRLYNLDVYGYQIYDKMGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAGIVIF
NOV44d  AYRLYNLDVYGYQIYDKMGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAVEYTL
NOV44e  AYRLYNLDVYGYQIYDKMGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAVEYTL
NOV44f  AYRLYNLDVYGYQIYDKMGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAVEYTL
NOV44g  AYRLYNLDVYGYQIYDKMGIYGSVPYLLAHKLGRTIGIFWLNASETLVEINTEPAVEYTL

NOV44a  TQMGPVAAKQK-VRSRTHVHWMSESGIIDVFLLTGPTPSDVFKQYSHLTGTQAMPPLFSL
NOV44b  TQMGPVAAKQK-VRSRTHVHWMSESGIIDVFLLTGPTPSDVFKQYSHLTGIQAMPPLFSL
NOV44c  GPVSLIYQSQGDTPLTTHVHWMSESGIIDVFLLTGPTPSDVFKQYSHLTGTQAMPPLFSL
NOV44d  TQMGPVAAKQK-VRSRTHVHWMSESGIIDVFLLTGPTPSDVFKQYSHLTGTQAMPPLFSL
```

```
NOV44e  TQMGPVAAKQK-VGSRTHVHWMSESGIIDVFLLTGPTPSDVFKQYSHLTGTQAMPPLFSL
NOV44f  TQMGPVAAKQK-VRSRTHVHWMSESGIIDVFLLTGPTPSDVFKQYSHLTGTQAMPPLFSL
NOV44g  TQMGPVAAKQK-VRSRTHVHWMSESGIIDVFLLTGPTPSDVFKQYSHLTGTQAMPPLFSL

NOV44a  GYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYFTWDKNRFPNPKRMQE
NOV44b  GYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYFTWDKNRFPNPKRMQE
NOV44c  GYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYFTWDKNRFPNPKRMQE
NOV44d  GYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYFTWDKNRFPNPKRMQE
NOV44e  GYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYFTWDKNRFPNPKRMQE
NOV44f  GYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYFTWDKNRFPNPKRMQE
NOV44g  GYHQCRWNYEDEQDVKAVDAGFDEHDIPYDAMWLDIEHTEGKRYFTWDKNRFPNPKRMQE

NOV44a  LLRSKKR-KLVVISDPHIKIDPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGLSSYLDFT
NOV44b  LLRSKKRKVLVVISDPHIKIDPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGLSSYLDFT
NOV44c  LLRSKKR-KLVVISDPHIKIEPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGMKSYLDFT
NOV44d  LLRSKKR-KLVVISDPHIKIDPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGLSSYLDFT
NOV44e  LLRSKKR-KLVVISDPHIKIDPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGLSSYLDFT
NOV44f  LLRSKKR-KLVVISDPHIKIEPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGLSSYLDFT
NOV44g  LLRSKKR-KLVVISDPHIKIEPDYSVYVKAKDQGFFVKNQEGEDFEGVCWPGLSSYLDFT

NOV44a  NPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHN
NOV44b  NPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHN
NOV44c  NPKVREWYSSMFSS-NCDGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHN
NOV44d  NPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHN
NOV44e  NPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHN
NOV44f  NPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHN
NOV44g  NPKVREWYSSLFAFPVYQGSTDILFLWNDMNEPSVFRGPEQTMQKNAIHHGNWEHRELHN

NOV44a  IYGFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYGGAVWTGDNTAEWSNLKISIPM
NOV44b  IYGFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYG-AVWTGDNTAEWSNLKISIPM
NOV44c  IYGFY--MATAEGLIKRSKGKERPFVLTRSFFAGSQKYG-AVWTGDNTAEWSNLKISIPM
NOV44d  IYGFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYG-AVWTGDNTAEWSNLKISIPM
NOV44e  IYGFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYG-AVWTGDNTAEWSNLKISIPM
NOV44f  IYGFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYG-AVWTGDNTAEWSNLKISIPM
NOV44g  IYGFYHQMATAEGLIKRSKGKERPFVLTRSFFAGSQKYG-AVWTGDNTAEWSNLKISIPM

NOV44a  LLTLSITGISFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEE
NOV44b  LLTLSITGISFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEE
NOV44c  LLTLSITGISFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEE
NOV44d  LLTLSITGISFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEE
NOV44e  LLTLSITGVSFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEE
NOV44f  LLTLSITGISFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEE
NOV44g  LLTLSITGISFCGADIGGFIGNPETELLVRWYQAGAYQPFFRGHATMNTKRREPWLFGEE

NOV44a  HTRLIREAIRERYGLLPYWYSLFYHAHVASQPVMR-PLWVEFPDELKTFDMEDEYMLG--
NOV44b  HTRLIREAIRERYGLLPYWYSLFYHAHVASQPVMRRPLWVEFPDELKTFDMEDEYMLGLG
NOV44c  HTRLIREAIRERYGLLPYWYSLFYHAHVASQPVMR-PLWVEFPDELKTFDMEDEYMLG--
NOV44d  HTRLIREAIRERYGLLPYWYSLFYHAHVASQPVMR-PLWVEFPDELKTFDMEDEYMLG--
NOV44e  HTRLIREAIRERYGLLPYWYSLFYHAHVASQPVMR-PLWVEFPDELKTFDMEDEYMLG--
NOV44f  HTRLIREAIRERYGLLPYWYSLFYHAHVASQPVMR-PLWVEFPDELKTFDMEDEYMLG--
NOV44g  HTRLIREAIRERYGLLPYWYSLFYHAHVASQPVMR-PLWVEFPDELKTFDMEDEYMLG--

NOV44a  SALLVHPVTEPKATTVDVFLPGSNE-VWYDYKTFAHWEGGCTVKIPVALDTIPVFQRGGS
NOV44b  SALLVHPVTEPKATTVDVFLPGSNE-VWYDYKTFAHWEGGCTVKIPVALDTIPVFQRGGS
```

```
NOV44c  SALLVHPVTEPKATTVDVFLPGSNEVVWYDYKTFAHWEGGCTVKIPVLLQ-IPVFQRGGS
NOV44d  SALLVHPVTEPKATTVDVFLPGSNE-VWYDYKTFAHWEGGCTVKIPVALDTIPVFQRGGS
NOV44e  SALLVHPVTEPKATTVDVFLPGSNE-VWYDYKTFAHWEGGCTVKIPVALDTIPVFQRGGS
NOV44f  SALLVHPVTEPKATTVDVFLPGSNE-VWYDYKTFAHWEGGCTVKIPVALDTIPVFQRGGS
NOV44g  SALLVHPVTEPKATTVDVFLPGSNE-VWYDYKTFAHWEGGCTVKIPVALDTIPVFQRGGS

NOV44a  VIPIKTTVGKSTGWMTESSYGLRVALSTK-GSSVGELYLDDGHSFQYLHQKQFLHRKFSF
NOV44b  VIPIKTTVGKSTGWMTESSYGLRVALST--GSSVGELYLDDGHSFQYLHQKQFLHRKFSF
NOV44c  VIPIKTTVGKSTGWMTESSYGLRVALSTLQGSSVGELYLDDGHSFQYLHQKQFLHRKFSF
NOV44d  VIPIKTTVGKSTGWMTESSYGLRVALSTK-GSSVGELYLDDGHSFQYLHQKQFLHRKFSF
NOV44e  VIPIKTTVGKSTGWMTESSYGLRVALSTK-GSSVGELYLDDGHSFQYLHQKQFLHRKFSF
NOV44f  VIPIKTTVGKSTGWMTESSYGLRVALSTQ-GSSVGELYLDDGHSFQYLHQKQFLHRKFSF
NOV44g  VIPIKTTVGKSTGWMTESSYGLRVALSTQ-GSSVGELYLDDGHSFQYLHQKQFLHRKFSF

NOV44a  CSSVLINSSFADQRGHYPSKCVVEKILVLGFRKEPSSVTTHSSGDGKDQPVAFTYCAKTS
NOV44b  CSSVLINSSFADQRGHYPSKCVVEKILVLGFRKEPSSVTTHSS-DGKDQPVAFTYCAKTS
NOV44c  CSSVLVA-SSPVSQGHLHTPLSMTKALLFTVSS-PASVKMRLH-YSPEKRARFSHCAKTS
NOV44d  CSSVLIN-SFADQRGHYPSKCVVEKILVLGFRKEPSSVTTHSS-DGKDQPVAFTYCAKTS
NOV44e  CSSVLIN-SFADQRGHYPSKCVVEKILVLGFRKEPSSVTTHSS-DGKDQPVAFTYCAKTS
NOV44f  CSSVLIN-SFADQRGHYPSKCVVEKILVLGFRKEPSSVTTHSS-DGKDQPVAFTYCAKTS
NOV44g  CSSVLIN-SFADQRGHYPSKCVVEKILVLGFRKEPSSVTTHSS-DGKDQPVAFTYCAKTS

NOV44a  ILSLEKLSLNIATDWEVRII
NOV44b  ILSLEKLSLNIATDWEVRII
NOV44c  ILSLEKLSLNIATDWEVRII
NOV44d  ILSLEKLSLNIATDWEVRII
NOV44e  ILSLEKLSLNIATDWEVRII
NOV44f  ILSLEKLSLNIATDWEVRII
NOV44g  ILSLEKLSLNIATDWEVRII

NOV44a  (SEQ ID NO:  660)
NOV44b  (SEQ ID NO:  662)
NOV44c  (SEQ ID NO:  664)
NOV44d  (SEQ ID NO:  666)
NOV44e  (SEQ ID NO:  668)
NOV44f  (SEQ ID NO:  670)
NOV44g  (SEQ ID NO:  672)
```

Further analysis of the NOV44a protein yielded the following properties shown in Table 44C.

Table 44C. Protein Sequence Properties NOV44a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 13;  peak value   5.66
      PSG score:   1.26

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -7.02
```

```
      possible cleavage site: between 51 and 52

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -2.07   Transmembrane  337 - 353
      PERIPHERAL  Likelihood =  3.82 (at 496)
      ALOM score:  -2.07  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 344
      Charge difference: -2.0   C(-3.0) - N(-1.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 2   (cytoplasmic tail 1 to 337)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  2.91
      Hyd Moment(95):  0.96    G content:       4
      D/E content:     1       S/T content:     3
      Score: -6.13

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 27  GRT|IG

NUCDISC: discrimination of nuclear localization signals
      pat4: KKRK (5) at  167
      pat7: none
      bipartite: none
      content of basic residues:  10.4%
      NLS Score: -0.16

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : MGIYGSV

Prenylation motif: none
```

```
memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: too long tail

Dileucine motif in the tail: found
      LL at 10
      LL at 74
      LL at 163

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

-------------------------

Final Results (k = 9/23):

       30.4 %: cytoplasmic
       30.4 %: mitochondrial
       17.4 %: Golgi
        8.7 %: endoplasmic reticulum
        4.3 %: extracellular, including cell wall
        4.3 %: nuclear
        4.3 %: vesicles of secretory system

>> prediction for CG55752-07 is cyt (k=23)
```

A search of the NOV44a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 44D.

Table 44D. Geneseq Results for NOV44a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV44a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABG69611	Human NOV11b protein - Homo sapiens, 914 aa. [WO200250277-A2, 27-JUN-2002]	1..656 262..914	653/656 (99%) 653/656 (99%)	0.0
ABG69613	Human NOV11d protein - Homo sapiens, 912 aa. [WO200250277-A2, 27-JUN-2002]	1..656 260..912	651/656 (99%) 653/656 (99%)	0.0
ABG69612	Human NOV11c protein - Homo sapiens, 914 aa. [WO200250277-A2, 27-JUN-2002]	1..656 262..914	651/656 (99%) 652/656 (99%)	0.0
AAG79779	Carbohydrate-associated protein (CHOP)-1 - Homo sapiens, 912 aa. [WO200297060-A2, 05-DEC-2002]	1..652 262..910	648/652 (99%) 648/652 (99%)	0.0
ABP52437	Human carbohydrate metabolism enzyme CME-2 protein SEQ ID NO:2 - Homo sapiens, 914 aa. [WO200261059-A2, 08-AUG-2002]	1..656 262..914	634/656 (96%) 636/656 (96%)	0.0

In a BLAST search of public sequence databases, the NOV44a protein was found to have homology to the proteins shown in the BLASTP data in Table 44E.

Table 44E. Public BLASTP Results for NOV44a				
Protein Accession Number	**Protein/Organism/Length**	**NOV44a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8IWZ0	Glucosidase - Homo sapiens (Human), 769 aa (fragment).	1..656 117..769	653/656 (99%) 653/656 (99%)	0.0
Q8TET4	FLJ00088 protein - Homo sapiens (Human), 925 aa (fragment).	1..656 273..925	653/656 (99%) 653/656 (99%)	0.0
Q8IZM5	Neutral alpha glucosidase C type 2 - Homo sapiens (Human), 914 aa.	1..656 262..914	653/656 (99%) 653/656 (99%)	0.0
AAN74755	Neutral alpha glucosidase C - synthetic construct, 914 aa.	1..656 262..914	652/656 (99%) 653/656 (99%)	0.0
Q8IZM4	Neutral alpha-glucosidase C type 3 - Homo sapiens (Human), 914 aa.	1..656 262..914	652/656 (99%) 653/656 (99%)	0.0

PFam analysis indicates that the NOV44a protein contains the domains shown in the Table 44F.

Table 44F. Domain Analysis of NOV44a			
Pfam Domain	**NOV44a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Glyco_hydro_31	1..616	259/856 (30%) 517/856 (60%)	3.4e-235

Example 45.

The NOV45 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 45A.

Table 45A. NOV45 Sequence Analysis		
NOV45a, CG55778-03 DNA Sequence	SEQ ID NO: 679	752 bp
	ORF Start: ATG at 23	ORF Stop: TAA at 653
GGCGGGGCGGCCGGCGGCGGCC**ATG**GGAGATATCCCAGCCGTGGGCCTCAGCTCCTGGAAGCAGGCTT CTCCAGGTAAAGTGACCGAGGCAGTGAAAGAGGCCATTGACGCAGGGTACCGGCACTTCGACTGTGCT TACTTTTACCACAATGAGAGGGAGGTTGGAGCAGGGATCCGTTGCAAGATCAAGGAAGGCGCTGTAAG ACGGGAGGATCTGTTCATTGCCACTAAGCTGTGGTGCACCTGCCATAAGAAGTCCTTGGTGGAAACAG CATGCAGAAAGAGTCTCAAGGCCTTGAAGCTGAACTATTTGGACCTCTACCTCATACACTGGCCCATG		

GGTTTCAAGCCTCCTCATCCAGAATGGATCATGAGCTGCAGTGAACTTTCCTTCTGCCTCTCACATCC
TCGAGTGCAGGACTTGCCTCTGGACGAGAGCAACATGGTTATTCCCAGTGACACGGACTTCCTGGACA
CGTGGGAGATTTTGATCCGATTTCAAATCCAGAGGAATGTGATAGTGATCCCCGGATCTATCACCCCA
AGTCACATTAAAGAGAATATCCAGGTGTTTGATTTTGAATTAACACAGCACGATATGGATAACATCCT
CAGCCTAAACAAGAATCTCCGACTGGCCATGTTCCCCATGTAAATATGGCTCCTTCTTTTTAAAACAG
AGGGAAGAATATACAGATTGAATGATTGGTGTCTGAATAGAACTAAAAATCACAAAGACTATCCTTTC
CACA

NOV45a, CG55778-03 Protein Sequence	SEQ ID NO: 680	210 aa	MW at 24136.7kD

MGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKIKEGAVRREDLFIA
TKLWCTCHKKSLVETACRKSLKALKLNYLDLYLIHWPMGFKPPHPEWIMSCSELSFCLSHPRVQDLPL
DESNMVIPSDTDFLDTWEILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNILSLNKNLR
LAMFPM

NOV45b, CG55778-06 DNA Sequence	SEQ ID NO: 681	569 bp
	ORF Start: ATG at 1	ORF Stop: TGA at 550

ATGGGAGATATCCCAGCCGTGGGCCTCAGCTCCTGGAAGGCTTCTCCAGGAAAAGTGACCGAGGCAGT
GAAAGAGGCCATTGACGCAGGGTACCGGCACTTCGACTGTGCTTACTTTTACCACAATGAGAGGGAGG
TTGGAGCAGGGATCCGTTGCAAGATCAAGGAAGGCGCTGTAAGACGGGAGGATCTGTTCATTGCCACT
AAGCCTCCTCATCCAGAATGGATCATGAGCTGCAGTGAACTTTCCTTCTGCCTCTCACATCCTCGAGT
GCAGGACTTGCCTCTGGACGAGAGCAACATGGTTATTCCCAGTGACACGGACTTCCTGGACACGTGGG
AGATTTTGATCCGATTTCAAATCCAGAGGAATGTGATAGTGATCCCCGGATCTATCACCCCAAGTCAC
ATTAAAGAGAATATCCAGGTGTTTGATTTTGAATTAACACAGCACGATATGGATAACATCCTCAGCCT
AAACAGGAATCTCCGACTGGCCATGTTCCCCATAACTAAAAATCACAAAGACTATCCTTTCCACATAG
AATACTGAGGACGCTTCCCCTTCCT

NOV45b, CG55778-06 Protein Sequence	SEQ ID NO: 682	183 aa	MW at 21057.8kD

MGDIPAVGLSSWKASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKIKEGAVRREDLFIAT
KPPHPEWIMSCSELSFCLSHPRVQDLPLDESNMVIPSDTDFLDTWEILIRFQIQRNVIVIPGSITPSH
IKENIQVFDFELTQHDMDNILSLNRNLRLAMFPITKNHKDYPFHIEY

NOV45c, 275480984 DNA Sequence	SEQ ID NO: 683	655 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCAGATCTCCCACCATGGGAGATATCCCAGCCGTGGGCCTCAGCTCCTGGAAGCAGGCTTCTCCAG
GGAAAGTGACCGAGGCAGTGAAAGAGGCCATTGACGCAGGGTACCGGCACTTCGACTGTGCTTACTTT
TACCACAATGAGAGGGAGGTTGGAGCAGGGATCCGTTGCAAGATCAAGGAAGGCGCTGTAAGACGGGA
GGATCTGTTCATTGCCACTAAGCTGTGGTGCACCTGCCATAAGAAGTCCTTGGTGGAAACAGCATGCA
GAAAGAGTCTCAAGGCCTTGAAGCTGAACTATTTGGACCTCTACCTCATACACTGGCCCATGGGTTTC
AAGCCTCCTCATCCAGAATGGATCATGAGCTGCAGTGAACTTTCCTTCTGCCTCTCACATCCTCGAGT
GCAGGACTTGCCTCTGGACGAGAGCAACGTGGTTATTCCCAGTGACACGGACTTCCTGGACACGTGGG
AGATTTTGATCCGATTTCAAATCCAGAGGAATGTGATAGTGATCCCCGGATCTATCACCCCAAGTCAC
ATTAAAGAGAATATCCAGGTGTTTGATTTTGAATTAACACAGCACGATATGGATAACATCCTCAGCCT
AAACAGGAACCTCCGACTGGCCATGTTCCCCATGGTCGACGGC

NOV45c, 275480984 Protein Sequence	SEQ ID NO: 684	218 aa	MW at 24946.5kD

TRSPTMGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKIKEGAVRRE
DLFIATKLWCTCHKKSLVETACRKSLKALKLNYLDLYLIHWPMGFKPPHPEWIMSCSELSFCLSHPRV
QDLPLDESNVVIPSDTDFLDTWEILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNILSL
NRNLRLAMFPMVDG

NOV45d, CG55778-01 DNA Sequence	SEQ ID NO: 685	956 bp
	ORF Start: ATG at 31	ORF Stop: TGA at 937

GGCGGGGCGGCGGGGCGGCCGGCGGCGGCCATGGGAGATATCCCAGCCGTGGGCCTCAGCTCCTGGAA
GCAGGCTTCTCCAGGGAAAGTGACCGAGGCAGTGAAAGAGGCCATTGACGCAGGGTACCGGCACTTCG
ACTGTGCTTACTTTTACCACAATGAGAGGGAGGTTGGAGCAGGGATCCGTTGCAAGATCAAGGAAGGC

GCTGTAAGACGGGAGGATCTGTTCATTGCCACTAAGCTGTGGTGCACCTGCCATAAGAAGTCCTTGGT
GGAAACAGCATGCAGAAAGAGTCTCAAGGCCTTGAAGCTGAACTATTTGGACCTCTACCTCATACACT
GGCCCATGGGTTTCAAGCCTCGAGTGCAGGACTTGCCTCTGGACGAGAGCAACATGGTTATTCCCAGT
GACACGGACTTCCTGGACACGTGGGAGGCCATGGAGGACCTGGTGATCACCGGGCTGGTGAAGAACAT
CGGGGTGTCAAACTTCAACCATGAACAGCTTGAGAGGCTTTTGAATAAGCCTGGGTTGAGGTTCAAGC
CACTAACCAACCAGATTGAGTGCCACCCATATCTTACTCAGAAGAATCTGATCAGTTTTTGCCAATCC
AGAGATGTGTCCGTGACTGCTTACCGTCCTCTTGGTGGCTCTAGTGAGGGGGTTGACCTGATAGACAA
CCCTGTGATCAAGAGGATTGCAAAGGAGCACGGCAAGTCTCCTGCTCAGATTTTGATCCGATTTCAAA
TCCAGAGGAATGTGATAGTGATCCCCGGATCTATCACCCCAAGTCACATTAAAGAGAATATCCAGGTG
TTTGATTTTGAATTAACACAGCACGATATGGATAACATCCTCAGCCTAAACAGGAATCTCCGACTGGC
CATGTTCCCCAGAACTAAAAATCACAAAGACTATCCTTTCCACATAGAATAC**TGA**GGACGCTTCCCCT
TCCT

NOV45d, CG55778-01 Protein Sequence	SEQ ID NO: 686	302 aa	MW at 34561.5kD

MGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKIKEGAVRREDLFIA
TKLWCTCHKKSLVETACRKSLKALKLNYLDLYLIHWPMGFKPRVQDLPLDESNMVIPSDTDFLDTWEA
MEDLVITGLVKNIGVSNFNHEQLERLLNKPGLRFKPLTNQIECHPYLTQKNLISFCQSRDVSVTAYRP
LGGSSEGVDLIDNPVIKRIAKEHGKSPAQILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDM
DNILSLNRNLRLAMFPRTKNHKDYPFHIEY

NOV45e, CG55778-02 DNA Sequence	SEQ ID NO: 687	875 bp
	ORF Start: ATG at 23	ORF Stop: TAA at 776

GGCGGGGCGGCCGGCGGCGGCC**ATG**GGAGATATCCCAGCCGTGGGCCTCAGCTCCTGGAAGCAGGCTT
CTCCAGGAAAAGTGACCGAGGCAGTGAAAGAGGCCATTGACGCAGGGTACCGGCACTTCGACTGTGCT
TACTTTTACCACAATGAGAGGGAGGTTGGAGCAGGGATCCGTTGCAAGATCAAGGAAGGCGCTGTAAG
ACGGGAGGATCTGTTCATTGCCACTAAGCTGTGGTGCACCTGCCATAAGAAGTCCTTGGTGGAAACAG
CATGCAGAAAGGGTCTCAAGGCCTTGAAGCTGAACTATTTGGACCTCTACCTCATACACTGGCCCATG
GGTTTCAAGCCTCCTCATCCAGAATGGATCATGAGCTGCAGTGAACTTTCCTTCTGCCTCTCACATCC
TCGAGTGCAGGACTTGCCTCTGGACGAGAGCAACATGGTTATTCCCAGTGACACGGACTTCCTGGACA
CGTGGGAGGCCATGGAGGACCTGGTGATCACCGGGCTGGTGAAGAACATCGGGGTGTCAAACTTCAAC
CATGAACAGCTTGAGAGGCTTTTGAATAAGCCTGGGTTGAGGTTCAAGCCACTAACCAACCAGATTTT
GATCCGATTTCAAATCCAGAGGAATGTGATAGTGATCCCCGGATCTATCACCCCAAGTCACATTAAAG
AGAATATCCAGGTGTTTGATTTTGAATTAACACAGCACGATATGGATAACATCCTCAGCCTAAACAGG
AATCTCCGACTGGCCATGTTCCCCATG**TAA**ATATGGCTCCTTCTTTTTAAAACAGAGGGAAGAATATA
CAGATTGAATGATTGGTGTCTGAATAGAACTAAAAATCACAAAGACTATCCTTTCCACA

NOV45e, CG55778-02 Protein Sequence	SEQ ID NO: 688	251 aa	MW at 28765.1kD

MGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKIKEGAVRREDLFIA
TKLWCTCHKKSLVETACRKGLKALKLNYLDLYLIHWPMGFKPPHPEWIMSCSELSFCLSHPRVQDLPL
DESNMVIPSDTDFLDTWEAMEDLVITGLVKNIGVSNFNHEQLERLLNKPGLRFKPLTNQILIRFQIQR
NVIVIPGSITPSHIKENIQVFDFELTQHDMDNILSLNRNLRLAMFPM

NOV45f, CG55778-04 DNA Sequence	SEQ ID NO: 689	785 bp
	ORF Start: ATG at 31	ORF Stop: TGA at 766

GGCGGGGCGGCGGGGCGGCCGGCGGCGGCC**ATG**GGAGATATCCCAGCCGTGGGCCTCAGCTCCTGGAA
GCAGGCTTCTCCAGGGAAAGTGACCGAGGCAGTGAAAGAGGCCATTGACGCAGGGTACCGGCACTTCG
ACTGTGCTTACTTTTACCACAATGAGAGGGAGGTTGGAGCAGGGATCCGTTGCAAGATCAAGGAAGGC
GCTGTAAGACGGGAGGATCTGTTCATTGCCACTAAGCTGTGGTGCACCTGCCATAAGAAGTCCTTGGT
GGAAACAGCATGCAGAAAGAGTCTCAAGGCCTTGAAGCTGAACTATTTGGACCTCTACCTCATACACT
GGCCCATGGGTTTCAAGCCTCGAGTGCAGGACTTGCCTCTGGACGAGAGCAACATGGTTATTCCCAGT
GACACGGACTTCCTGGACACGTGGGAGGCCATGGAGGACCTGGTGATCACCGGGCTGGTGAAGAACAT
CGGGGTGTCAAACTTCAACCATGAACAGCTTGAGAGGCTTTTGAATAAGCCTGGGTTGAGGTTCAAGC
CACTAACCAACCAGATTTTGATCCGATTTCAAATCCAGAGGAATGTGATAGTGATCCCCGGATCTATC
ACCCCAAGTCACATTAAAGAGAATATCCAGGTGTTTGATTTTGAATTAACACAGCACGATATGGATAA
CATCCTCAGCCTAAACAGGAATCTCCGACTGGCCATGTTCCCCAGAACTAAAAATCACAAAGACTATC
CTTTCCACATAGAATAC**TGA**GGACGCTTCCCCTTCCT

NOV45f, CG55778-04 Protein Sequence	SEQ ID NO: 690	245 aa	MW at 28311.4kD

MGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKIKEGAVRREDLFIA
TKLWCTCHKKSLVETACRKSLKALKLNYLDLYLIHWPMGFKPRVQDLPLDESNMVIPSDTDFLDTWEA
MEDLVITGLVKNIGVSNFNHEQLERLLNKPGLRFKPLTNQILIRFQIQRNVIVIPGSITPSHIKENIQ
VFDFELTQHDMDNILSLNRNLRLAMFPRTKNHKDYPFHIEY

NOV45g, CG55778-05 DNA Sequence	SEQ ID NO: 691	937 bp
	ORF Start: ATG at 31	ORF Stop: TAA at 838

GGCGGGGCGGCGGGGCGGCCGGCGGCGGCC**ATG**GGAGATATCCCAGCCGTGGGCCTCAGCTCCTGGAA
GCAGGCTTCTCCAGGGAAAGTGACCGAGGCAGTGAAAGAGGCCATTGACGCAGGGTACCGGCACTTCG
ACTGTGCTTACTTTTACCACAATGAGAGGGAGGTTGGAGCAGGGATCCGTTGCAAGATCAAGGAAGGC
GCTGTAAGACGGGAGGATCTGTTCATTGCCACTAAGCCTCCTCATCCAGAATGGATCATGAGCTGCAG
TGAACTTTCCTTCTGCCTCTCACATCCTCGAGTGCAGGACTTGCCTCTGGACGAGAGCAACATGGTTA
TTCCCAGTGACACGGACTTCCTGGACACGTGGGAGGCCATGGAGGACCTGGTGATCACCGGGCTGGTG
AAGAACATCGGGGTGTCAAACTTCAACCATGAACAGCTTGAGAGGCTTTTGAATAAGCCTGGGTTGAG
GTTCAAGCCACTAACCAACCAGATTGAGTGCCACCCATATCTTACTCAGAAGAATCTGATCAGTTTTT
GCCAATCCAGAGATGTGTCCGTGACTGCTTACCGTCCTCTTGGTGGCTCGTGTGAGGGGGTTGACCTG
ATAGACAACCCTGTGATCAAGAGGATTGCAAAGGAGCACGGCAAGTCTCCTGCTCAGATTTTGATCCG
ATTTCAAATCCAGAGGAATGTGATAGTGATCCCCGGATCTATCACCCCAAGTCACATTAAAGAGAATA
TCCAGGTGTTTGATTTTGAATTAACACAGCACGATATGGATAACATCCTCAGCCTAAACAGGAATCTC
CGACTGGCCATGTTCCCCATGT**AAA**TATGGCTCCTTCTTTTTAAAACAGAGGGAAGAATATACAGATT
GAATGATTGGTGTCTGAATAGAACTAAAAATCACAAAGACTATCCTTTCCACA

NOV45g, CG55778-05 Protein Sequence	SEQ ID NO: 692	269 aa	MW at 30426.6kD

MGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKIKEGAVRREDLFIA
TKPPHPEWIMSCSELSFCLSHPRVQDLPLDESNMVIPSDTDFLDTWEAMEDLVITGLVKNIGVSNFNH
EQLERLLNKPGLRFKPLTNQIECHPYLTQKNLISFCQSRDVSVTAYRPLGGSCEGVDLIDNPVIKRIA
KEHGKSPAQILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNILSLNRNLRLAMFPM

NOV45h, CG55778-07 DNA Sequence	SEQ ID NO: 693	1527 bp
	ORF Start: ATG at 21	ORF Stop: TGA at 981

CGGGGCGGCCGGCGGCGGCC**ATG**GGAGATATCCCAGCCGTGGGCCTCAGCTCCTGGAAGGCTTCTCCA
GGGAAAGTGACCGAGGCAGTGAAAGAGGCCATTGACGCAGGGTACCGGCACTTCGACTGTGCTTACTT
TTACCACAATGAGAGGGAGGTTGGAGCAGGGATCCGTTGCAAGATCAAGGAAGGCGCTGTAAGACGGG
AGGATCTGTTCATTGCCACTAAGCTGTGGTGCACCTGCCATAAGAAGTCCTTGGTGGAAACAGCATGC
AGAAAGAGTCTCAAGGCCTTGAAGCTGAACTATTTGGACCTCTACCTCATACACTGGCCCATGGGTTT
CAAGCCTCCTCATCCAGAATGGATCATGAGCTGCAGTGAACTTTCCTTCTGCCTCTCACATCCTCGAG
TGCAGGACTTGCCTCTGGACGAGAGCAACATGGTTATTCCCAGTGACACGGACTTCCTGGACACGTGG
GAGGCCATGGAGGACCTGGTGATCACCGGGCTGGTGAAGAACATCGGGGTGTCAAACTTCAACCATGA
ACAGCTTGAGAGGCTTTTGAATAAGCCTGGGTTGAGGTTCAAGCCACTAACCAACCAGATTGAGTGCC
ACCCATATCTTACTCAGAAGAATCTGATCAGTTTTTGCCAATCCAGAGATGTGTCCGTGACTGCTTAC
CGTCCTCTTGGTGGCTCGTGTGAGGGGGTTGACCTGATAGACAACCCTGTGATCAAGAGGATTGCAAA
GGAGCACGGCAAGTCTCCTGCTCAGATTTTGATCCGATTTCAAATCCAGAGGAATGTGATAGTGATCC
CCGGATCTATCACCCCAAGTCACATTAAAGAGAATATCCAGGTGTTTGATTTTGAATTAACACAGCAC
GATATGGATAACATCCTCAGCCTAAACAGGAATCTCCGACTGGCCATGTTCCCCATAACTAAAAATCA
CAATGACTATCCTTTCCACATAGAATACT**GA**GGACGCTTCCCCTTCCTTGTTTCTGCTCAGCCCAGAT
GCACAGACACTATTGGCAATGTTGACCCTCCTCTGTCATCACAGCGCCAGGGCAGCTGTGCCTGGGAC
AGGAGCCACACAGTCAGAGGGGGATGTAAGAGCCACCTTCTCTGACAAATCTGGAGAATTGAGTGTGT
TCTAAGTGAAGGCAATGGGGTTTCTCCAAGACAGCCTGTGTGGCCTCTACTCTGAACAAATACACTGA
TGAGTCATCAGTGAAATTTGCCTTCACATTTTAAGAAAACTTTATCTTATGGAGTTATTTAAGCCATC
TACAGAGCTGAGGAAACAGTGTAATGTGTCTCTGCCCCATTGCGCAGCTCCACCCATTGTGCCCCAGG
CCAGCCCGCGTCACCTACACTTCCTTCTGTGCCCTGCCAGTGACCCCCAGGTTATTCTAAAGCAGAGT
CCTTCCCTTCCCCCAGTGAGAAGGAAAATGGGATAAGTCTGGGACACTGTTTCAGTTCAATAAAGAGG
CTTTTTTCTTCCTTAAAAAAAAAAAAAAAA

NOV45h, CG55778-07 Protein Sequence	SEQ ID NO: 694	320 aa	MW at 36574.8kD

MGDIPAVGLSSWKASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKIKEGAVRREDLFIAT
KLWCTCHKKSLVETACRKSLKALKLNYLDLYLIHWPMGFKPPHPEWIMSCSELSFCLSHPRVQDLPLD
ESNMVIPSDTDFLDTWEAMEDLVITGLVKNIGVSNFNHEQLERLLNKPGLRFKPLTNQIECHPYLTQK
NLISFCQSRDVSVTAYRPLGGSCEGVDLIDNPVIKRIAKEHGKSPAQILIRFQIQRNVIVIPGSITPS
HIKENIQVFDFELTQHDMDNILSLNRNLRLAMFPITKNHNDYPFHIEY

NOV45i, CG55778-08 DNA Sequence	SEQ ID NO: 695	1680 bp
	ORF Start: ATG at 101	ORF Stop: TAA at 1022

GGCACGAGGACGGCGGGTCGCCAGCGCCTCAGTAGCTCGCGCGGTGCCTGTCGGTAGTCGCGTGCGGG
GCGGCGGGGCGGCGGGGCGGCCGGCGGCGGCCATGGGAGATATCCCAGCCGTGGGCCTCAGCTCCTGG
AAGGCTTCTCCAGGGAAAGTGACCGAGGCAGTGAAAGAGGCCATTGACGCAGGGTACCGGCACTTCGA
CTGTGCTTACTTTTACCACAATGAGAGGGAGGTTGGAGCAGGGATCCGTTGCAAGATCAAGGAAGGCG
CTGTAAGACGGGAGGATCTGTTCATTGCCACTAAGCTGTGGTGCACCTGCCATAAGAAGTCCTTGGTG
GAAACAGCATGCAGAAAGAGTCTCAAGGCCTTGAAGCTGAACTATTTGGACCTCTACCTCATACACTG
GCCCATGGGTTTCAAGCCTCCTCATCCAGAATGGATCATGAGCTGCAGTGAACTTTCCTTCTGCCTCT
CACATCCTCGAGTGCAGGACTTGCCTCTGGACGAGAGCAACATGGTTATTCCCAGTGACACGGACTTC
CTGGACACGTGGGAGGCCATGGAGGACCTGGTGATCACCGGGCTGGTGAAGAACATCGGGGTGTCAAA
CTTCAACCATGAACAGCTTGAGAGGCTTTTGAATAAGCCTGGGTTGAGGTTCAAGCCACTAACCAACC
AGATTGAGTGCCACCCATATCTTACTCAGAAGAATCTGATCAGTTTTTGCCAATCCAGAGATGTGTCC
GTGACTGCTTACCGTCCTCTTGGTGGCTCGTGTGAGGGGGTTGACCTGATAGACAACCCTGTGATCAA
GAGGATTGCAAAGGAGCACGGCAAGTCTCCTGCTCAGATTTTGATCCGATTTCAAATCCAGAGGAATG
TGATAGTGATCCCCGGATCTATCACCCCAAGTCACATTAAAGAGAATATCCAGGTGTTTGATTTTGAA
TTAACACAGCACGATATGGATAACATCCTCAGCCTAAACAGGAATCTCCGACTGGCCATGTTCCCCAT
GTAAATATGGCTCCTTCTTTTTAAAACAGAGGGAAGAATATACAGATTGAATGATTGGTGTCTGAATA
GAACTAAAAATCACAAAGACTATCCTTTCCACATAGAATACTGAGGACGCTTCCCCTTCCTTGTTTCT
GCTCAGCCCAGATGCACAGACACTATTGGCAATGTTGACCCTCCTCTGTCATCACAGCGCCAGGGCAG
CTGTGCCTGGGACAGGAGCCACACAGTCAGAGGGGGATGTAAGAGCCACCTTCTCTGACAAATCTGGA
GAATTGAGTGTGTTCTAAGTGAAGGCAATGGGGTTTCTCCAAGACAGCCTGTGTGGCCTCTACTCTGA
ACAAATACACTGATGAGTCATCAGTGAAATTTGCCTTCACATTTTAAGAAAACTTTATCTTATGGAGT
TATTTAAGCCATCTACAGAGCTGAGGAAACAGTGTAATGTGTCTCTGCCCCATTGCGCAGCTCCACCC
ATTGTGCCCCAGGCCAGCCCGCGTCACCTACACTTCCTTCTGTGCCCTGCCAGTGACCCCCAGGTTAT
TCTAAAGCAGAGTCCTTCCCTTCCCCCAGTGAGAAGGAAAATGGGATAAGTCTGGGACACTGTTTCAG
TTCAATAAAGAGGCTTTTTTCTTCCTTAAAAAAAAAAAAAAAAAAAA

NOV45i, CG55778-08 Protein Sequence	SEQ ID NO: 696	307 aa	MW at 34933.0kD

MGDIPAVGLSSWKASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKIKEGAVRREDLFIAT
KLWCTCHKKSLVETACRKSLKALKLNYLDLYLIHWPMGFKPPHPEWIMSCSELSFCLSHPRVQDLPLD
ESNMVIPSDTDFLDTWEAMEDLVITGLVKNIGVSNFNHEQLERLLNKPGLRFKPLTNQIECHPYLTQK
NLISFCQSRDVSVTAYRPLGGSCEGVDLIDNPVIKRIAKEHGKSPAQILIRFQIQRNVIVIPGSITPS
HIKENIQVFDFELTQHDMDNILSLNRNLRLAMFPM

NOV45j, SNP13375813 of CG55778-03, DNA Sequence	SEQ ID NO: 697	752 bp
	ORF Start: ATG at 23	ORF Stop: TAA at 653
	SNP Pos: 624	SNP Change: A to G

GGCGGGGCGGCCGGCGGCGGCCATGGGAGATATCCCAGCCGTGGGCCTCAGCTCCTGGAAGCAGGCTT
CTCCAGGTAAAGTGACCGAGGCAGTGAAAGAGGCCATTGACGCAGGGTACCGGCACTTCGACTGTGCT
TACTTTTACCACAATGAGAGGGAGGTTGGAGCAGGGATCCGTTGCAAGATCAAGGAAGGCGCTGTAAG
ACGGGAGGATCTGTTCATTGCCACTAAGCTGTGGTGCACCTGCCATAAGAAGTCCTTGGTGGAAACAG
CATGCAGAAAGAGTCTCAAGGCCTTGAAGCTGAACTATTTGGACCTCTACCTCATACACTGGCCCATG
GGTTTCAAGCCTCCTCATCCAGAATGGATCATGAGCTGCAGTGAACTTTCCTTCTGCCTCTCACATCC
TCGAGTGCAGGACTTGCCTCTGGACGAGAGCAACATGGTTATTCCCAGTGACACGGACTTCCTGGACA
CGTGGGAGATTTTGATCCGATTTCAAATCCAGAGGAATGTGATAGTGATCCCCGGATCTATCACCCCA
AGTCACATTAAAGAGAATATCCAGGTGTTTGATTTTGAATTAACACAGCACGATATGGATAACATCCT

CAGCCTAAACAGGAATCTCCGACTGGCCATGTTCCCCATGTAAATATGGCTCCTTCTTTTTAAAACAG
AGGGAAGAATATACAGATTGAATGATTGGTGTCTGAATAGAACTAAAAATCACAAAGACTATCCTTTC
CACA

NOV45j, SNP13375813 of CG55778-03, Protein Sequence	SEQ ID NO: 698	210 aa	MW at 24164.7kD
	SNP Pos: 201		SNP Change: Lys to Arg

MGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKIKEGAVRREDLFIA
TKLWCTCHKKSLVETACRKSLKALKLNYLDLYLIHWPMGFKPPHPEWIMSCSELSFCLSHPRVQDLPL
DESNMVIPSDTDFLDTWEILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNILSLNRNLR
LAMFPM

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 45B.

Table 45B. Comparison of the NOV45 protein sequences.

```
NOV45a  -----MGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKI
NOV45b  -----MGDIPAVGLSSWKAS-PGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKI
NOV45c  TRSPTMGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKI
NOV45d  -----MGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKI
NOV45e  -----MGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKI
NOV45f  -----MGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKI
NOV45g  -----MGDIPAVGLSSWKQASPGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKI
NOV45h  -----MGDIPAVGLSSWKAS-PGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKI
NOV45i  -----MGDIPAVGLSSWKAS-PGKVTEAVKEAIDAGYRHFDCAYFYHNEREVGAGIRCKI

NOV45a  KEGAVRREDLFIATKLWCTCHKKSLVETACRKSLKALKLNYLDLYLIHWPMGFKPPHPEW
NOV45b  KEGAVRREDLFIAT-----------------------------------------KPPHPEW
NOV45c  KEGAVRREDLFIATKLWCTCHKKSLVETACRKSLKALKLNYLDLYLIHWPMGFKPPHPEW
NOV45d  KEGAVRREDLFIATKLWCTCHKKSLVETACR-------------------KSLKALKLNY
NOV45e  KEGAVRREDLFIATKLWCTCHKKSLVETACRKGLKALKLNYLDLYLIHWPMGFKPPHPEW
NOV45f  KEGAVRREDLFIATKLWCTCHKKSLVETACR-------------------KSLKALKLNY
NOV45g  KEGAVRREDLFIAT-----------------------------------------KPPHPEW
NOV45h  KEGAVRREDLFIATKLWCTCHKKSLVETACRKSLKALKLNYLDLYLIHWPMGFKPPHPEW
NOV45i  KEGAVRREDLFIATKLWCTCHKKSLVETACRKSLKALKLNYLDLYLIHWPMGFKPPHPEW

NOV45a  IMSCSELSFCLSHPRVQDLPLDESNMVIPSDTDFLDTWE----------------------
NOV45b  IMSCSELSFCLSHPRVQDLPLDESNMVIPSDTDFLDTWE----------------------
NOV45c  IMSCSELSFCLSHPRVQDLPLDESNVVIPSDTDFLDTWE----------------------
NOV45d  LDLYLIHWPMGFKPRVQDLPLDESNMVIPSDTDFLDTWEAMEDLVITGLVKNIGVSNFNH
NOV45e  IMSCSELSFCLSHPRVQDLPLDESNMVIPSDTDFLDTWEAMEDLVITGLVKNIGVSNFNH
NOV45f  LDLYLIHWPMGFKPRVQDLPLDESNMVIPSDTDFLDTWEAMEDLVITGLVKNIGVSNFNH
NOV45g  IMSCSELSFCLSHPRVQDLPLDESNMVIPSDTDFLDTWEAMEDLVITGLVKNIGVSNFNH
NOV45h  IMSCSELSFCLSHPRVQDLPLDESNMVIPSDTDFLDTWEAMEDLVITGLVKNIGVSNFNH
NOV45i  IMSCSELSFCLSHPRVQDLPLDESNMVIPSDTDFLDTWEAMEDLVITGLVKNIGVSNFNH

NOV45a  ------------------------------------------------------------
NOV45b  ------------------------------------------------------------IL-
NOV45c  ------------------------------------------------------------
NOV45d  EQLERLLNKPGLRFKPLTNQIECHPYLTQKNLISFCQSRDVSVTAYRPLGGSSEGVDLID
NOV45e  EQLERLLNKPGLRFKPLTN-----------------------------------------
NOV45f  EQLERLLNKPGLRFKPLT------------------------------------------
NOV45g  EQLERLLNKPGLRFKPLTNQIECHPYLTQKNLISFCQSRDVSVTAYRPLGGSCEGVDLID
NOV45h  EQLERLLNKPGLRFKPLTNQIECHPYLTQKNLISFCQSRDVSVTAYRPLGGSCEGVDLID
```

```
NOV45i  EQLERLLNKPGLRFKPLTNQIECHPYLTQKNLISFCQSRDVSVTAYRPLGGSCEGVDLID

NOV45a  ------------------ILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNIL
NOV45b  ------------------I--RFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNIL
NOV45c  ------------------ILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNIL
NOV45d  NPVIKRIAKEHGKSPAQILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNIL
NOV45e  ----------------QILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNIL
NOV45f  --------N-------QILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNIL
NOV45g  NPVIKRIAKEHGKSPAQILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNIL
NOV45h  NPVIKRIAKEHGKSPAQILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNIL
NOV45i  NPVIKRIAKEHGKSPAQILIRFQIQRNVIVIPGSITPSHIKENIQVFDFELTQHDMDNIL

NOV45a  SLNKNLRLAMFPM-------------
NOV45b  SLNRNLRLAMFPITKNHKDYPFHIEY
NOV45c  SLNRNLRLAMFPMVDG----------
NOV45d  SLNRNLRLAMFPRTKNHKDYPFHIEY
NOV45e  SLNRNLRLAMFPM-------------
NOV45f  SLNRNLRLAMFPRTKNHKDYPFHIEY
NOV45g  SLNRNLRLAMFPM-------------
NOV45h  SLNRNLRLAMFPITKNHNDYPFHIEY
NOV45i  SLNRNLRLAMFPM-------------

NOV45a  (SEQ ID NO:  680)
NOV45b  (SEQ ID NO:  682)
NOV45c  (SEQ ID NO:  684)
NOV45d  (SEQ ID NO:  686)
NOV45e  (SEQ ID NO:  688)
NOV45f  (SEQ ID NO:  690)
NOV45g  (SEQ ID NO:  692)
NOV45h  (SEQ ID NO:  694)
NOV45i  (SEQ ID NO:  696)
```

Further analysis of the NOV45a protein yielded the following properties shown in Table 45C.

Table 45C. Protein Sequence Properties NOV45a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 3;  pos.chg 0;  neg.chg 1
      H-region:  length 9;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -6.56
      possible cleavage site: between 18 and 19

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
```

```
      Tentative number of TMS(s) for the threshold  0.5:    0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  6.21 (at 155)
      ALOM score:   6.21  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  5.33
      Hyd Moment(95):  5.59     G content:       3
      D/E content:     2        S/T content:     4
      Score: -7.44

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  11.4%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
```

```
      Prediction: cytoplasmic
      Reliability: 89

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       52.2 %: cytoplasmic
       34.8 %: nuclear
       13.0 %: mitochondrial

>> prediction for CG55778-03 is cyt (k=23)
```

A search of the NOV45a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 45D.

Table 45D. Geneseq Results for NOV45a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV45a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABP53592	Human NOV16c protein SEQ ID NO:48 - Homo sapiens, 210 aa. [WO200262999-A2, 15-AUG-2002]	1..210 1..210	210/210 (100%) 210/210 (100%)	e-124
ABP53591	Human NOV16b protein SEQ ID NO:46 - Homo sapiens, 251 aa. [WO200262999-A2, 15-AUG-2002]	1..210 1..251	208/251 (82%) 209/251 (82%)	e-117
AAM79279	Human protein SEQ ID NO 1941 - Homo sapiens, 263 aa. [WO200157190-A2, 09-AUG-2001]	1..210 1..250	205/251 (81%) 207/251 (81%)	e-114
ABU11795	Human MDDT polypeptide SEQ ID 742 - Homo sapiens, 215 aa. [WO200279449-A2, 10-OCT-2002]	1..156 31..185	152/156 (97%) 154/156 (98%)	5e-89
AAM80263	Human protein SEQ ID NO 3909 - Homo sapiens, 264 aa. [WO200157190-A2, 09-AUG-2001]	1..156 11..165	151/156 (96%) 152/156 (96%)	3e-88

In a BLAST search of public sequence databases, the NOV45a protein was found to have homology to the proteins shown in the BLASTP data in Table 45E.

Table 45E. Public BLASTP Results for NOV45a

Protein Accession Number	Protein/Organism/Length	NOV45a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
BAC54567	Aldo-keto reductase related protein 3 - Homo sapiens (Human), 222 aa.	1..210 1..209	207/210 (98%) 209/210 (98%)	e-121
BAC54565	Aldo-keto reductase related protein 1 - Homo sapiens (Human), 250 aa.	1..210 1..250	208/251 (82%) 209/251 (82%)	e-115
BAC54566	Aldo-keto reductase related protein 2 - Homo sapiens (Human), 263 aa.	1..210 1..250	207/251 (82%) 209/251 (82%)	e-115
Q9BU71	Similar to aldo-keto reductase - Homo sapiens (Human), 307 aa.	1..156 1..155	153/156 (98%) 154/156 (98%)	6e-89
Q96JD6	Aldo-keto reductase loopADR - Homo sapiens (Human), 320 aa.	1..156 1..155	153/156 (98%) 154/156 (98%)	6e-89

PFam analysis indicates that the NOV45a protein contains the domains shown in the Table 45F.

Table 45F. Domain Analysis of NOV45a

Pfam Domain	NOV45a Match Region	Identities/ Similarities for the Matched Region	Expect Value
aldo_ket_red	2..203	87/369 (24%) 157/369 (43%)	8.2e-37

Example 46.

The NOV46 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 46A.

Table 46A. NOV46 Sequence Analysis

NOV46a, CG55794-03 DNA Sequence	SEQ ID NO: 699	1222 bp
	ORF Start: ATG at 41	ORF Stop: TAA at 1163

CCAGAGAGGCCCAGAATTTTCTAACTTACTGTGTGGCAGAATGAAGCCTCTGCTTGAAACCCTTTATC
TTTTGGGGATGCTGGTTCCTGGAGGGCTGGGATATGATAGATCCTTAGCCCAACACAGACAAGAGATT
GTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGA
GATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAG
TGACCTATGAGACCCACCCCATATATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATC
ATTTGGATGGACTGTGGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAA
AGAAATTCTACAAAACCATAAAGACAACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATG
TCCTTCCAGTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCC
CGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCTTCTTGGTG
TAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAG
AGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCAC
TCTTATGGGCAGTTAATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAAT
GATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGAT
CGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCC
TTCTCATATACGTTTGAGCTGAGGGACAGTGGAACATATGGGTTTGTTCTGCCAGAAGCTCAGATCCA
GCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGC
ACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCT
CTTCTC**TAA**GTGCATTCTGCCCAGGCCTGCTCAACCCCAGTGGCATGAGTGTGGCTTGGAGGAACG

NOV46a, CG55794-03 Protein Sequence	SEQ ID NO: 700	374 aa	MW at 42580.0kD

MKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYK
EVVTQHFLGVTYETHPIYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSRIR
KLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTF
CGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKA
KYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLSV
LDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46b, 210223559 DNA Sequence	SEQ ID NO: 701	1579 bp
	ORF Start: at 67	ORF Stop: TAA at 1396

CCAATTGACAAGCTTTTGATTTTAACGACTTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATT
ATTCGAAACGATGAGATTTCCTTCAATTTTTACTGCTGTTTTATTCGCAGCATCCTCCGCATTAGCTG
CTCCAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCGGCTGAAGCTGTCATCGGTTACTCA
GATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTT
TATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGATATGATA
GATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTAT
TCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAA
GGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTGAAGATCA
GCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGAATGGATT
GCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAACTCAAGTATACG
CAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGGTTATATCTACACTT
GGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACATGTTTTGGGACGGAT
CTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATT
CTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGG
ATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGCTACACC
AAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAAAGC
AAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTCAA
GAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGGAACGTAT
GGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGT
CCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTATGC
TGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTCT**TAA**GCGGCCGCCAGCTTTCTAGAACAAAAACTC
ATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATTGAGTTTGTAGCCTTAG
ACATGACTGTTCCTCAGTTCAAGTTGGGCACTTACGAGAAGACCGGTCTTGCTAGATTCTAATCAAGA

GGATGTCAGAATGCC			
NOV46b, 210223559 Protein Sequence	SEQ ID NO: 702	443 aa	MW at 49842.7kD

LFETMRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSNSTNNGLL
FINTTIASIAAKEEGVSLEKRYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKY
KEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSI
RKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQT
FCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALK
AKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLS
VLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46c, 210223626 DNA Sequence	SEQ ID NO: 703	1607 bp
	ORF Start: at 74	ORF Stop: TAA at 1421

ACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTTTTAACGACAACTTGAGAAGATCAAAAAACA
ACTAATTATTCGAAACGATGAGATTTCCTTCAATTTTTACTGCTGTTTTATTCGCAGCATCCTCCGCA
TTAGCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCGGCTGAAGCTGTCATCGG
TTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGT
TATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGA
TATGATAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGA
GACGTATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGA
AGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTG
AAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGA
ATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAACTCAA
GTATACGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGGTTATATC
TACACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACATGTTTTGG
GACGGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATC
AAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGC
AAGAAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGG
CTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCAAATGCAT
TGAAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTCATCAGGG
TCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGG
AACGTATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGC
TGTCAGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGACATCTGCC
ACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTCCACCATCACCACCATCAC**TAA**GCGGC
CGCCAGCTTTCTAGAACAAAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATC
ATCATCATTGAGTTTGTAGCCTTAGACATGACTGTTCCTCAGTTCAAGTTGGGCACTTACGAGAAGAC
CGGTCTTGCTAGATTCTAATCAAGAGGATGTCAGAATGCCATT

NOV46c, 210223626 Protein Sequence	SEQ ID NO: 704	449 aa	MW at 50665.6kD

LFETMRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSNSTNNGLL
FINTTIASIAAKEEGVSLEKRYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKY
KEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSI
RKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQT
FCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALK
AKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLS
VLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLLHHHHHH

NOV46d, 171095097 DNA Sequence	SEQ ID NO: 705	1148 bp
	ORF Start: at 2	ORF Stop: TAA at 1127

CACCATGAAGCCTCTGCTTGAAACCCTTTATCTTTTGGGGATGCTGGTTCCTGGAGGGCTGGGATATG
ATAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACG
TATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTA
CAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTGAAGA
TCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGAATGG
ATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAACTCAAGTAT
ACGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGGTTATATCTACA

CTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACATGTTTTGGGACG
GATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAAC
ATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGA
AGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGCTAC
ACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAA
AGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTCATCAGGGTCTT
CAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGGAACG
TATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTC
AGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTA
TGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTC**TAA**GCGGCCGCTCGAGTCTAGA

NOV46d, 171095097 Protein Sequence	SEQ ID NO: 706	375 aa	MW at 42630.1kD

TMKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKY
KEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSI
RKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQT
FCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALK
AKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLS
VLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46e, 183852229 DNA Sequence	SEQ ID NO: 707	1148 bp
	ORF Start: at 3	ORF Stop: TAA at 1146

CCACCATGAAGCCTCTGCTTGAAACCCTTTATCTTTTGGGGATGCTGGTTCCTGGAGGGCTGGGATAT
GATAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGAC
GTATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGT
ACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTGAAG
ATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGAATG
GATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAACTCAAGTA
TACGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGGTTATATCTAC
ACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACATGTTTTGGGAC
GGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAA
CATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAG
AAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGCTA
CACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCAAATGCATTGA
AAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTCATCAGGGTCT
TCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGGAAC
GTATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGT
CAGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGACATCTGCCACT
ATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTCCACCATCACCACCATCAC**TAA**

NOV46e, 183852229 Protein Sequence	SEQ ID NO: 708	381 aa	MW at 43452.9kD

TMKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKY
KEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSI
RKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQT
FCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALK
AKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLS
VLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLLHHHHHH

NOV46f, 183852264 DNA Sequence	SEQ ID NO: 709	1151 bp
	ORF Start: at 3	ORF Stop: TAA at 1149

CCACCATGGGCCACCATCACCACCATCACAAGCCTCTGCTTGAAACCCTTTATCTTTTGGGGATGCTG
GTTCCTGGAGGGCTGGGATATGATAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGT
GAGTCCATGGAGCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGA
TGAGAGAGATCAGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACC
CACCCCATGTATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTG
TGGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAA
ACCATAAAGACAACTCAAGTATACGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTT

AACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAA
TAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCT
CTAGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTT
GCCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTT
AATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGAC
AGAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATT
TTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTT
TGAGCTGAGGGACAGTGGAACGTATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGG
AGACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCT
GGAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTC**TAA**

NOV46f, 183852264 Protein Sequence	SEQ ID NO: 710	382 aa	MW at 43510.0kD

TMGHHHHHHKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWM
REISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQN
HKDNSSIRKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGAS
RNCQDQTFCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQ
KAANALKAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEE
TMEAVLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46g, 183852410 DNA Sequence	SEQ ID NO: 711	1162 bp
	ORF Start: at 2	ORF Stop: TAA at 1148

CACCATGGGCCACCATCACCACCATCACAAGCCTCTGCTTGAAACCCTTTATCTTTTGGGGATGCTGG
TTCCTGGAGGGCTGGGATATGATAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTG
AGTCCATGGAGCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGAT
GAGAGAGATCAGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCC
ACCCCATGTATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGT
GGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAA
CCATAAAGACAACTCAAGTATACGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTA
ACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAAT
AATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTC
TAGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTG
CCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTA
ATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACA
GAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTT
TATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTT
GAGCTGAGGGACAGTGGAACGTATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGA
GACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTG
GAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTC**TAA**GCGGCC
GCTTTC

NOV46g, 183852410 Protein Sequence	SEQ ID NO: 712	382 aa	MW at 43510.0kD

TMGHHHHHHKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWM
REISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQN
HKDNSSIRKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGAS
RNCQDQTFCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQ
KAANALKAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEE
TMEAVLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46h, 183523337 DNA Sequence	SEQ ID NO: 713	1162 bp
	ORF Start: at 2	ORF Stop: TAA at 1148

CACCATGGGCCACCATCACCACCATCACAAGCCTCTGCTTGAAACCCTTTATCTTTTGGGGATGCTGG
TTCCTGGAGGGCTGGGATATGATAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTG
AGTCCATGGAGCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGAT
GAGAGAGATCAGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCC
ACCCCATGTATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGT
GGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAA

CCATAAAGACAACTCAAGTATACGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTA
ACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAAT
AATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTC
TAGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTG
CCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTA
ATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACA
GAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTT
TATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTT
GAGCTGAGGGACAGTGGAACGTATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGA
GACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTG
GAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTCTAAGCGGCC
GCTTTC

NOV46h, 183523337 Protein Sequence	SEQ ID NO: 714	382 aa	MW at 43510.0kD

TMGHHHHHHKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWM
REISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQN
HKDNSSIRKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGAS
RNCQDQTFCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQ
KAANALKAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEE
TMEAVLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46i, CG55794-01 DNA Sequence	SEQ ID NO: 715	1196 bp
	ORF Start: ATG at 16	ORF Stop: TAA at 1138

TTACTGTGTGGCAGAATGAAGCCTCTGCTTGAAACCCTTTATCTTTTGGGGATGCTGGTTCCTGGAGG
GCTGGGATATGATAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGA
GCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGAGATCAATGAGTGGATGAGAGAGATC
AGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATATA
TTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGGCTGTGGAATTCACG
CCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGAC
AACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGG
TTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACAT
GTTTTGGGACGGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGC
CAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCAT
AGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACAC
CTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCA
AATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTC
ATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGG
ACAGTGGAACATATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAG
GCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGAC
ATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTCTAAGTGCATCCTGCCCAGG
CCTGCTCAACCCCAGTGGCATGAGTGTGGCTGGAGGAACG

NOV46i, CG55794-01 Protein Sequence	SEQ ID NO: 716	374 aa	MW at 42472.9kD

MKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEINEWMREISEKYK
EVVTQHFLGVTYETHPIYYLKISQPSGNPKKIIWMGCGIHAREWIAPAFCQWFVKEILQNHKDNSRIR
KLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTF
CGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKA
KYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLSV
LDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46j, CG55794-02 DNA Sequence	SEQ ID NO: 717	1008 bp
	ORF Start: at 1	ORF Stop: end of sequence

TATGATAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGA
GACGTATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGA
AGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTG
AAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGA

ATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAACTCAA
GTATACGCAAGCTCCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGGTTATATC
TACACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACATGTTTTGG
GACGGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATC
AAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGC
AAGAAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGG
CTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCAAATGCAT
TGAAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTCATCAGGG
TCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGG
AACATATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGC
TGTCAGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGG

NOV46j, CG55794-02 Protein Sequence	SEQ ID NO: 718	336 aa	MW at 38597.1kD

YDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYL
KISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETKAVASFIES
KKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGTNYRVGSSADILYASSG
SSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLSVLDDVYAKHWHSDSAGR

NOV46k, CG55794-04 DNA Sequence	SEQ ID NO: 719	1579 bp
	ORF Start: at 334	ORF Stop: TAA at 1396

CCAATTGACAAGCTTTTGATTTTAACGACTTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATT
ATTCGAAACGATGAGATTTCCTTCAATTTTTACTGCTGTTTTATTCGCAGCATCCTCCGCATTAGCTG
CTCCAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCGGCTGAAGCTGTCATCGGTTACTCA
GATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTT
TATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGATATGATA
GATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTAT
TCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAA
GGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTGAAGATCA
GCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGAATGGATT
GCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAACTCAAGTATACG
CAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGGTTATATCTACACTT
GGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACATGTTTTGGGACGGAT
CTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATT
CTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGG
ATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGCTACACC
AAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAAAGC
AAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTCAA
GAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGGAACGTAT
GGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGT
CCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTATGC
TGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTC**TAA**GCGGCCGCCAGCTTTCTAGAACAAAAACTC
ATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATCATTGAGTTTGTAGCCTTAG
ACATGACTGTTCCTCAGTTCAAGTTGGGCACTTACGAGAAGACCGGTCTTGCTAGATTCTAATCAAGA
GGATGTCAGAATGCC

NOV46k, CG55794-04 Protein Sequence	SEQ ID NO: 720	354 aa	MW at 40431.3kD

YDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYL
KISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETKAVASFIES
KKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGTNYRVGSSADILYASSG
SSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLSVLDDVYAKHWHSDSAGRVTSA
TMLLGLLVSCMSLL

NOV46l, CG55794-05 DNA Sequence	SEQ ID NO: 721	1607 bp
	ORF Start: at 341	ORF Stop: at 1403

ACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTTTTAACGACAACTTGAGAAGATCAAAAAACA
ACTAATTATTCGAAACGATGAGATTTCCTTCAATTTTTACTGCTGTTTTATTCGCAGCATCCTCCGCA
TTAGCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCGGCTGAAGCTGTCATCGG
TTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGT
TATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGA
TATGATAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGA
GACGTATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGA
AGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTG
AAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGA
ATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAACTCAA
GTATACGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGGTTATATC
TACACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACATGTTTTGG
GACGGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATC
AAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGC
AAGAAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGG
CTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCAAATGCAT
TGAAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTCATCAGGG
TCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGG
AACGTATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGC
TGTCAGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGACATCTGCC
ACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTCCACCATCACCACCATCACTAAGCGGC
CGCCAGCTTTCTAGAACAAAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATC
ATCATCATTGAGTTTGTAGCCTTAGACATGACTGTTCCTCAGTTCAAGTTGGGCACTTACGAGAAGAC
CGGTCTTGCTAGATTCTAATCAAGAGGATGTCAGAATGCCATT

NOV46l, CG55794-05 Protein Sequence	SEQ ID NO: 722	354 aa	MW at 40431.3kD

YDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYL
KISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETKAVASFIES
KKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGTNYRVGSSADILYASSG
SSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLSVLDDVYAKHWHSDSAGRVTSA
TMLLGLLVSCMSLL

NOV46m, CG55794-06 DNA Sequence	SEQ ID NO: 723	977 bp
	ORF Start: ATG at 41	ORF Stop: TAG at 671

CCAGAGAGGCCCAGAATTTTCTAACTTACTGTGTGGCAGA**ATG**AAGCCTCTGCTTGAAACCCTTTATC
TTTTGGGGATGCTGGTTCCTGGAGGGCTGGGATATGATAGATCCTTAGCCCAACACAGACAAGAGATT
GTGGACAAGTCAGTGAGTCCATGGAGCCTGGAAACGTATTCCTATAACATATACCACCCCATGGGAGA
GATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAG
TGACCTATGAGACCCACCCCATATATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATC
ATTTGGATGGACTGTGGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAA
AGAAATTCTACAAAACCATAAAGACAACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATG
TCCTTCCAGTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCC
CGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCTTCTTGGTG
TAATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGA
TCGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCC
CTTCTCATATACGTTTGAGCTGAGGGACAGTGGAACATATGGGTTTGTTCTGCCAGAAGCTCAGATCC
AGCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGG
CACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTC
TCTTCTCTAAGTGCATCCTGCCCAG

NOV46m, CG55794-06 Protein Sequence	SEQ ID NO: 724	210 aa	MW at 24848.2kD

MKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYK

EVVTQHFLGVTYETHPIYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSRIR
KLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCNSSWTEGSKCIES
KVWNQL

NOV46n, CG55794-07 DNA Sequence	SEQ ID NO: 725	1378 bp
	ORF Start: ATG at 259	ORF Stop: TAA at 1225

ACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGTGAAAATACATCAGCATGTGGGAAAGAGCAACG
TTGATCGTCTTCACGGAAAGGCTGAGGACCCTGCGCCTACCACATGTTGGCCAGGGTGAGCAAGCAGT
GAAAGAGAAAACACTTTTTTCAAAAAGCCAACTGATCCTTAGCCCAACACAGACAAGAGATTGTGGAC
AAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGAGATCTA
TGAGTGGATGAGAGAGATCAGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCT
ATGAGACCCACCCCATATATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGG
ATGGACTGTGGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAAT
TCTACAAAACCATAAAGACAACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATGTCCTTC
CAGTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCA
CCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCTTCTTGGTGTAGTAT
TGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTA
AAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTAT
GGGCAGTTAATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCA
AGTTGGACAGAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTG
CAGATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCA
TATACGTTTGAGCTGAGGGACAGTGGAACATATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCAC
CTGTGAGGAGACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGG
ACAGTGCTGGAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTC
TAAGTGCATTCTGCCCAGGCCTGCTCAACCCCAGTGGCATGAGTGTGGCTGGAGGAACGGTGTGTTAT
GGTTGTAAAGAAACCAAATAATTTAACTAAAAATACTTCCTATTTCAATAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAA

NOV46n, CG55794-07 Protein Sequence	SEQ ID NO: 726	322 aa	MW at 36617.2kD

MGEIYEWMREISEKYKEVVTQHFLGVTYETHPIYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQW
FVKEILQNHKDNSRIRKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNA
SWCSIGASRNCQDQTFCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNH
PEMIQVGQKAANALKAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEA
QIQPTCEETMEAVLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46o, CG55794-08 DNA Sequence	SEQ ID NO: 727	1128 bp
	ORF Start: at 1	ORF Stop: TAA at 1126

ACCATGAAGCCTCTGCTTGAAACCCTTTATCTTTTGGGGATGCTGGTTCCTGGAGGGCTGGGATATGA
TAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGT
ATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTAC
AAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTGAAGAT
CAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGAATGGA
TTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAACTCAAGTATA
CGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGGTTATATCTACAC
TTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACATGTTTTGGGACGG
ATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACA
TTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAA
GGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGCTACA
CCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAAA
GCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTC
AAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGGAACGT
ATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCA
GTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTAT
GCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTC**TAA**

NOV46o, CG55794-08 Protein Sequence	SEQ ID NO: 728	375 aa	MW at 42630.1kD

TMKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKY
KEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSI
RKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQT
FCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALK
AKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLS
VLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46p, CG55794-09 DNA Sequence	SEQ ID NO: 729	1151 bp
	ORF Start: at 3	ORF Stop: TAA at 1149

CCACCATGGGCCACCATCACCACCATCACAAGCCTCTGCTTGAAACCCTTTATCTTTTGGGGATGCTG
GTTCCTGGAGGGCTGGGATATGATAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGT
GAGTCCATGGAGCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGA
TGAGAGAGATCAGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACC
CACCCCATGTATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTG
TGGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAA
ACCATAAAGACAACTCAAGTATACGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTT
AACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAA
TAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCT
CTAGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTT
GCCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTT
AATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGAC
AGAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATT
TTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTT
TGAGCTGAGGGACAGTGGAACGTATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGG
AGACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCT
GGAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTC**TAA**

NOV46p, CG55794-09 Protein Sequence	SEQ ID NO: 730	382 aa	MW at 43510.0kD

TMGHHHHHHKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWM
REISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQN
HKDNSSIRKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGAS
RNCQDQTFCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQ
KAANALKAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEE
TMEAVLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46q, CG55794-10 DNA Sequence	SEQ ID NO: 731	1148 bp
	ORF Start: at 3	ORF Stop: TAA at 1146

CCACCATGAAGCCTCTGCTTGAAACCCTTTATCTTTTGGGGATGCTGGTTCCTGGAGGGCTGGGATAT
GATAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGAC
GTATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGT
ACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTGAAG
ATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGAATG
GATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAACTCAAGTA
TACGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGGTTATATCTAC
ACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACATGTTTTGGGAC
GGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAA
CATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAG
AAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGCTA
CACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCAAATGCATTGA
AAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTCATCAGGGTCT
TCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGGAAC
GTATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGT
CAGTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGACATCTGCCACT
ATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTCCACCATCACCACCATCAC**TAA**

NOV46q, CG55794-10 Protein Sequence	SEQ ID NO: 732	381 aa	MW at 43452.9kD

TMKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKY
KEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSI
RKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQT
FCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALK
AKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLS
VLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLLHHHHHH

NOV46r, CG55794-11 DNA Sequence	SEQ ID NO: 733	1146 bp
	ORF Start: at 1	ORF Stop: TAA at 1144

ACCATGGTAAGCGCTATTGTTTTATATGTGCTTTTGGCGGCGGCGGCGCATTCTGCCTTTGCGTATGA
TAGATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGT
ATTCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTAC
AAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTGAAGAT
CAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGAATGGA
TTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAACTCAAGTATA
CGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGGTTATATCTACAC
TTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACATGTTTTGGGACGG
ATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACA
TTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAA
GGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGCTACA
CCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAAA
GCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTC
AAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGGAACGT
ATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCA
GTCCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTAT
GCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTCCACCATCACCACCATCAC**TAA**

NOV46r, CG55794-11 Protein Sequence	SEQ ID NO: 734	381 aa	MW at 43355.7kD

TMVSAIVLYVLLAAAAHSAFAYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKY
KEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSI
RKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQT
FCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALK
AKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLS
VLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLLHHHHHH

NOV46s, CG55794-12 DNA Sequence	SEQ ID NO: 735	1161 bp
	ORF Start: at 1	ORF Stop: TAA at 1159

ATTATTATACCTCCCACCATCGGGCGCGGATCCACCATGGTAAGCGCTATTGTTTTATATGTGCTTTT
GGCGGCGGCGGCGCATTCTGCCTTTGCGTATGATAGATCCTTAGCCCAACACAGACAAGAGATTGTGG
ACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGAGATC
TATGAGTGGATGAGAGAGATCAGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGAC
CTATGAGACCCACCCCATGTATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTT
GGATGGACTGTGGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAA
ATTCTACAAAACCATAAAGACAACTCAAGTATACGCAAGCTTCTTAGGAACCTGGACTTCTATGTCCT
TCCAGTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTT
CACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGT
ATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGAC
TAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCACTCTT
ATGGGCAGTTAATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATT
CAAGTTGGACAGAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAG
TGCAGATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCT
CATATACGTTTGAGCTGAGGGACAGTGGAACGTATGGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCC
ACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGCACTC
GGACAGTGCTGGAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTC
TCTAA

NOV46s, CG55794-12 Protein Sequence	SEQ ID NO: 736	386 aa	MW at 43638.1kD

IIIPPTIGRGSTMVSAIVLYVLLAAAAHSAFAYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEI
YEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKE
ILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCS
IGASRNCQDQTFCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMI
QVGQKAANALKAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQP
TCEETMEAVLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46t, CG55794-13 DNA Sequence	SEQ ID NO: 737	1212 bp
	ORF Start: at 1	ORF Stop: TAA at 1210

CGCGGATCCACCATGGTAAGCGCTATTGTTTTATATGTGCTTTTGGCGGCGGCGGCGCATTCTGCCTT
TGCGAAGCCTCTGCTTGAAACCCTTTATCTTTTGGGGATGCTGGTTCCTGGAGGGCTGGGATATGATA
GATCCTTAGCCCAACACAGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTAT
TCCTATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAA
GGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTGAAGATCA
GCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGAATGGATT
GCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAACTCAAGTATACG
CAAGCTTCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATGGTTATATCTACACTT
GGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGGCACATGTTTTGGGACGGAT
CTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATT
CTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGG
ATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGCTACACC
AAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAAAGC
AAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTCAA
GAGATTGGGCCCGAGACATTGGGATTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGGAACGTAT
GGGTTTGTTCTGCCAGAAGCTCAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGT
CCTGGATGATGTGTATGCGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTATGC
TGCTGGGCCTGCTGGTGTCCTGCATGTCTCTTCTCCACCATCACCACCATCAC**TAA**

NOV46t, CG55794-13 Protein Sequence	SEQ ID NO: 738	403 aa	MW at 45622.4kD

RGSTMVSAIVLYVLLAAAAHSAFAKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETY
SYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCGIHAREWI
APAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTD
LNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYT
KNKSSNHPEMIQVGQKAANALKAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTY
GFVLPEAQIQPTCEETMEAVLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLLHHHHHH

NOV46u, SNP13375362 of CG55794-03, DNA Sequence	SEQ ID NO: 739	1222 bp
	ORF Start: ATG at 41	ORF Stop: TAA at 1163
	SNP Pos: 240	SNP Change: A to G

CCAGAGAGGCCCAGAATTTTCTAACTTACTGTGTGGCAGAATGAAGCCTCTGCTTGAAACCCTTTATC
TTTTGGGGATGCTGGTTCCTGGAGGGCTGGGATATGATAGATCCTTAGCCCAACACAGACAAGAGATT
GTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGA
GATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTGCAAGGAAGTGGTGACACAGCATTTCCTAGGAG
TGACCTATGAGACCCACCCCATATATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATC
ATTTGGATGGACTGTGGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAA
AGAAATTCTACAAAACCATAAAGACAACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATG
TCCTTCCAGTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCC
CGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCTTCTTGGTG
TAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAG
AGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCAC
TCTTATGGGCAGTTAATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAAT
GATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGAT
CGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCC
TTCTCATATACGTTTGAGCTGAGGGACAGTGGAACATATGGGTTTGTTCTGCCAGAAGCTCAGATCCA

GCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGC
ACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCT
CTTCTC**TAA**GTGCATTCTGCCCAGGCCTGCTCAACCCCAGTGGCATGAGTGTGGCTTGGAGGAACG

NOV46u, SNP13375362 of CG55794-03, Protein Sequence	SEQ ID NO: 740	374 aa	MW at 42520.0kD
	SNP Pos: 67		SNP Change: Tyr to Cys

MKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKCK
EVVTQHFLGVTYETHPIYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSRIR
KLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTF
CGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKA
KYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLSV
LDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46v, SNP13379598 of CG55794-03, DNA Sequence	SEQ ID NO: 741	1222 bp
	ORF Start: ATG at 41	ORF Stop: TAA at 1163
	SNP Pos: 988	SNP Change: A to G

CCAGAGAGGCCCAGAATTTTCTAACTTACTGTGTGGCAGA**ATG**AAGCCTCTGCTTGAAACCCTTTATC
TTTTGGGGATGCTGGTTCCTGGAGGGCTGGGATATGATAGATCCTTAGCCCAACACAGACAAGAGATT
GTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGA
GATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAG
TGACCTATGAGACCCACCCCATATATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATC
ATTTGGATGGACTGTGGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAA
AGAAATTCTACAAAACCATAAAGACAACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATG
TCCTTCCAGTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCC
CGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCTTCTTGGTG
TAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAG
AGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCAC
TCTTATGGGCAGTTAATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAAT
GATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGAT
CGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCC
TTCTCATATACGTTTGAGCTGAGGGACAGTGGAACGTATGGGTTTGTTCTGCCAGAAGCTCAGATCCA
GCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGC
ACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCT
CTTCTC**TAA**GTGCATTCTGCCCAGGCCTGCTCAACCCCAGTGGCATGAGTGTGGCTTGGAGGAACG

NOV46v, SNP13379598 of CG55794-03, Protein Sequence	SEQ ID NO: 742	374 aa	MW at 42580.0kD
	SNP Pos: 316		SNP Change: Thr to Thr

MKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYK
EVVTQHFLGVTYETHPIYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSRIR
KLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTF
CGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKA
KYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSG**T**YGFVLPEAQIQPTCEETMEAVLSV
LDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL

NOV46w, SNP13375066 of CG55794-03, DNA Sequence	SEQ ID NO: 743	1222 bp
	ORF Start: ATG at 41	ORF Stop: TAA at 1163
	SNP Pos: 1152	SNP Change: T to C

CCAGAGAGGCCCAGAATTTTCTAACTTACTGTGTGGCAGA**ATG**AAGCCTCTGCTTGAAACCCTTTATC
TTTTGGGGATGCTGGTTCCTGGAGGGCTGGGATATGATAGATCCTTAGCCCAACACAGACAAGAGATT
GTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGA
GATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAG
TGACCTATGAGACCCACCCCATATATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATC
ATTTGGATGGACTGTGGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAA
AGAAATTCTACAAAACCATAAAGACAACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATG
TCCTTCCAGTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCC
CGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCTTCTTGGTG
TAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAG

AGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCAC
TCTTATGGGCAGTTAATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAAT
GATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGAT
CGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCC
TTCTCATATACGTTTGAGCTGAGGGACAGTGGAACATATGGGTTTGTTCTGCCAGAAGCTCAGATCCA
GCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGC
ACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCACGTCT
CTTCTCTAAGTGCATTCTGCCCAGGCCTGCTCAACCCCAGTGGCATGAGTGTGGCTTGGAGGAACG

NOV46w, SNP13375066 of CG55794-03, Protein Sequence	SEQ ID NO: 744	374 aa	MW at 42549.9kD
	SNP Pos: 371		SNP Change: Met to Thr

MKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYK
EVVTQHFLGVTYETHPIYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSRIR
KLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTF
CGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKA
KYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLSV
LDDVYAKHWHSDSAGRVTSATMLLGLLVSCTSLL

NOV46x, SNP13375067 of CG55794-03, DNA Sequence	SEQ ID NO: 745	1222 bp
	ORF Start: ATG at 41	ORF Stop: TAA at 1163
	SNP Pos: 1161	SNP Change: T to C

CCAGAGAGGCCCAGAATTTTCTAACTTACTGTGTGGCAGAATGAAGCCTCTGCTTGAAACCCTTTATC
TTTTGGGGATGCTGGTTCCTGGAGGGCTGGGATATGATAGATCCTTAGCCCAACACAGACAAGAGATT
GTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAACATATACCACCCCATGGGAGA
GATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAAGGAAGTGGTGACACAGCATTTCCTAGGAG
TGACCTATGAGACCCACCCCATATATTATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATC
ATTTGGATGGACTGTGGAATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAA
AGAAATTCTACAAAACCATAAAGACAACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATG
TCCTTCCAGTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCC
CGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCTTCTTGGTG
TAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAG
AGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCAC
TCTTATGGGCAGTTAATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAAT
GATTCAAGTTGGACAGAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGAT
CGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCC
TTCTCATATACGTTTGAGCTGAGGGACAGTGGAACATATGGGTTTGTTCTGCCAGAAGCTCAGATCCA
GCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGTCCTGGATGATGTGTATGCGAAACACTGGC
ACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCTGGTGTCCTGCATGTCT
CTTCCCTAAGTGCATTCTGCCCAGGCCTGCTCAACCCCAGTGGCATGAGTGTGGCTTGGAGGAACG

NOV46x, SNP13375067 of CG55794-03, Protein Sequence	SEQ ID NO: 746	374 aa	MW at 42564.0kD
	SNP Pos: 374		SNP Change: Leu to Pro

MKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYK
EVVTQHFLGVTYETHPIYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSRIR
KLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTF
CGTGPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKA
KYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLSV
LDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLP

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 46B.

Table 46B. Comparison of the NOV46 protein sequences.

```
NOV46a  ------------------------------------------------------------
NOV46b  LFETMRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFS
```

```
NOV46c  LFETMRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFS
NOV46d  ------------------------------------------------------------
NOV46e  ------------------------------------------------------------
NOV46f  ----------------------------------------------------------TM
NOV46g  ----------------------------------------------------------TM
NOV46h  ----------------------------------------------------------TM
NOV46i  ------------------------------------------------------------
NOV46j  ------------------------------------------------------------
NOV46k  ------------------------------------------------------------
NOV46l  ------------------------------------------------------------
NOV46m  ------------------------------------------------------------
NOV46n  ------------------------------------------------------------
NOV46o  ------------------------------------------------------------
NOV46p  ----------------------------------------------------------TM
NOV46q  ------------------------------------------------------------
NOV46r  ------------------------------------------------------------
NOV46s  ---------------------------------------------------IIIPPTI
NOV46t  ------------------------------------------------------------

NOV46a  ------MKPLLETLYLLGMLVPG-----------------GLGYDRSLAQHRQEIVDKS
NOV46b  NSTNNGLLFINTTIASIAAKEEGVSL--------------EKRYDRSLAQHRQEIVDKS
NOV46c  NSTNNGLLFINTTIASIAAKEEGVSL--------------EKRYDRSLAQHRQEIVDKS
NOV46d  ----TMKPLLETLYLLGMLVPGG-----------------LGYDRSLAQHRQEIVDKS
NOV46e  ----TMKPLLETLYLLGMLVPGG-----------------LGYDRSLAQHRQEIVDKS
NOV46f  GHHHHHHKPLLETLYLLGMLVPG-----------------GLGYDRSLAQHRQEIVDKS
NOV46g  GHHHHHHKPLLETLYLLGMLVPG-----------------GLGYDRSLAQHRQEIVDKS
NOV46h  GHHHHHHKPLLETLYLLGMLVPG-----------------GLGYDRSLAQHRQEIVDKS
NOV46i  ------MKPLLETLYLLGMLVPG-----------------GLGYDRSLAQHRQEIVDKS
NOV46j  ------------------------------------------YDRSLAQHRQEIVDKS
NOV46k  ------------------------------------------YDRSLAQHRQEIVDKS
NOV46l  ------------------------------------------YDRSLAQHRQEIVDKS
NOV46m  ------MKPLLETLYLLGMLVPG-----------------GLGYDRSLAQHRQEIVDKS
NOV46n  ------------------------------------------------------------
NOV46o  ----TMKPLLETLYLLGMLVPGG-----------------LGYDRSLAQHRQEIVDKS
NOV46p  GHHHHHHKPLLETLYLLGMLVPG-----------------GLGYDRSLAQHRQEIVDKS
NOV46q  ----TMKPLLETLYLLGMLVPGG-----------------LGYDRSLAQHRQEIVDKS
NOV46r  ----TMVSAIVLYVLLAAAAHS-----------------AFAYDRSLAQHRQEIVDKS
NOV46s  GRGSTMVSAIVLYVLLAAAAHS-----------------AFAYDRSLAQHRQEIVDKS
NOV46t  -RGSTMVSAIVLYVLLAAAAHSAFAKPLLETLYLLGMLVPGGLGYDRSLAQHRQEIVDKS

NOV46a  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPIYYLKISQPSGN
NOV46b  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46c  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46d  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46e  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46f  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46g  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46h  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46i  VSPWSLETYSYNIYHPMGEINEWMREISEKYKEVVTQHFLGVTYETHPIYYLKISQPSGN
NOV46j  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46k  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46l  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46m  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPIYYLKISQPSGN
NOV46n  ---------------MGEIYEWMREISEKYKEVVTQHFLGVTYETHPIYYLKISQPSGN
```

```
NOV46o  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46p  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46q  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46r  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46s  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN
NOV46t  VSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGN

NOV46a  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSRIRKLLRNLDFYVLPVLNIDGYI
NOV46b  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46c  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46d  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46e  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46f  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46g  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46h  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46i  PKKIIWMGCGIHAREWIAPAFCQWFVKEILQNHKDNSRIRKLLRNLDFYVLPVLNIDGYI
NOV46j  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46k  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46l  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46m  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSRIRKLLRNLDFYVLPVLNIDGYI
NOV46n  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSRIRKLLRNLDFYVLPVLNIDGYI
NOV46o  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46p  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46q  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46r  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46s  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI
NOV46t  PKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYI

NOV46a  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46b  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46c  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46d  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46e  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46f  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46g  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46h  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46i  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46j  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46k  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46l  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46m  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCNSSWTEGSKCIESKVWNQL------
NOV46n  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46o  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46p  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46q  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46r  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46s  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK
NOV46t  YTWTTDRLWRKSRSPHNNGTCFGTDLNRNFNASWCSIGASRNCQDQTFCGTGPVSEPETK

NOV46a  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46b  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46c  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46d  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46e  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
```

```
NOV46f  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46g  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46h  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46i  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46j  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46k  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46l  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46m  ------------------------------------------------------------
NOV46n  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46o  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46p  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46q  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46r  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46s  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT
NOV46t  AVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGT

NOV46a  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46b  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46c  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46d  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46e  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46f  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46g  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46h  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46i  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46j  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46k  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46l  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46m  ------------------------------------------------------------
NOV46n  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46o  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46p  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46q  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46r  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46s  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA
NOV46t  NYRVGSSADILYASSGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEA

NOV46a  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46b  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46c  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLLHHHHHH
NOV46d  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46e  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLLHHHHHH
NOV46f  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46g  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46h  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46i  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46j  VLSVLDDVYAKHWHSDSAGR----------------------
NOV46k  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46l  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46m  ------------------------------------------
NOV46n  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46o  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46p  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46q  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLLHHHHHH
```

```
NOV46r  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLLHHHHHH
NOV46s  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLL------
NOV46t  VLSVLDDVYAKHWHSDSAGRVTSATMLLGLLVSCMSLLHHHHHH

NOV46a  (SEQ ID NO:  700)
NOV46b  (SEQ ID NO:  702)
NOV46c  (SEQ ID NO:  704)
NOV46d  (SEQ ID NO:  706)
NOV46e  (SEQ ID NO:  708)
NOV46f  (SEQ ID NO:  710)
NOV46g  (SEQ ID NO:  712)
NOV46h  (SEQ ID NO:  714)
NOV46i  (SEQ ID NO:  716)
NOV46j  (SEQ ID NO:  718)
NOV46k  (SEQ ID NO:  720)
NOV46l  (SEQ ID NO:  722)
NOV46m  (SEQ ID NO:  724)
NOV46n  (SEQ ID NO:  726)
NOV46o  (SEQ ID NO:  728)
NOV46p  (SEQ ID NO:  730)
NOV46q  (SEQ ID NO:  732)
NOV46r  (SEQ ID NO:  734)
NOV46s  (SEQ ID NO:  736)
NOV46t  (SEQ ID NO:  738)
```

Further analysis of the NOV46a protein yielded the following properties shown in Table 46C.

Table 46C. Protein Sequence Properties NOV46a

SignalP analysis:	Cleavage site between residues 21 and 22

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 6;  pos.chg 1;  neg.chg 1
      H-region:  length 15;  peak value   9.15
      PSG score:   4.75

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -0.52
      possible cleavage site: between 20 and 21

>>> Seems to have a cleavable signal peptide (1 to 20)

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 21
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -2.34   Transmembrane  357 - 373
      PERIPHERAL  Likelihood =  4.08 (at 229)
      ALOM score:  -2.34  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
```

```
      Center position for calculation: 10
      Charge difference: -0.5   C( 0.5) - N( 1.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 1a  (cytoplasmic tail 374 to 374)

>>> Seems to be GPI anchored

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  8.63
      Hyd Moment(95): 12.49    G content:       4
      D/E content:     2       S/T content:     1
      Score: -7.55

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   9.9%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none
```

```
checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 70.6

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: extracellular, including cell wall
       22.2 %: endoplasmic reticulum
       11.1 %: Golgi
       11.1 %: plasma membrane

>> prediction for CG55794-03 is exc (k=9)
```

A search of the NOV46a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 46D.

Table 46D. Geneseq Results for NOV46a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV46a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAG66547	Human secreted metallocarboxypeptidase-like polypeptide - Homo sapiens, 374 aa. [WO200157265-A1, 09-AUG-2001]	1..374 1..374	374/374 (100%) 374/374 (100%)	0.0
AAG66565	Human secreted metallocarboxypeptidase-like variant polypeptide - Homo sapiens, 374 aa. [WO200157265-A1, 09-AUG-2001]	1..374 1..374	373/374 (99%) 374/374 (99%)	0.0
AAU82703	Amino acid sequence of novel human protease #2 - Homo sapiens, 374 aa. [WO200200860-A2, 03-JAN-2002]	1..374 1..374	372/374 (99%) 373/374 (99%)	0.0
AAB74682	Human protease and protease inhibitor PPIM-15 - Homo sapiens, 362 aa. [WO200110903-A2, 15-FEB-2001]	13..374 1..362	360/362 (99%) 361/362 (99%)	0.0
AAG66560	Human secreted metallocarboxypeptidase-like polypeptide - Homo sapiens, 354 aa. [WO200157265-A1, 09-AUG-2001]	21..374 1..354	354/354 (100%) 354/354 (100%)	0.0

In a BLAST search of public sequence databases, the NOV46a protein was found to have homology to the proteins shown in the BLASTP data in Table 46E.

Table 46E. Public BLASTP Results for NOV46a				
Protein Accession Number	**Protein/Organism/Length**	**NOV46a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8IVL8	Zn-carboxypeptidase - Homo sapiens (Human), 374 aa.	1..374 1..374	372/374 (99%) 373/374 (99%)	0.0
Q9PUF2	Carboxypeptidase homolog - Bothrops jararaca (Jararaca), 416 aa.	13..337 82..405	152/326 (46%) 219/326 (66%)	7e-87
Q8AXN3	Carboxypeptidase B - Paralichthys olivaceus (Flounder), 408 aa (fragment).	41..337 100..396	146/297 (49%) 203/297 (68%)	8e-86
CAA03381	SEQUENCE 85 FROM PATENT WO9620011 - unidentified, 349 aa.	45..352 26..336	148/312 (47%) 204/312 (64%)	1e-85
Q9JHH6	Carboxypeptidase R (Thrombin-activatable fibrinolysis inhibitor) (1110032P04Rik protein) - Mus musculus (Mouse), 422 aa.	30..350 103..417	145/323 (44%) 210/323 (64%)	2e-85

PFam analysis indicates that the NOV46a protein contains the domains shown in the Table 46F.

Table 46F. Domain Analysis of NOV46a			
Pfam Domain	**NOV46a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Zn_carbOpept	48..249	82/215 (38%) 168/215 (78%)	1.2e-74

Example 47.

The NOV47 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 47A.

Table 47A. NOV47 Sequence Analysis

NOV47a, CG55806-04 DNA Sequence	SEQ ID NO: 747	1300 bp
	ORF Start: ATG at 20	ORF Stop: TGA at 1274

TGACTGTATCGCCGGATTC**ATG**CAGCGCGTGAACATGATCATGGCAGAATCACCAGGCCTCATCACCA
TCTGCCTTTTAGGATATCTACTCAGTGCTGAATGTACAGTTTTTCTTGATCATGAAAACGCCAACAAA
ATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAGGAGTTTGTTCATGGGAACCTTGAGAG
AGAATGTATGGAAGAAAAGTGCAGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGAAAGAACAA
CTGAATTTTGGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGT
TGCAAGGATGTCATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATT
AGTGCCATTTCCATGTGGAAGAGTTTCTGTTTCACAAACTTCTAAGCTCACCCGTGCTGAGACTGTTT
TTCCTGATGTGGACTATGTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACC
CAATCATTTAATGACTTCACTCGGGTTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCCTTGGCA
GGTTGTTTTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAA
CTGCTGCCCACTGTGTTGAAACTGGTGTTAAAATTACAGTTGTCGCAGGTGAACATAATATTGAGGAG
ACAGAACATACAGAGCAAAAGCGAAATGTGATTCGAATTATTCCTCACCACAACTACAATGCAGCTAT
TAATAAGTACAACCATGACATTGCCCTTCTGGAACTGGACGAACCCTTAGTGCTAAACAGCTACGTTA
CACCTATTTGCATTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTAAGT
GGCTGGGGAAGAGTCTTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGT
TGACCGAGCCACATGTCTTCGATCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCC
ATGAAGGAGGTAGAGATTCATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGGAAGGGACC
AGTTTCTTAACTGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATAC
CAAGGTATCCCGGTATGTCAACTGGATTAAGGAAAAAACAAAGCTCACTTGACTGCAGCCAAGCTAAT
TCCGGAAG

NOV47a, CG55806-04 Protein Sequence	SEQ ID NO: 748	418 aa	MW at 47031.8kD

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVHGNLERECMEEK
CSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDVINSYECWCPFGFEGKNCELVPFPCG
RVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGK
VDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHD
IALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCL
RSTKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYV
NWIKEKTKLT

NOV47b, CG55806-01 DNA Sequence	SEQ ID NO: 749	1438 bp
	ORF Start: ATG at 2	ORF Stop: TAA at 1184

T**ATG**CAGCGCGTGAACATGATCATGGCAGAATCACCAGGCCTCATCACCATCTGCCTTTTAGGATATC
TACTCAGTGCTGAATGTACAGTTTTTCTTGATCATGAAAACGCCAACAAAATTCTGAATCGGCCAAAG
AGGTATAATTCAGGTAAATTGGAAGAGTTTGTTCAAGGGAACCTTGAGAGAGAATGTCTGGAGGAAAA
GTGTAGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGAAAGAACAACTGAATTTTGGAAGCAGT
ATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGTTGCAAGGATGACATTAAT
TCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATTAGATGTGGACTATGTAAA
TTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCACTC
GGGTTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCCTTGGCAGGTTGTTTTGAATGGTAAAGTT
GATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAACTGCTGCCCACTGTGTTGAAAC
TGGTGTTAAAATTACAGTTGTCGCAGGTGAACATAATATTGAGGAGACAGAACATACAGAGCAAAAGC
GAAATGTGATTCGAATTATTCCTCACCACAACTACAATGCAGCTATTAATAAGTACAACCATGACATT
GCCCTTCTGGAACTGGACGAACCCTTAGTGCTAAACAGCTACGTTACACCTATTTGCATTGCTGACAA
GGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTAAGTGGCTGGGGAAGAGTCTTCCACA
AAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACCGAGCCACATGTCTTCGA
TCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAGATTCATG
TCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGGAAGGGACCAGTTTCTTAACTGGAATTATTA
GCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATACCAAGGTATCCCGGTATGTCAAC
TGGATTAAGGAAAAAACAAAGCTCACT**TAA**TGAAAGATGGATTTCCAAGGTTAATTCATTGGAATTGA
AAATTAACAGGGCCTCTCACTAACTAATCACTTTCCCATCTTTTGTTAGATTTGAATATATACATTCT
ATGATCATTGCTTTTTCTCTTTACAGGGGAGAATTTCATATTTTACCTGAGCAAATTGATTAGAAAAT

GGAACCACTAGAGGAATATAATGTGTTAGGAAATTACAGTCATTTCTAAGGGCCCAGCCTTGACAAAT
TGTGAGTAAA

NOV47b, CG55806-01 Protein Sequence	SEQ ID NO: 750	394 aa	MW at 44431.7kD

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECLEEK
CSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVDYVN
STEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVET
GVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHDIALLELDEPLVLNSYVTPICIADK
EYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSC
QGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

NOV47c, CG55806-02 DNA Sequence	SEQ ID NO: 751	2807 bp
	ORF Start: ATG at 20	ORF Stop: TAA at 1403

CAATCTGCTAGCAAAGGTTATGCAGCGCGTGAACATGATCATGGCAGAATCACCAGGCCTGATCACCA
TCTGCCTTTTAGGATATCTACTCAGTGCTGAATGTACAGTTTTTCTTGATCATGAAAACGCCAACAAA
ATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAAGAGTTTGTTCAAGGGAACCTTGAGAG
AGAATGTATGGAAGAAAAGTGTAGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGAAAGAACAA
CTGAATTTTGGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGT
TGCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATT
AGATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAGTGCTGATAACAAGG
TGGTTTGCTCCTGTACTGAGGGATATCGACTTGCAGAAAACCAGAAGTCCTGTGAACCAGCAGTGCCA
TTTCCATGTGGAAGAGTTTCTGTTTCACAAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTTCCTGA
TGTGGACTATGTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCCAATCAT
TTAATGACTTCACTCGGGTTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCCTTGGCAGGTTGTT
TTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAACTGCTGC
CCACTGTGTTGAAACTGGTGTTAAAATTACAGTTGTCGCAGGTGAACATAATATTGAGGAGACAGAAC
ATACAGAGCAAAAGCGAAATGTGATTCGAATTATTCCTCACCACAACTACAATGCAGCTATTAATAAG
TACAACCATGACATTGCCCTTCTGGAACTGGACGAACCCTTAGTGCTAAACAGCTACGTTACACCTAT
TTGCATTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTAAGTGGCTGGG
GAAGAGTCTTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACCGA
GCCACATGTCTTCGATCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGG
AGGTAGAGATTCATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGGAAGGGACCAGTTTCT
TAACTGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATACCAAGGTA
TCCCGGTATGTCAACTGGATTAAGGAAAAAACAAAGCTCACTTAATGAAAGATGGATTTCCAAGGTTA
ATTCATTGGAATTGAAAATTAACAGGGCCTCTCACTAACTAATCACTTTCCCATCTTTTGTTAGATTT
GAATATATACATTCTATGATCATTGCTTTTTCTCTTTACAGGGGAGAATTTCATATTTTACCTGAGCA
AATTGATTAGAAAATGGAACCACTAGAGGAATATAATGTGTTAGGAAATTACAGTCATTTCTGAAGCC
CAGCCCTTGACAAAATTGTGAAGTTAAATTCTCCACTCTGTCCATCAGATACTATGGTTCTCCACTAT
GGCAACTAACTCACTCAATTTTCCCTCCTTAGCAGCATTCCATCTTCCCGATCTTCTTTGCTTCTCCA
ACCAAAACATCATGTTTATTAGTTCTGTATACAGTACAGGATCTTTGGTCTACTCTATCACGAGAAGG
CTCAGTACCACACTCATGAAGAAAGAACACAGGAGTAGCTGAGAGGCTAAAACTCATCAAAAACACTA
CTCCTTTTCCTCTACCCTATTCCTCAATCTTTTACCTTTTCCAAATTCCCAATTCCCCAAATCAGTTT
TTCTCTTTCTTACTCCCTCTCTCCCTTTTACCCTCCATGGTCGTTAAAGGAGAGATGGGGAGCATCAT
TCTGTTATACTTCTGTACACAGTTATACATGTCTATCAAACCCAGACTTGCTTCCATAGTGGAGACTT
GCTTTTCAGAACATAGGGATGAAGTAAGGTGCCTGAAAAGTTTGGGGGAAAAGTTTCTTTCAGAGAGT
TAAGTTATTTTATATATATAATATATATATAAAATATATAATATACAATATAAATATATAGTGTGTGT
GTGTATGCGTGTGTGTAGACACACACGCATACACACATATAATGGAAGCAATAAGCCATTCTTAAGAG
CTTGTATGGTTATGCAGGTCTGACTAGGCATGATTTCACGAAGGCAAGATTGGCATATCAGTTGTAAC
TAAAAAAGCTGACATTGACCCAGACATATTGTACTCTTTCTAAAAATAATAATAATAATGCTAACAGA
AAGAAGAGAACCGTTCGTTTGCAATCTACAGCTAGTAGAGACTTGAGGAAGAATTCAACAGTGTGTCT
TCAGCAGTGTTCAGAGCCAAGCAAGAAGTTGAAGTTGCCTAGACCAGAGGACATAAGTATCATGTCTC
CTTTAACTAGCATACCCCGAAGTGGAGAAGGATGCAGCAGGCTCAAAGGCATAAGTCATTCCAATCAG
CCAACTAAGTTGTCCTTTTCTGGTTTCGTGTTCACCATGGAACATTTTGATTATAGTTAATCCTTCTA
TCTTGAATCTTCTAGAGAGTTGCTGACCAACTGACGTATGTTTCCCTTTGTGAATTAATAAACTGGTG
TTCTGGTTCAAAAAAAAAA

NOV47c, CG55806-02 Protein Sequence	SEQ ID NO: 752	461 aa	MW at 51778.0kD

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEEK
CSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNI
KNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNS
TEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETG
VKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHDIALLELDEPLVLNSYVTPICIADKE
YTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSCQ
GDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

NOV47d, CG55806-03 DNA Sequence	SEQ ID NO: 753	1612 bp
	ORF Start: ATG at 22	ORF Stop: TAA at 1405

TTGACTGTATCGCCGGAATTC**ATG**CAGCGCGTGAACATGATCATGGCAGAATCACCAGGCCTCATCAC
CATCTGCCTTTTAGGATATCTACTCAGTGCTGAATGTACAGTTTTTCTTGATCATGAAAACGCCAACA
AAATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAAGAGTTTGTTCAAGGGAACCTTGAG
AGAGAATGTATGGAAGAAAAGTGTAGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGAAAGAAC
AACTGAATTTTGGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCA
GTTGCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAA
TTAGATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAGTGCTGATAACAA
GGTGGTTTGCTCCTGTACTGAGGGATATCGACTTGCAGAAAACCAGAAGTCCTGTGAACCAGCAGTGC
CATTTCCATGTGGAAGAGTTTCTGTTTCACAAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTTCCT
GATGTGGACTATGTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCCAATC
ATTTAATGACTTCACTCGGGTTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCCTTGGCAGGTTG
TTTTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAACTGCT
GCCCACTGTGTTGAAACTGGTGTTAAAATTACAGTTGTCGCAGGTGAACATAATATTGAGGAGACAGA
ACATACAGAGCAAAAGCGAAATGTGATTCGAATTATTCCTCACCACAACTACAATGCAGCTATTAATA
AGTACAACCATGACATTGCCCTTCTGGAACTGGACGAACCCTTAGTGCTAAACAGCTACGTTACACCT
ATTTGCATTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTAAGTGGCTG
GGGAAGAGTCTTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACC
GAGCCACATGTCTTCGATCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAA
GGAGGTAGAGATTCATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGGAAGGGACCAGTTT
CTTAACTGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATACCAAGG
TATCCCGGTATGTCAACTGGATTAAGGAAAAAACAAAGCTCACT**TAA**TGAAAGATGGATTTCCAAGGT
TAATTCATTGAAATTGAAAATTAATAGGGCCTCTCACTAACTAATCACTTTCCCATCTTTTGTTAGAT
TTGAATATATACATTCTATGATCATTGCTTTTTCTCTTTACAGGGGAGAATTTCATATTTTACCTGAG
CAAATTGATTAGAAAATGGAACCACTAGAGGAATATAATGTGTTAGGA

NOV47d, CG55806-03 Protein Sequence	SEQ ID NO: 754	461 aa	MW at 51778.0kD

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEEK
CSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVTCNI
KNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVDYVNS
TEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVETG
VKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHDIALLELDEPLVLNSYVTPICIADKE
YTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSCQ
GDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

NOV47e, SNP13382503 of CG55806-04, DNA Sequence	SEQ ID NO: 755	1300 bp
	ORF Start: ATG at 20	ORF Stop: TGA at 1274
	SNP Pos: 470	SNP Change: A to G

TGACTGTATCGCCGGATTC**ATG**CAGCGCGTGAACATGATCATGGCAGAATCACCAGGCCTCATCACCA
TCTGCCTTTTAGGATATCTACTCAGTGCTGAATGTACAGTTTTTCTTGATCATGAAAACGCCAACAAA
ATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAGGAGTTTGTTCATGGGAACCTTGAGAG
AGAATGTATGGAAGAAAAGTGCAGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGAAAGAACAA
CTGAATTTTGGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGT
TGCAAGGATGTCATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATT

AGTGCCATTTCCATGTGGAAGAGTTTCTGTTTCACAAACTTCTAAGCTCACCCGTGCTGAGGCTGTTT
TTCCTGATGTGGACTATGTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACC
CAATCATTTAATGACTTCACTCGGGTTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCCTTGGCA
GGTTGTTTTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAA
CTGCTGCCCACTGTGTTGAAACTGGTGTTAAAATTACAGTTGTCGCAGGTGAACATAATATTGAGGAG
ACAGAACATACAGAGCAAAAGCGAAATGTGATTCGAATTATTCCTCACCACAACTACAATGCAGCTAT
TAATAAGTACAACCATGACATTGCCCTTCTGGAACTGGACGAACCCTTAGTGCTAAACAGCTACGTTA
CACCTATTTGCATTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTAAGT
GGCTGGGGAAGAGTCTTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGT
TGACCGAGCCACATGTCTTCGATCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCC
ATGAAGGAGGTAGAGATTCATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGGAAGGGACC
AGTTTCTTAACTGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATAC
CAAGGTATCCCGGTATGTCAACTGGATTAAGGAAAAAACAAAGCTCACTTGACTGCAGCCAAGCTAAT
TCCGGAAG

NOV47e, SNP13382503 of CG55806-04, Protein Sequence	SEQ ID NO: 756	418 aa	MW at 47001.8kD
	SNP Pos: 151		SNP Change: Thr to Ala

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVHGNLERECMEEK
CSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDVINSYECWCPFGFEGKNCELVPFPCG
RVSVSQTSKLTRAEAVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGK
VDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHD
IALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCL
RSTKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYV
NWIKEKTKLT

NOV47f, SNP13382492 of CG55806-04, DNA Sequence	SEQ ID NO: 757	1300 bp
	ORF Start: at 2	ORF Stop: at 1253
	SNP Pos: 673	SNP Change: G to A

TGACTGTATCGCCGGATTCATGCAGCGCGTGAACATGATCATGGCAGAATCACCAGGCCTCATCACCA
TCTGCCTTTTAGGATATCTACTCAGTGCTGAATGTACAGTTTTTCTTGATCATGAAAACGCCAACAAA
ATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAGGAGTTTGTTCATGGGAACCTTGAGAG
AGAATGTATGGAAGAAAAGTGCAGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGAAAGAACAA
CTGAATTTTGGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGT
TGCAAGGATGTCATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATT
AGTGCCATTTCCATGTGGAAGAGTTTCTGTTTCACAAACTTCTAAGCTCACCCGTGCTGAGACTGTTT
TTCCTGATGTGGACTATGTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACC
CAATCATTTAATGACTTCACTCGGGTTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCCTTGGCA
GGTTGTTTTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGAATTGTAA
CTGCTGCCCACTGTGTTGAAACTGGTGTTAAAATTACAGTTGTCGCAGGTGAACATAATATTGAGGAG
ACAGAACATACAGAGCAAAAGCGAAATGTGATTCGAATTATTCCTCACCACAACTACAATGCAGCTAT
TAATAAGTACAACCATGACATTGCCCTTCTGGAACTGGACGAACCCTTAGTGCTAAACAGCTACGTTA
CACCTATTTGCATTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTAAGT
GGCTGGGGAAGAGTCTTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGT
TGACCGAGCCACATGTCTTCGATCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCC
ATGAAGGAGGTAGAGATTCATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGGAAGGGACC
AGTTTCTTAACTGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATAC
CAAGGTATCCCGGTATGTCAACTGGATTAAGGAAAAAACAAAGCTCACTTGACTGCAGCCAAGCTAAT
TCCGGAAG

NOV47f, SNP13382492 of CG55806-04, Protein Sequence	SEQ ID NO: 758	417 aa	MW at 46845.6kD
	SNP Pos: 218		SNP Change: Trp to End

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVHGNLERECMEEK
CSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDVINSYECWCPFGFEGKNCELVPFPCG
RVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGK
VDAFCGGSIVNEKIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHDI
ALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLR
STKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVN
WIKEKTKLT

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 47B.

Table 47B. Comparison of the NOV47 protein sequences.

```
NOV47a  MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVHGNL
NOV47b  MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNL
NOV47c  MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNL
NOV47d  MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNL

NOV47a  ERECMEEKCSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDVINSYECWCP
NOV47b  ERECLEEKCSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCP
NOV47c  ERECMEEKCSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCP
NOV47d  ERECMEEKCSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCP

NOV47a  FGFEGKNCELVP----------------------------------------FP--CGR
NOV47b  FGFEGKNCELDV----------------------------------------D------
NOV47c  FGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGR
NOV47d  FGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGR

NOV47a  VSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPW
NOV47b  ------------------YVNSTEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPW
NOV47c  VSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPW
NOV47d  VSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPW

NOV47a  QVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRII
NOV47b  QVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRII
NOV47c  QVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRII
NOV47d  QVVLNGKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRII

NOV47a  PHHNYNAAINKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVF
NOV47b  PHHNYNAAINKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVF
NOV47c  PHHNYNAAINKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVF
NOV47d  PHHNYNAAINKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVF

NOV47a  HKGRSALVLQYLRVPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVE
NOV47b  HKGRSALVLQYLRVPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVE
NOV47c  HKGRSALVLQYLRVPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVE
NOV47d  HKGRSALVLQYLRVPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVE

NOV47a  GTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT
NOV47b  GTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT
NOV47c  GTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT
NOV47d  GTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT

NOV47a  (SEQ ID NO:  748)
NOV47b  (SEQ ID NO:  750)
NOV47c  (SEQ ID NO:  752)
NOV47d  (SEQ ID NO:  754)
```

Further analysis of the NOV47a protein yielded the following properties shown in Table 47C.

Table 47C. Protein Sequence Properties NOV47a	
SignalP analysis:	Cleavage site between residues 26 and 27
PSORT II analysis:	

```
PSG:  a new signal peptide prediction method
      N-region:  length 10;  pos.chg 1;  neg.chg 1
      H-region:  length 16;  peak value   9.84
      PSG score:   5.44

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -2.46
      possible cleavage site: between 25 and 26

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -2.50   Transmembrane   14 -  30
      PERIPHERAL  Likelihood =  2.07 (at 219)
      ALOM score:  -2.50  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 21
      Charge difference: -2.5   C(-1.5) - N( 1.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 2   (cytoplasmic tail 1 to 14)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  4.83
      Hyd Moment(95):  6.44    G content:       2
      D/E content:     2       S/T content:     3
      Score: -6.32

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 13  QRV|NM

NUCDISC: discrimination of nuclear localization signals
      pat4: RPKR (4) at   43
      pat7: none
      bipartite: none
      content of basic residues:  10.0%
      NLS Score: -0.22

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: QRVN
```

```
      KKXX-like motif in the C-terminus: KTKL

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
  58 N  0.51
  59 L  0.51
  60 E  0.51
  61 R  0.51
  62 E  0.51
  63 C  0.51
  64 M  0.51
  65 E  0.51
  66 E  0.51
  67 K  0.51
  68 C  0.51
  69 S  0.51
  70 F  0.51
  71 E  0.51
  72 E  0.51
  73 A  0.51
  74 R  0.51
  75 E  0.51
  76 V  0.51
```

```
  77 F   0.51
  78 E   0.51
  79 N   0.51
  80 T   0.51
  81 E   0.51
  82 R   0.51
  83 T   0.51
  84 T   0.51
  85 E   0.51
      total: 28 residues

--------------------------

Final Results (k = 9/23):

       34.8 %: mitochondrial
       30.4 %: cytoplasmic
        8.7 %: Golgi
        8.7 %: nuclear
        4.3 %: vacuolar
        4.3 %: extracellular, including cell wall
        4.3 %: vesicles of secretory system
        4.3 %: endoplasmic reticulum

>> prediction for CG55806-04 is mit (k=23)
```

A search of the NOV47a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 47D.

Table 47D. Geneseq Results for NOV47a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV47a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABB99529	Amino acid sequence of human Factor IX - Homo sapiens, 461 aa. [WO200286091-A2, 31-OCT-2002]	1..418 1..461	416/461 (90%) 416/461 (90%)	0.0
ABB81908	Protein relating to the invention #1 - Unidentified, 461 aa. [RU2181147-C2, 10-APR-2002]	1..418 1..461	416/461 (90%) 416/461 (90%)	0.0
AAY97295	Human clotting factor IX - Homo sapiens, 461 aa. [WO200049147-A1, 24-AUG-2000]	1..418 1..461	416/461 (90%) 416/461 (90%)	0.0
AAO21524	Protein of human factor IX - Homo sapiens, 461 aa. [WO200240544-A2, 23-MAY-2002]	1..418 1..461	408/462 (88%) 409/462 (88%)	0.0
ABP53571	Human NOV3 protein SEQ ID NO:6 - Homo sapiens, 394 aa. [WO200262999-A2, 15-AUG-2002]	1..418 1..394	391/418 (93%) 392/418 (93%)	0.0

In a BLAST search of public sequence databases, the NOV47a protein was found to have homology to the proteins shown in the BLASTP data in Table 47E.

Table 47E. Public BLASTP Results for NOV47a				
Protein Accession Number	**Protein/Organism/Length**	**NOV47a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
CAA00205	FACTOR IX - Homo sapiens (Human), 461 aa.	1..418 1..461	416/461 (90%) 416/461 (90%)	0.0
Q95ND7	Coagulation factor XI - Pan troglodytes (Chimpanzee), 461 aa.	1..418 1..461	415/461 (90%) 416/461 (90%)	0.0
P00740	Coagulation factor IX precursor (EC 3.4.21.22) (Christmas factor) - Homo sapiens (Human), 461 aa.	1..418 1..461	415/461 (90%) 415/461 (90%)	0.0
Q14316	F9 (Coagulation factor IX (Plasma THROMBOPLASTIC component, christmas disease,HAEMOPHILIA B)) (Factor IX) - Homo sapiens (Human), 456 aa.	6..418 1..456	411/456 (90%) 411/456 (90%)	0.0
CAA01607	FACTOR IX PROTEIN - Homo sapiens (Human), 456 aa.	6..418 1..456	408/456 (89%) 408/456 (89%)	0.0

PFam analysis indicates that the NOV47a protein contains the domains shown in the Table 47F.

Table 47F. Domain Analysis of NOV47a			
Pfam Domain	**NOV47a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
gla	52..93	26/42 (62%) 40/42 (95%)	1.1e-18
EGF	97..128	16/47 (34%) 25/47 (53%)	1.9e-06
trypsin	184..411	109/263 (41%) 197/263 (75%)	4.4e-92

Example 48.

The NOV48 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 48A.

Table 48A. NOV48 Sequence Analysis

NOV48a, CG55828-02 DNA Sequence	SEQ ID NO: 759	1439 bp
	ORF Start: ATG at 35	ORF Stop: TGA at 1406

TCATCCACAGCCATCATATAAAGGTTTTGGCATCATGTTTGGGAAGAAAAAGAAAAAGATTGAAATAT
CTGGCCCGTCCAACTTTGAACACAGGGTTCATACTGGGTTTGATCCACAAGAGCAGAAGTTTACCGGC
CTTCCCCAGCAGTGGCACAGCCTGTTAGCAGATACGGCCAACAGGCCAAAGCCTATGGTGGACCCTTC
ATGCATCACACCCATCCAGCTGGCTCCTATGAAGGTGGATTACGATCGAGCACAGATGGTCCTCAGCC
CTCCACTGTCAGGGTCTGACACCTACCCCAGGGGCCCTGCCAAACTACCTCAAAGTCAAAGCAAATCG
GGCTATTCCTCAAGCAGTCACCAGTACCCGTCTGGGTACCACAAAGCCACCTTGTACCATCACCCCTC
CCTGCAGAGCAGTTCGCAGTACATCTCCACGGCTTCCTACCTGAGCTCCCTCAGCCTCTCATCCAGCA
CCTACCCGCCGCCCAGCTGGGGCTCCTCCTCCGACCAGCAGCCCTCCAGGGTGTCCCATGAACAGTTT
CGGGCGGCCCTGCAGCTGGTGGTCAGCCCAGGAGACCCCAGGGAATACTTGGCCAACTTTATCAAAAT
CGGGGAAGGCTCAACCGGCATCGTATGCATCGCCACCGAGAAACACACAGGGAAACAAGTTGCAGTGA
AGAAAATGGACCTCCGGAAGCAACAGAGACGAGAACTGCTTTTCAATGAGGTCGTGATCATGCGGGAT
TACCACCATGACAATGTGGTTGACATGTACAGCAGCTACCTTGTCGGCGATGAGCTCTGGGTGGTCAT
GGAGTTTCTAGAAGGTGGTGCCTTGACAGACATTGTGACTCACACCAGAATGAATGAAGAACAGATAG
CTACTGTCTGCCTGTCAGTTCTGAGAGCTCTCTCCTACCTTCATAACCAAGGAGTGATTCACAGGGAC
ATAAAAAGTGACTCCATTCTCCTGACAAGCGATGGCCGGATAAAGTTGTCTGATTTTGGTTTCTGTGC
TCAAGTTTCCAAAGAGGTGCCGAAGAGGAAATCATTGGTTGGCACTCCCTACTGGATGGCCCCTGAGG
TGATTTCTAGGCTACCTTATGGGACAGAGGTGGACATCTGGTCCCTCGGGATCATGGTGATAGAAATG
ATTGATGGCGAGCCCCCCTACTTCAATGAGCCTCCCCTCCAGGCGATGCGGAGGATCCGGGACAGTTT
ACCTCCAAGAGTGAAGGACCTACACAAGGTTTCTTCAGTGCTCCGGGGATTCCTAGACTTGATGTTGG
TGAGGGAGCCCTCTCAGAGAGCAACAGCCCAGGAACTCCTCGGACATCCATTCTTAAAACTAGCAGGT
CCACCGTCTTGCATCGTCCCCCTCATGAGACAATACAGGCATCACTGAGCAGAGGATTCGTGTAGGTG
GCAAAGCTAGA

NOV48a, CG55828-02 Protein Sequence	SEQ ID NO: 760	457 aa	MW at 51352.3kD

MFGKKKKKIEISGPSNFEHRVHTGFDPQEQKFTGLPQQWHSLLADTANRPKPMVDPSCITPIQLAPMK
VDYDRAQMVLSPPLSGSDTYPRGPAKLPQSQSKSGYSSSSHQYPSGYHKATLYHHPSLQSSSQYISTA
SYLSSLSLSSSTYPPPSWGSSSDQQPSRVSHEQFRAALQLVVSPGDPREYLANFIKIGEGSTGIVCIA
TEKHTGKQVAVKKMDLRKQQRRELLFNEVVIMRDYHHDNVVDMYSSYLVGDELWVVMEFLEGGALTDI
VTHTRMNEEQIATVCLSVLRALSYLHNQGVIHRDIKSDSILLTSDGRIKLSDFGFCAQVSKEVPKRKS
LVGTPYWMAPEVISRLPYGTEVDIWSLGIMVIEMIDGEPPYFNEPPLQAMRRIRDSLPPRVKDLHKVS
SVLRGFLDLMLVREPSQRATAQELLGHPFLKLAGPPSCIVPLMRQYRHH

NOV48b, CG55828-01 DNA Sequence	SEQ ID NO: 761	4796 bp
	ORF Start: ATG at 549	ORF Stop: TGA at 1833

NCGAAAACCCGGAGCAGCTGCGTACGCTCATGGACAGTCCTCCGAGGGGCGAAGCCGGGCAGCTGGGC
ATGCTCAGTAGCTGGGGGAGGTTTGGGTGGAGAGTAGAAAGCTGTGGCTCTGCCTCTCATCCCCTCCC
GCTGGCCCCCGCCCCCCTTGCCCCTACCCAGCCAGTAGTAGTTCCCCAGCGTGCGCCCGGGGAGACCG
GGAACATGGCGCTGGGAGCGCTGTAGCAGCTGAGAAGGGGCTGAGGCACCGCCGCTTCGCTGACAGCC
GGCCACCAGATGTTCATGCATTCTAGAGAAAGTGGAAAACTTAGAAGCCTAATTAATGACTGTCTTCT
GGACCTCTGAGACCATGTTTCTAGTGTTTTCCGTGGAATATTATCAGAAATACACTGTGGTGAAATGC
TTCCACCTCTTGCTAAAATGAACACTGAGGAAAAATGAAGAAGACTGACAAGCACCAGCGAAAAGTTG
CAGAATAGAAATAGCCACACTCCTCTGGAGTCTTTAATTCATCCACAGCCATCATATAAAGGTTTTGG
CATCATGTTTGGGAAGAAAAAGAAAAAGATTGAAATATCTGGCCCGTCCAACTTTGAACACAGGGTTC
ATACTGGGTTTGATCCACAAGAGCAGAAGTTTACCGGCCTTCCCCAGCAGTGGCACAGCCTGTTAGCA
GATACGGCCAACAGGCCAAAGCCTATGGTGGACCCTTCATGCATCACACCCATCCAGCTGGCTCCTAT
GAAGACAATCGTTAGAGGAAACAAACCCTGCAAGGAAACCTCCATCAACGGCCTGCTAGAGGATTTTG
ACAACATCTCGGTGACTCGCTCCAACTCCCTAAGGAAAGAAAGCCCACCCACCCCAGATCAGGGAGCC
TCCAGCCACGGTCCAGGCCACGCGGAAGAAAATGGCTTCATCACCTTCTCCCAGTATTCCAGCGAATC
CGATACTACTGCTGACTACACGACCGAAAAGTACAGGGAGAAGAGTCTCTATGGAGATGATCTGGATC
CGTATTATAGAGGCAGCCACGCAGCCAAGCAAAATGGGCACGTAATGAAAATGAAGCACGGGGAGGCC
TACTATTCTGAGGTGAAGCCTTTGAAATCCGATTTTGCCAGATTTTCTGCCGATTATCACTCACATTT
GGACTCACTGAGCAAACCAAGTGAATACAGTGACCTCAAGTGGGAGTATCAGAGAGCCTCGAGTAGCT

CCCCTCTGGATTATTCATTCCAATTCACACCTTCTAGAACTGCAGGGACCAGCGGGTGCTCCAAGGAG
AGCCTGGCGTACAGTGAAAGTGAATGGGGACCCAGCCTGGATGACTATGACAGGAGGCCAAAGTCTTC
GTACCTGAATCAGACAAGCCCTCAGCCCACCATGCGGCAGAGGTCCAGGTCAGGCTCGGGACTCCAGG
AACCGATGATGCCATTTGGAGCAAGTGCATTTAAAACCCATCCCCAAGGACACTCCTACAACTCCTAC
ACCTACCCTCGCTTGTCCGAGCCCACAATGTGCATTCCAAAGGTGGATTACGATCCAGCACAGATGGT
CCTCAGCCCTCCACTGTCAGGGTCTGACACCTACCCCAGGGGCCCTGCCAAACTACCTCAAAGTCAAA
GCAAATCGGGCTATTCCTCAAGCAGTCACCAGTACCCGTCTGGGTACCACAAAGCCACCTTGTACCAT
CACCCCTCCCTGCAGAGCAGTTCGCAGTACATCTCCACGGCTTCCTACCTGAGCTCCCTCAGCCTCTC
ATCCAGCATACCCGCCGCCCAGCTGGGGCTCCTCCTCCGACCAGCAGCCCTCCAGGGTGTCCCA**TGA**A
CAGTTTCGGGCGGCCCTGCAGCTGGTGGTCAGCCCAGGAGACCCCAGGGAATACTTGGCCAACTTTAT
CAAAATCGGGGAAGGCTCAACCGGCATCGTATGCATCGCCACCGAGAAACACACAGGGAAACAAGTTG
CAGTGAAGAAAATGGACCTCCGGAAGCAACAGAGACGAGAACTGCTTTTCAATGAGGTCGTGATCATG
CGGGATTACCACCATGACAATGTGGTTGACATGTACAGCAGCTACCTTGTCGGCGATGAGCTCTGGGT
GGTCATGGAGTTTCTAGAAGGTGGTGCCTTGACAGACATTGTGACTCACACCAGAATGAATGAAGAAC
AGATAGCTACTGTCTGCCTGTCAGTTCTGAGAGCTCTCTCCTACCTTCATAACCAAGGAGTGATTCAC
AGGGACATAAAAAGTGACTCCATCCTCCTGACAAGCGATGGCCGGATAAAGTTGTCTGATTTTGGTTT
CTGTGCTCAAGTTTCCAAAGAGGTGCCGAAGAGGAAATCATTGGTTGGCACTCCCTACTGGATGGCCC
CTGAGGTGATTTCTAGGCTACCTTATGGGACAGAGGTGGACATCTGGTCCCTCGGGATCATGGTGATA
GAAATGATTGATGGCGAGCCCCCCTACTTCAATGAGCCTCCCCTCCAGGCGATGCGGAGGATCCGGGA
CAGTTTACCTCCAAGAGTGAAGGACCTACACAAGGTTTCTTCAGTGCTCCGGGGATTCCTAGACTTGA
TGTTGGTGAGGGAGCCCTCTCAGAGAGCAACAGCCCAGGAACTCCTCGGACATCCATTCTTAAAACTA
GCAGGTCCACCGTCTTGCATCGTCCCCCTCATGAGACAATACAGGCATCACTGAGCAGAGGATTCGTG
TAGGTGGCAAAGCTAGATGAGGACATGAGAATAATTCAGGAGAACAAAAGGAAACACAGAACATGCAA
AAGGCCTGTGCATTCTAGACCAGCCAATTGGTGGGACAGCGTGATGACCGGCAGGGTTCAACAGACCA
GGGCATCTTCTTGTGTCTTAAACAGGCATCTCTCCACTGACAGCCGGTGTGGTCACTTGGAGCACGGC
TTTAATAAGTCATTATTATATTTTTCAGCCCTTCATCCAGCAAATCAGAAGGACTCAGTACAAACTCC
GTTATGATATATCCTAGCCACATGCAGGGTAACATGTAGGATTTTCTATATTGAAAGAATACTTTTCT
GGCAAAAAAAAAAAAAAAAAGAAAGAAAGGAAAACAAAAAGCACTTTTTTCTTAATGGTAGCAGTATA
ATGTATTTTGCAACGAATTTGTAATTTTTCTGTACGATAGTTTTGATAATTTATAGTACTTTGATGTC
ATGTAGCCATTGTATCAGTTGAAGTAATACTTGTTTACTAGAGGAGTTTGAACAAAGCCTTTCCTACT
TTTTTATCCCTTTAAGAGAACCAATGATTCTTTAGGAACTTTGAATACTGAATGACTCTCAATCACCG

TCAGCTTTAGTAAAATCTCTTTCTTATCCTAACAAGTGTCTTATTTGGTGGAAGAAGAATTAAGAGTG
ATGGTGATGGTGTGCACGTTTCATTAATCCAACCAAAAATAATGAAATAAAATTTGAGCCACAGTATA
CCACTCCTTGGGATAAAGTTAAATATTTTTAAAGATCACATTTTCCATGAACGCCTCTAGTAGCAAAC
CATTCTTTTGCACACCACAATGTTTCCCTCAGTGCCCTTTCTCAAATGGGTACAATGTTCCCTTGTGG
CCAAATTTCCCTCCCAGGGAGCAATTTCAGTGCTAGGATCATTGGATTCAGTTCCCAAAATAGAATGT
TTCAGTGAGACCATGAGAATTCCAGGCTCACAGAGGGAGAGGAGAGAACAGGGCAAGACGTTTGGTTT
CATTTGTCACCATTTTTAAAACTCTGTATGCTAGCACACCAAACTCTTGTCTATATTTACCTTTGTAC
CACAGTATTAATCGCTATTGTTCATGTATCGTGCTGGAAGTCTGAACTGACTCTAGAGGATGAATTAG
CAAGAGGGTATTTTACCAGGTATGATCTGACTTCAGTTGTGCCCATGTTATAATGTGTTTCCGACATA
GGAGAGTCGTGCTGCTGTCTAGATCTTCTTGAATGTTGATAAAAATGAATGACTACTACAATACATTT
TGTGTTGCTTGTTGGATGAATTTGCATGTTAACTGTAGGCCAATATAGATTTGCCTTTAAAACTCTGG
AAGAGCTACATAGTCATCATTAGTTTCTATTAATTATGCATCAGACAAAAGCCATTTGTTACCAAACT
GGGAAAACAGAGGCTTTTCTTAACTATTTCACATACTGTAACAAATATGAATTTAAATTTGTGATAGC
GCTCTGGTTGCTCTAAGCATAATTAAGAATTTTTGTAATTAATAGGTTGCTAATTATTTATCACTGCT
AAAAAGGAAAAAAGGCATAAAATGACCTTCTACTGATTAGATTTTCAGTTTTCTTTCAAACTGGAAAT
GCCTCCATAAATATGATCTATGATTTTGCTTCATAAAACAGCAAATCAATGTTTTATGTAAAATATTA
AAGCATTAATATAAATATGTGAGAATAAAAACAATCTAAATCCAGAAAATGGCAGTCCTAAATGTTCA
TGAGACAGATTGTATTAATTTAACCAGGACTATGTAGAAGTAGAAAGAAAAGAAAAAGAAAATCTTTT
TTAAACCAGAATAAACATTAAAAACTATTGCAGAAAATAGTGGATTTTGGATTCCAAACATTTTCGAC
AGTGTAATGGAAATTTTTCTGTAATTTTCTTACCATCGGGTATTTTTTAAAGTATTCATTGAGTTTAC
CAAAAGTTACTGTAGCTTAAAAGGTTTTGTGAGCACTAACTATTGGCAGAAACTGCATTTGCAAATAA
AAATAAATGTTTGCCTTTTAAAAAAAAAAAAAAAAA

NOV48b, CG55828-01 Protein Sequence	SEQ ID NO: 762	428 aa	MW at 47439.0kD

MFGKKKKKIEISGPSNFEHRVHTGFDPQEQKFTGLPQQWHSLLADTANRPKPMVDPSCITPIQLAPMK
TIVRGNKPCKETSINGLLEDFDNISVTRSNSLRKESPPTPDQGASSHGPGHAEENGFITFSQYSSESD
TTADYTTEKYREKSLYGDDLDPYYRGSHAAKQNGHVMKMKHGEAYYSEVKPLKSDFARFSADYHSHLD

SLSKPSEYSDLKWEYQRASSSSPLDYSFQFTPSRTAGTSGCSKESLAYSESEWGPSLDDYDRRPKSSY
LNQTSPQPTMRQRSRSGSGLQEPMMPFGASAFKTHPQGHSYNSYTYPRLSEPTMCIPKVDYDPAQMVL
SPPLSGSDTYPRGPAKLPQSQSKSGYSSSSHQYPSGYHKATLYHHPSLQSSSQYISTASYLSSLSLSS
SIPAAQLGLLLRPAALQGVP

NOV48c, SNP13379517 of CG55828-02, DNA Sequence	SEQ ID NO: 763	1439 bp
	ORF Start: ATG at 35	ORF Stop: TGA at 1406
	SNP Pos: 149	SNP Change: T to C

TCATCCACAGCCATCATATAAAGGTTTTGGCATCATGTTTGGGAAGAAAAAGAAAAAGATTGAAATAT
CTGGCCCGTCCAACTTTGAACACAGGGTTCATACTGGGTTTGATCCACAAGAGCAGAAGTTTACCGGC
CTTCCCCAGCAGCGGCACAGCCTGTTAGCAGATACGGCCAACAGGCCAAAGCCTATGGTGGACCCTTC
ATGCATCACACCCATCCAGCTGGCTCCTATGAAGGTGGATTACGATCGAGCACAGATGGTCCTCAGCC
CTCCACTGTCAGGGTCTGACACCTACCCCAGGGGCCCTGCCAAACTACCTCAAAGTCAAAGCAAATCG
GGCTATTCCTCAAGCAGTCACCAGTACCCGTCTGGGTACCACAAAGCCACCTTGTACCATCACCCCTC
CCTGCAGAGCAGTTCGCAGTACATCTCCACGGCTTCCTACCTGAGCTCCCTCAGCCTCTCATCCAGCA
CCTACCCGCCGCCCAGCTGGGGCTCCTCCTCCGACCAGCAGCCCTCCAGGGTGTCCCATGAACAGTTT
CGGGCGGCCCTGCAGCTGGTGGTCAGCCCAGGAGACCCCAGGGAATACTTGGCCAACTTTATCAAAAT
CGGGGAAGGCTCAACCGGCATCGTATGCATCGCCACCGAGAAACACACAGGGAAACAAGTTGCAGTGA
AGAAAATGGACCTCCGGAAGCAACAGAGACGAGAACTGCTTTTCAATGAGGTCGTGATCATGCGGGAT
TACCACCATGACAATGTGGTTGACATGTACAGCAGCTACCTTGTCGGCGATGAGCTCTGGGTGGTCAT
GGAGTTTCTAGAAGGTGGTGCCTTGACAGACATTGTGACTCACACCAGAATGAATGAAGAACAGATAG
CTACTGTCTGCCTGTCAGTTCTGAGAGCTCTCTCCTACCTTCATAACCAAGGAGTGATTCACAGGGAC
ATAAAAAGTGACTCCATTCTCCTGACAAGCGATGGCCGGATAAAGTTGTCTGATTTTGGTTTCTGTGC
TCAAGTTTCCAAAGAGGTGCCGAAGAGGAAATCATTGGTTGGCACTCCCTACTGGATGGCCCCTGAGG
TGATTTCTAGGCTACCTTATGGGACAGAGGTGGACATCTGGTCCCTCGGGATCATGGTGATAGAAATG
ATTGATGGCGAGCCCCCCTACTTCAATGAGCCTCCCCTCCAGGCGATGCGGAGGATCCGGGACAGTTT
ACCTCCAAGAGTGAAGGACCTACACAAGGTTTCTTCAGTGCTCCGGGGATTCCTAGACTTGATGTTGG
TGAGGGAGCCCTCTCAGAGAGCAACAGCCCAGGAACTCCTCGGACATCCATTCTTAAAACTAGCAGGT
CCACCGTCTTGCATCGTCCCCCTCATGAGACAATACAGGCATCACTGAGCAGAGGATTCGTGTAGGTG
GCAAAGCTAGA

NOV48c, SNP13379517 of CG55828-02, Protein Sequence	SEQ ID NO: 764	457 aa	MW at 51322.3kD
	SNP Pos: 39		SNP Change: Trp to Arg

MFGKKKKKIEISGPSNFEHRVHTGFDPQEQKFTGLPQQRHSLLADTANRPKPMVDPSCITPIQLAPMK
VDYDRAQMVLSPPLSGSDTYPRGPAKLPQSQSKSGYSSSSHQYPSGYHKATLYHHPSLQSSSQYISTA
SYLSSLSLSSSTYPPPSWGSSSDQQPSRVSHEQFRAALQLVVSPGDPREYLANFIKIGEGSTGIVCIA
TEKHTGKQVAVKKMDLRKQQRRELLFNEVVIMRDYHHDNVVDMYSSYLVGDELWVVMEFLEGGALTDI
VTHTRMNEEQIATVCLSVLRALSYLHNQGVIHRDIKSDSILLTSDGRIKLSDFGFCAQVSKEVPKRKS
LVGTPYWMAPEVISRLPYGTEVDIWSLGIMVIEMIDGEPPYFNEPPLQAMRRIRDSLPPRVKDLHKVS
SVLRGFLDLMLVREPSQRATAQELLGHPFLKLAGPPSCIVPLMRQYRHH

NOV48d, SNP13376535 of CG55828-02, DNA Sequence	SEQ ID NO: 765	1439 bp
	ORF Start: ATG at 35	ORF Stop: TGA at 1406
	SNP Pos: 222	SNP Change: A to G

TCATCCACAGCCATCATATAAAGGTTTTGGCATCATGTTTGGGAAGAAAAAGAAAAAGATTGAAATAT
CTGGCCCGTCCAACTTTGAACACAGGGTTCATACTGGGTTTGATCCACAAGAGCAGAAGTTTACCGGC
CTTCCCCAGCAGTGGCACAGCCTGTTAGCAGATACGGCCAACAGGCCAAAGCCTATGGTGGACCCTTC
ATGCATCACACCCATCCGGCTGGCTCCTATGAAGGTGGATTACGATCGAGCACAGATGGTCCTCAGCC
CTCCACTGTCAGGGTCTGACACCTACCCCAGGGGCCCTGCCAAACTACCTCAAAGTCAAAGCAAATCG
GGCTATTCCTCAAGCAGTCACCAGTACCCGTCTGGGTACCACAAAGCCACCTTGTACCATCACCCCTC
CCTGCAGAGCAGTTCGCAGTACATCTCCACGGCTTCCTACCTGAGCTCCCTCAGCCTCTCATCCAGCA
CCTACCCGCCGCCCAGCTGGGGCTCCTCCTCCGACCAGCAGCCCTCCAGGGTGTCCCATGAACAGTTT
CGGGCGGCCCTGCAGCTGGTGGTCAGCCCAGGAGACCCCAGGGAATACTTGGCCAACTTTATCAAAAT
CGGGGAAGGCTCAACCGGCATCGTATGCATCGCCACCGAGAAACACACAGGGAAACAAGTTGCAGTGA
AGAAAATGGACCTCCGGAAGCAACAGAGACGAGAACTGCTTTTCAATGAGGTCGTGATCATGCGGGAT
TACCACCATGACAATGTGGTTGACATGTACAGCAGCTACCTTGTCGGCGATGAGCTCTGGGTGGTCAT
GGAGTTTCTAGAAGGTGGTGCCTTGACAGACATTGTGACTCACACCAGAATGAATGAAGAACAGATAG

CTACTGTCTGCCTGTCAGTTCTGAGAGCTCTCTCCTACCTTCATAACCAAGGAGTGATTCACAGGGAC
ATAAAAAGTGACTCCATTCTCCTGACAAGCGATGGCCGGATAAAGTTGTCTGATTTTGGTTTCTGTGC
TCAAGTTTCCAAAGAGGTGCCGAAGAGGAAATCATTGGTTGGCACTCCCTACTGGATGGCCCCTGAGG
TGATTTCTAGGCTACCTTATGGGACAGAGGTGGACATCTGGTCCCTCGGGATCATGGTGATAGAAATG
ATTGATGGCGAGCCCCCCTACTTCAATGAGCCTCCCCTCCAGGCGATGCGGAGGATCCGGGACAGTTT
ACCTCCAAGAGTGAAGGACCTACACAAGGTTTCTTCAGTGCTCCGGGGATTCCTAGACTTGATGTTGG
TGAGGGAGCCCTCTCAGAGAGCAACAGCCCAGGAACTCCTCGGACATCCATTCTTAAAACTAGCAGGT
CCACCGTCTTGCATCGTCCCCCTCATGAGACAATACAGGCATCACTGAGCAGAGGATTCGTGTAGGTG
GCAAAGCTAGA

NOV48d, SNP13376535 of CG55828-02, Protein Sequence	SEQ ID NO: 766	457 aa	MW at 51380.3kD
	SNP Pos: 63		SNP Change: Gln to Arg

MFGKKKKKIEISGPSNFEHRVHTGFDPQEQKFTGLPQQWHSLLADTANRPKPMVDPSCITPIRLAPMK
VDYDRAQMVLSPPLSGSDTYPRGPAKLPQSQSKSGYSSSSHQYPSGYHKATLYHHPSLQSSSQYISTA
SYLSSLSLSSSTYPPPSWGSSSDQQPSRVSHEQFRAALQLVVSPGDPREYLANFIKIGEGSTGIVCIA
TEKHTGKQVAVKKMDLRKQQRRELLFNEVVIMRDYHHDNVVDMYSSYLVGDELWVVMEFLEGGALTDI
VTHTRMNEEQIATVCLSVLRALSYLHNQGVIHRDIKSDSILLTSDGRIKLSDFGFCAQVSKEVPKRKS
LVGTPYWMAPEVISRLPYGTEVDIWSLGIMVIEMIDGEPPYFNEPPLQAMRRIRDSLPPRVKDLHKVS
SVLRGFLDLMLVREPSQRATAQELLGHPFLKLAGPPSCIVPLMRQYRHH

NOV48e, SNP13382500 of CG55828-02, DNA Sequence	SEQ ID NO: 767	1439 bp
	ORF Start: ATG at 35	ORF Stop: TGA at 1406
	SNP Pos: 970	SNP Change: T to C

TCATCCACAGCCATCATATAAAGGTTTTGGCATCATGTTTGGGAAGAAAAAGAAAAAGATTGAAATAT
CTGGCCCGTCCAACTTTGAACACAGGGTTCATACTGGGTTTGATCCACAAGAGCAGAAGTTTACCGGC
CTTCCCCAGCAGTGGCACAGCCTGTTAGCAGATACGGCCAACAGGCCAAAGCCTATGGTGGACCCTTC
ATGCATCACACCCATCCAGCTGGCTCCTATGAAGGTGGATTACGATCGAGCACAGATGGTCCTCAGCC
CTCCACTGTCAGGGTCTGACACCTACCCCAGGGGCCCTGCCAAACTACCTCAAAGTCAAAGCAAATCG
GGCTATTCCTCAAGCAGTCACCAGTACCCGTCTGGGTACCACAAAGCCACCTTGTACCATCACCCCTC
CCTGCAGAGCAGTTCGCAGTACATCTCCACGGCTTCCTACCTGAGCTCCCTCAGCCTCTCATCCAGCA
CCTACCCGCCGCCCAGCTGGGGCTCCTCCTCCGACCAGCAGCCCTCCAGGGTGTCCCATGAACAGTTT
CGGGCGGCCCTGCAGCTGGTGGTCAGCCCAGGAGACCCCAGGGAATACTTGGCCAACTTTATCAAAAT
CGGGGAAGGCTCAACCGGCATCGTATGCATCGCCACCGAGAAACACACAGGGAAACAAGTTGCAGTGA
AGAAAATGGACCTCCGGAAGCAACAGAGACGAGAACTGCTTTTCAATGAGGTCGTGATCATGCGGGAT
TACCACCATGACAATGTGGTTGACATGTACAGCAGCTACCTTGTCGGCGATGAGCTCTGGGTGGTCAT
GGAGTTTCTAGAAGGTGGTGCCTTGACAGACATTGTGACTCACACCAGAATGAATGAAGAACAGATAG
CTACTGTCTGCCTGTCAGTTCTGAGAGCTCTCTCCTACCTTCATAACCAAGGAGTGATTCACAGGGAC
ATAAAAAGTGACTCCATCCTCCTGACAAGCGATGGCCGGATAAAGTTGTCTGATTTTGGTTTCTGTGC
TCAAGTTTCCAAAGAGGTGCCGAAGAGGAAATCATTGGTTGGCACTCCCTACTGGATGGCCCCTGAGG
TGATTTCTAGGCTACCTTATGGGACAGAGGTGGACATCTGGTCCCTCGGGATCATGGTGATAGAAATG
ATTGATGGCGAGCCCCCCTACTTCAATGAGCCTCCCCTCCAGGCGATGCGGAGGATCCGGGACAGTTT
ACCTCCAAGAGTGAAGGACCTACACAAGGTTTCTTCAGTGCTCCGGGGATTCCTAGACTTGATGTTGG
TGAGGGAGCCCTCTCAGAGAGCAACAGCCCAGGAACTCCTCGGACATCCATTCTTAAAACTAGCAGGT
CCACCGTCTTGCATCGTCCCCCTCATGAGACAATACAGGCATCACTGAGCAGAGGATTCGTGTAGGTG
GCAAAGCTAGA

NOV48e, SNP13382500 of CG55828-02, Protein Sequence	SEQ ID NO: 768	457 aa	MW at 51352.3kD
	SNP Pos: 312		SNP Change: Ile to Ile

MFGKKKKKIEISGPSNFEHRVHTGFDPQEQKFTGLPQQWHSLLADTANRPKPMVDPSCITPIQLAPMK
VDYDRAQMVLSPPLSGSDTYPRGPAKLPQSQSKSGYSSSSHQYPSGYHKATLYHHPSLQSSSQYISTA
SYLSSLSLSSSTYPPPSWGSSSDQQPSRVSHEQFRAALQLVVSPGDPREYLANFIKIGEGSTGIVCIA
TEKHTGKQVAVKKMDLRKQQRRELLFNEVVIMRDYHHDNVVDMYSSYLVGDELWVVMEFLEGGALTDI
VTHTRMNEEQIATVCLSVLRALSYLHNQGVIHRDIKSDSILLTSDGRIKLSDFGFCAQVSKEVPKRKS
LVGTPYWMAPEVISRLPYGTEVDIWSLGIMVIEMIDGEPPYFNEPPLQAMRRIRDSLPPRVKDLHKVS
SVLRGFLDLMLVREPSQRATAQELLGHPFLKLAGPPSCIVPLMRQYRHH

NOV48f, SNP13375705 of CG55828-02, DNA Sequence	SEQ ID NO: 769	1439 bp
	ORF Start: ATG at 35	ORF Stop: TGA at 1406
	SNP Pos: 1375	SNP Change: C to T

TCATCCACAGCCATCATATAAAGGTTTTGGCATCATGTTTGGGAAGAAAAAGAAAAAGATTGAAATAT
CTGGCCCGTCCAACTTTGAACACAGGGTTCATACTGGGTTTGATCCACAAGAGCAGAAGTTTACCGGC
CTTCCCCAGCAGTGGCACAGCCTGTTAGCAGATACGGCCAACAGGCCAAAGCCTATGGTGGACCCTTC
ATGCATCACACCCATCCAGCTGGCTCCTATGAAGGTGGATTACGATCGAGCACAGATGGTCCTCAGCC
CTCCACTGTCAGGGTCTGACACCTACCCCAGGGGCCCTGCCAAACTACCTCAAAGTCAAAGCAAATCG
GGCTATTCCTCAAGCAGTCACCAGTACCCGTCTGGGTACCACAAAGCCACCTTGTACCATCACCCCTC
CCTGCAGAGCAGTTCGCAGTACATCTCCACGGCTTCCTACCTGAGCTCCCTCAGCCTCTCATCCAGCA
CCTACCCGCCGCCCAGCTGGGGCTCCTCCTCCGACCAGCAGCCCTCCAGGGTGTCCCATGAACAGTTT
CGGGCGGCCCTGCAGCTGGTGGTCAGCCCAGGAGACCCCAGGGAATACTTGGCCAACTTTATCAAAAT
CGGGGAAGGCTCAACCGGCATCGTATGCATCGCCACCGAGAAACACACAGGGAAACAAGTTGCAGTGA
AGAAAATGGACCTCCGGAAGCAACAGAGACGAGAACTGCTTTTCAATGAGGTCGTGATCATGCGGGAT
TACCACCATGACAATGTGGTTGACATGTACAGCAGCTACCTTGTCGGCGATGAGCTCTGGGTGGTCAT
GGAGTTTCTAGAAGGTGGTGCCTTGACAGACATTGTGACTCACACCAGAATGAATGAAGAACAGATAG
CTACTGTCTGCCTGTCAGTTCTGAGAGCTCTCTCCTACCTTCATAACCAAGGAGTGATTCACAGGGAC
ATAAAAAGTGACTCCATTCTCCTGACAAGCGATGGCCGGATAAAGTTGTCTGATTTTGGTTTCTGTGC
TCAAGTTTCCAAAGAGGTGCCGAAGAGGAAATCATTGGTTGGCACTCCCTACTGGATGGCCCCTGAGG
TGATTTCTAGGCTACCTTATGGGACAGAGGTGGACATCTGGTCCCTCGGGATCATGGTGATAGAAATG
ATTGATGGCGAGCCCCCCTACTTCAATGAGCCTCCCCTCCAGGCGATGCGGAGGATCCGGGACAGTTT
ACCTCCAAGAGTGAAGGACCTACACAAGGTTTCTTCAGTGCTCCGGGGATTCCTAGACTTGATGTTGG
TGAGGGAGCCCTCTCAGAGAGCAACAGCCCAGGAACTCCTCGGACATCCATTCTTAAAACTAGCAGGT
CCACCGTCTTGCATTGTCCCCCTCATGAGACAATACAGGCATCACTGAGCAGAGGATTCGTGTAGGTG
GCAAAGCTAGA

NOV48f, SNP13375705 of CG55828-02, Protein Sequence	SEQ ID NO: 770	457 aa	MW at 51352.3kD
	SNP Pos: 447		SNP Change: Ile to Ile

MFGKKKKKIEISGPSNFEHRVHTGFDPQEQKFTGLPQQWHSLLADTANRPKPMVDPSCITPIQLAPMK
VDYDRAQMVLSPPLSGSDTYPRGPAKLPQSQSKSGYSSSSHQYPSGYHKATLYHHPSLQSSSQYISTA
SYLSSLSLSSSTYPPPSWGSSSDQQPSRVSHEQFRAALQLVVSPGDPREYLANFIKIGEGSTGIVCIA
TEKHTGKQVAVKKMDLRKQQRRELLFNEVVIMRDYHHDNVVDMYSSYLVGDELWVVMEFLEGGALTDI
VTHTRMNEEQIATVCLSVLRALSYLHNQGVIHRDIKSDSILLTSDGRIKLSDFGFCAQVSKEVPKRKS
LVGTPYWMAPEVISRLPYGTEVDIWSLGIMVIEMIDGEPPYFNEPPLQAMRRIRDSLPPRVKDLHKVS
SVLRGFLDLMLVREPSQRATAQELLGHPFLKLAGPPSCIVPLMRQYRHH

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 48B.

Table 48B. Comparison of the NOV48 protein sequences.

```
NOV48a  MFGKKKKKIEISGPSNFEHRVHTGFDPQEQKFTGLPQQWHSLLADTANRPKPMVDPSCIT
NOV48b  MFGKKKKKIEISGPSNFEHRVHTGFDPQEQKFTGLPQQWHSLLADTANRPKPMVDPSCIT

NOV48a  PIQLAPMKVDYDRAQMVLSPPLSGSDTYPRGPAKLPQSQSKSGYSSSSHQYPSGYHKATL
NOV48b  PIQLAPMKTIVRGNKPCKETSING-LLEDFDNISVTRSNSLRKESPPTPDQGASSHGPGH

NOV48a  YHHPSLQSSSQYISTASYLSSLSLSSSTYPPPSWGSSSDQQPSRVSHEQFRAALQLVVSP
NOV48b  AEENGFITFSQYSSESD--TTADYTTEKYREKSLYGDDLDPYYRGSHAAKQNGHVMKMKH

NOV48a  GDPREYLANFIKIGEGSTGIVCIATEKHTGKQVAVKKMDLRKQQRRELLFNEVVIMRDYH
NOV48b  G-----EAYYSEVKPLKSDFARFSADYHSHLDSLSKPSEYS-DLKWEYQRASSSSPLDYS
```

```
NOV48a  HDNVVDMYS-SYLVGDELWVVMEFLEGGALTDIVTHTRMNEEQIATVCLSVLRALSYLHN
NOV48b  FQFTPSRTAGTSGCSKESLAYSESEWGPSLDDYDRRPKSS------------YLNQTSP

NOV48a  QGVIHRDIKSDSILLTSDGRIKLSDFGFCAQVSKEVPKRKSLVGTPYWMAPEVISRLPYG
NOV48b  QPTMRQRSRS------GSG-LQEPMMPFGASAFKTHPQGHSYNSYTYPRLSEPTMCIP--

NOV48a  TEVDIWSLGIMVIEMIDGEPPYFNEPPLQAMRRIRDSLPPRVKDLHKVSSVLRGFLDLML
NOV48b  -KVDYDPAQMVLSPPLSGSDTYPRGP--AKLPQSQSKSGYSSSSHQYPSGYHKATLYHHP

NOV48a  VREPS-QRATAQELLGHPFLKLAGPPSCIVPLMRQYRHH---
NOV48b  SLQSSSQYISTASYLSSLSLSSSIPAAQLGLLLRPAALQGVP

NOV48a  (SEQ ID NO:  760)
NOV48b  (SEQ ID NO:  762)
```

Further analysis of the NOV48a protein yielded the following properties shown in Table 48C.

Table 48C. Protein Sequence Properties NOV48a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 10;  pos.chg 5;  neg.chg 1
      H-region:  length 7;  peak value  -2.19
      PSG score:  -6.59

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -7.22
      possible cleavage site: between 44 and 45

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  2.54 (at 257)
      ALOM score:   2.54  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  3.66
      Hyd Moment(95):  6.70    G content:       2
      D/E content:     2       S/T content:     2
      Score: -7.66

Gavel: prediction of cleavage sites for mitochondrial preseq
```

```
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: KKKK (5) at    4
      pat4: KKKK (5) at    5
      pat4: PKRK (4) at  336
      pat7: PKRKSLV (5) at  336
      bipartite: none
      content of basic residues:  11.2%
      NLS Score:  0.84

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 76.7

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

-------------------------

Final Results (k = 9/23):
```

```
       78.3 %: nuclear
       13.0 %: cytoplasmic
        4.3 %: mitochondrial
        4.3 %: peroxisomal

>> prediction for CG55828-02 is nuc (k=23)
```

A search of the NOV48a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 48D.

Table 48D. Geneseq Results for NOV48a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV48a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB65705	Novel protein kinase, SEQ ID NO: 234 - Homo sapiens, 719 aa. [WO200073469-A2, 07-DEC-2000]	56..457 319..719	393/402 (97%) 393/402 (97%)	0.0
AAM38963	Human polypeptide SEQ ID NO 2108 - Homo sapiens, 719 aa. [WO200153312-A1, 26-JUL-2001]	56..457 319..719	392/402 (97%) 393/402 (97%)	0.0
AAE02187	Human p21-activated kinase 5 (PAK5) protein - Homo sapiens, 719 aa. [WO200136602-A2, 25-MAY-2001]	56..457 319..719	391/402 (97%) 392/402 (97%)	0.0
AAG67825	Human P21-active kinase 60 protein - Homo sapiens, 547 aa. [CN1298009-A, 06-JUN-2001]	56..457 147..547	355/402 (88%) 358/402 (88%)	0.0
ABG19308	Novel human diagnostic protein #19299 - Homo sapiens, 620 aa. [WO200175067-A2, 11-OCT-2001]	150..455 313..618	253/306 (82%) 276/306 (89%)	e-146

In a BLAST search of public sequence databases, the NOV48a protein was found to have homology to the proteins shown in the BLASTP data in Table 48E.

Table 48E. Public BLASTP Results for NOV48a				
Protein Accession Number	**Protein/Organism/Length**	**NOV48a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9P286	Serine/threonine-protein kinase PAK 7 (EC 2.7.1.-) (p21-activated kinase 7) (PAK-7) (PAK-5) - Homo sapiens (Human), 719 aa.	56..457 319..719	393/402 (97%) 393/402 (97%)	0.0
Q8TB93	P21(CDKN1A)-activated kinase 7 - Homo sapiens (Human), 719 aa.	56..457 319..719	392/402 (97%) 392/402 (97%)	0.0
Q8C015	Serine/threonine-protein kinase PAK 5 - Mus musculus (Mouse), 719 aa.	56..457 319..719	382/402 (95%) 389/402 (96%)	0.0
Q8BVB0	Serine/threonine-protein kinase PAK 5 - Mus musculus (Mouse), 719 aa.	56..457 319..719	381/402 (94%) 388/402 (95%)	0.0
Q9ULS8	Hypothetical protein KIAA1142 - Homo sapiens (Human), 467 aa (fragment).	1..455 30..465	317/460 (68%) 359/460 (77%)	e-174

PFam analysis indicates that the NOV48a protein contains the domains shown in the Table 48F.

Table 48F. Domain Analysis of NOV48a			
Pfam Domain	**NOV48a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
PBD	11..69	22/64 (34%) 44/64 (69%)	1.3e-10
pkinase	187..438	91/297 (31%) 200/297 (67%)	7.8e-83

Example 49.

The NOV49 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 49A.

Table 49A. NOV49 Sequence Analysis

NOV49a, CG55980-02 DNA Sequence	SEQ ID NO: 771	972 bp
	ORF Start: ATG at 18	ORF Stop: TAA at 954

CTCATATCTCCCTCATTATGTCTGTTCTCAATAACTCCGAAGTCAAGCTTTTCCTTCTGATTGGGATC
CCAGGACTGGAACATGCCCACATTTGGTTCTCCATCCCCATTTGCCTCATGTACCTGCTTGCCATCAT
GGGCAACTGCACCATTCTCTTTATTATAAAGACAGAGCCCTCGCTTCATGAGCCCATGTATTATTTCC
TTGCCATGTTGGCTGTCTCTGACATGGGCCTGTCCCTCTCCTCCCTTCCTACCATGTTGAGGGTCTTC
TTGTTCAATGCCATGGGAATTTCACCTAATGCCTGCTTTGCTCAAGAATTCTTCATTCATGGATTCAC
TGTCATGGAATCCTCAGTACTTCTAATTATGTCTTTGGACCGCTTTCTTGCCATTCACAATCCCTTAA
GATACAGTTCTATCCTCACTAGCAACAGGGTTGCTAAAATGGGACTTATTTTAGCCATTAGGAGCATT
CTCTTAGTGATTCCATTTCCCTTCACCTTAAGGAGATTAAAATATTGTCAAAAGAATCTTCTTTCTCA
CTCATACTGTCTTCATCAGGATACCATGAAGCTGGCCTGCTCTGACAACAAGACCAATGTCATCTATG
GCTTCTTCATTGCTCTCTGTACTATGCTGGACTTGGCACTGATTGTTTTGTCTTATGTGCTGATCTTG
AAGACTATACTCAGCATTGCATCTTTGGCAGAGAGGCTTAAGGCCCTAAATACCTGTGTCTCCCACAT
CTGTGCTGTGCTCACCTTCTATGTGCCCATCATCACCCTGGCTGCCATGCATCACTTTGCCAAGCACA
AAAGCCCTCTTGTTGTGATCCTTATTGCAGATATGTTCTTGTTGGTGCCGCCCCTTATGAACCCCATT
GTGTACTGTGTAAAGACTCGACAAATCTGGGAGAAGATCTTGGGGAAGTTGCTTAATGTATGTGGGAG
ATAAGAACTTGAACAATTAG

NOV49a, CG55980-02 Protein Sequence	SEQ ID NO: 772	312 aa	MW at 35131.1kD

MSVLNNSEVKLFLLIGIPGLEHAHIWFSIPICLMYLLAIMGNCTILFIIKTEPSLHEPMYYFLAMLAV
SDMGLSLSSLPTMLRVFLFNAMGISPNACFAQEFFIHGFTVMESSVLLIMSLDRFLAIHNPLRYSSIL
TSNRVAKMGLILAIRSILLVIPFPFTLRRLKYCQKNLLSHSYCLHQDTMKLACSDNKTNVIYGFFIAL
CTMLDLALIVLSYVLILKTILSIASLAERLKALNTCVSHICAVLTFYVPIITLAAMHHFAKHKSPLVV
ILIADMFLLVPPLMNPIVYCVKTRQIWEKILGKLLNVCGR

NOV49b, CG55980-01 DNA Sequence	SEQ ID NO: 773		972 bp
	ORF Start: ATG at 18		ORF Stop: TAA at 954

CTCATATCTCCCTCATTATGTCTGTTCTCAATAACTCCGAAGTCAAGCTTTTCCTTCTGATTGGGATC
CCAGGACTGGAACATGCCCACATTTGGTTCTCCATCCCCATTTGCCTCATGTACCTGCTTGCCATCAT
GGGCAACTGCACCATTCTCTTTATTATAAAGACAGAGCCCTCGCTTCATGAGCCCATGTATTATTTCC
TTGCCATGTTGGCTGTCTCTGACATGGGCCTGTCCCTCTCCTCCCTTCCTACCATGTTGAGGGTCTTC
TTGTTCAATGCCATGGGAATTTCACCTAATGCCTGCTTTGCTCAAGAATTCTTCATTCATGGATTCAC
TGTCATGGAATCCTCAGTACTTCTAATTATGTCTTTGGACCGCTTTCTTGCCATTCACAATCCCTTAA
GATACAGTTCTATCCTCACTAGCAACAGGGTTGCTAAAATGGGACTTATTTTAGCCATTAGGAGCATT
CTCTTAGTGATTCCATTTCCCTTCACCTTAAGGAGATTAAAATATTGTCAAAAGAATCTTCTTTCTCA
CTCATACTGTCTTCATCAGGATACCATGAAGCTGGCCTGCTCTGACAACAAGACCAATGTCATCTATG
GCTTCTTCATTGCTCTCTGTACTATGCTGGACTTGGCACTGATTGTTTTGTCTTATGTGCTGATCTTG
AAGACTATACTCAGCATTGCATCTTTGGCAGAGAGGCTTAAGGCCCTAAATACCTGTGTCTCCCACAT
CTGTGCTGTGCTCACCTTCTATGTGCCCATCATCACCCTGGCTGCCATGCATCACTTTGCCAAGCACA
AAAGCCCTCTTGTTGTGATCCTTATTGCAGATATGTTCTTGTTGGTGCCGCCCCTTATGAACCCCATT
GTGTACTGTGTAAAGACTCGACAAATCTGGGAGAAGATCTTGGGGAAGTTGCTTAATGTATGTGGGAG
ATAAGAACTTGAACAATTAG

NOV49b, CG55980-01 Protein Sequence	SEQ ID NO: 774	312 aa	MW at 35131.1kD

MSVLNNSEVKLFLLIGIPGLEHAHIWFSIPICLMYLLAIMGNCTILFIIKTEPSLHEPMYYFLAMLAV
SDMGLSLSSLPTMLRVFLFNAMGISPNACFAQEFFIHGFTVMESSVLLIMSLDRFLAIHNPLRYSSIL
TSNRVAKMGLILAIRSILLVIPFPFTLRRLKYCQKNLLSHSYCLHQDTMKLACSDNKTNVIYGFFIAL
CTMLDLALIVLSYVLILKTILSIASLAERLKALNTCVSHICAVLTFYVPIITLAAMHHFAKHKSPLVV
ILIADMFLLVPPLMNPIVYCVKTRQIWEKILGKLLNVCGR

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 49B.

Table 49B. Comparison of the NOV49 protein sequences.

```
NOV49a  MSVLNNSEVKLFLLIGIPGLEHAHIWFSIPICLMYLLAIMGNCTILFIIKTEPSLHEPMY
NOV49b  MSVLNNSEVKLFLLIGIPGLEHAHIWFSIPICLMYLLAIMGNCTILFIIKTEPSLHEPMY

NOV49a  YFLAMLAVSDMGLSLSSLPTMLRVFLFNAMGISPNACFAQEFFIHGFTVMESSVLLIMSL
NOV49b  YFLAMLAVSDMGLSLSSLPTMLRVFLFNAMGISPNACFAQEFFIHGFTVMESSVLLIMSL

NOV49a  DRFLAIHNPLRYSSILTSNRVAKMGLILAIRSILLVIPFPFTLRRLKYCQKNLLSHSYCL
NOV49b  DRFLAIHNPLRYSSILTSNRVAKMGLILAIRSILLVIPFPFTLRRLKYCQKNLLSHSYCL

NOV49a  HQDTMKLACSDNKTNVIYGFFIALCTMLDLALIVLSYVLILKTILSIASLAERLKALNTC
NOV49b  HQDTMKLACSDNKTNVIYGFFIALCTMLDLALIVLSYVLILKTILSIASLAERLKALNTC

NOV49a  VSHICAVLTFYVPIITLAAMHHFAKHKSPLVVILIADMFLLVPPLMNPIVYCVKTRQIWE
NOV49b  VSHICAVLTFYVPIITLAAMHHFAKHKSPLVVILIADMFLLVPPLMNPIVYCVKTRQIWE

NOV49a  KILGKLLNVCGR
NOV49b  KILGKLLNVCGR

NOV49a  (SEQ ID NO:  772)
NOV49b  (SEQ ID NO:  774)
```

Further analysis of the NOV49a protein yielded the following properties shown in Table 49C.

Table 49C. Protein Sequence Properties NOV49a

SignalP analysis:	Cleavage site between residues 42 and 43

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 10;  pos.chg 1;  neg.chg 1
      H-region:  length 10;  peak value  10.00
      PSG score:   5.60

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.28
      possible cleavage site: between 23 and 24

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   8
      INTEGRAL    Likelihood = -6.00   Transmembrane   33 -  49
      INTEGRAL    Likelihood =  0.32   Transmembrane   62 -  78
      INTEGRAL    Likelihood = -0.48   Transmembrane  102 - 118
      INTEGRAL    Likelihood = -5.63   Transmembrane  141 - 157
      INTEGRAL    Likelihood = -7.59   Transmembrane  199 - 215
      INTEGRAL    Likelihood =  0.42   Transmembrane  218 - 234
```

```
      INTEGRAL    Likelihood = -5.79   Transmembrane  244 - 260
      INTEGRAL    Likelihood = -8.17   Transmembrane  270 - 286
      PERIPHERAL  Likelihood =  2.12 (at 11)
      ALOM score:  -8.17  (number of TMSs: 8)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 40
      Charge difference: -1.5   C(-0.5) - N( 1.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  0.68
      Hyd Moment(95):  3.43     G content:       2
      D/E content:      2       S/T content:     2
      Score: -8.46

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.7%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none
```

```
checking 63 PROSITE DNA binding motifs:
      Leucine zipper pattern (PS00029):  *** found ***
         LKYCQKNLLSHSYCLHQDTMKL at 166
 none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       22.2 %: mitochondrial
       11.1 %: vesicles of secretory system

>> prediction for CG55980-02 is end (k=9)
```

A search of the NOV49a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 49D.

Table 49D. Geneseq Results for NOV49a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV49a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU85302	G-coupled olfactory receptor #163 - Homo sapiens, 312 aa. [WO200198526-A2, 27-DEC-2001]	1..312 1..312	312/312 (100%) 312/312 (100%)	e-178
AAU95741	Human olfactory and pheromone G protein-coupled receptor #228 - Homo sapiens, 312 aa. [WO200224726-A2, 28-MAR-2002]	1..312 1..312	312/312 (100%) 312/312 (100%)	e-178
AAB71167	Human GPCRX protein SEQ ID 10 - Homo sapiens, 312 aa. [WO200250275-A2, 27-JUN-2002]	1..312 1..312	312/312 (100%) 312/312 (100%)	e-178
AAB71166	Human GPCRX protein SEQ ID 8 - Homo sapiens, 312 aa. [WO200250275-A2, 27-JUN-2002]	1..312 1..312	312/312 (100%) 312/312 (100%)	e-178
AAB71344	Human GCREC-23 INCYTE ID 7472087CD1 SEQ ID 23 - Homo sapiens, 312 aa. [WO200263004-A2, 15-AUG-2002]	1..312 1..312	312/312 (100%) 312/312 (100%)	e-178

In a BLAST search of public sequence databases, the NOV49a protein was found to have homology to the proteins shown in the BLASTP data in Table 49E.

Table 49E. Public BLASTP Results for NOV49a				
Protein Accession Number	**Protein/Organism/Length**	**NOV49a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8NH64	Seven transmembrane helix receptor - Homo sapiens (Human), 312 aa.	1..312 1..312	312/312 (100%) 312/312 (100%)	e-178
Q8VH16	Olfactory receptor MOR8-1 - Mus musculus (Mouse), 318 aa.	1..306 8..315	242/308 (78%) 272/308 (87%)	e-141
Q8VH12	Olfactory receptor MOR8-3 - Mus musculus (Mouse), 312 aa.	3..306 2..307	233/306 (76%) 262/306 (85%)	e-134
CAD37707	Sequence 449 from Patent WO0224726 - Homo sapiens (Human), 313 aa.	1..306 1..308	222/308 (72%) 268/308 (86%)	e-130
Q8NGJ6	Seven transmembrane helix receptor - Homo sapiens (Human), 313 aa.	1..306 1..308	221/308 (71%) 267/308 (85%)	e-129

PFam analysis indicates that the NOV49a protein contains the domains shown in the Table 49F.

Table 49F. Domain Analysis of NOV49a			
Pfam Domain	**NOV49a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	41..291	57/277 (21%) 169/277 (61%)	7.4e-13

Example 50.

The NOV50 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 50A.

Table 50A. NOV50 Sequence Analysis			
NOV50a, CG55988-03 DNA Sequence	SEQ ID NO: 775		1630 bp
	ORF Start: ATG at 1		ORF Stop: TGA at 1528

ATGGGGTCCCGCCACTTCGAGGGGATTTATGACCACGTGGGGCACTTCGGCAGATTCCAGAGAGTCCT
CTATTTCATATGTGCCTTCCAGAACATCTCTTGTGGTATTCACTACTTGGCTTCTGTGTTCATGGGAG
TCACCCCTCATCATGTCTGCAGGCCCCCAGGCAATGTGAGTCAGGTTGTTTTCCATAATCACTCTAAT
TGGAGTTTGGAGGACACCGGGGCCCTGTTGTCTTCAGGCCAGAAAGATTATGTTACGGTGCAGTTGCA
GAATGGTGAGATCTGGGAGCTCTCAAGGTGTAGCAGGAATAAGAGGGAGAACACATCGAGTTTGGGCT
ATGAATACACTGGCAGTAAGAAAGAGTTTCCTTGTGTGGATGGCTACATATATGACCAGAACACATGG
AAAAGCACTGCGGTGACCCAGTGGAACCTGGTCTGTGACCGAAAATGGCTTGCAATGCTGATCCAGCC
CCTATTTATGTTTGGAGTCCTACTGGGATCGGTGACTTTTGGCTACTTTTCTGACAGGCTAGGACGCC
GGGTGGTCTTGTGGGCCACAAGCAGTAGCATGTTTTTGTTTGGAATAGCAGCGGCGTTTGCAGTTGAT
TATTACACCTTCATGGCTGCTCGCTTTTTTCTTGCCATGGTTGCAAGTGGCTATCTTGTGGTGGGGTT
TGTCTATGTGATGGAATTCATTGGCATGAAGTCTCGGACATGGGCGTCTGTCCATTTGCATTCCTTTT
TTGCAGTTGGAACCCTGCTGGTGGCTTTGACAGGATACTTGGTCAGGACCTGGTGGCTTTACCAGATG
ATCCTCTCCACAGTGACTGTCCCCTTTATCCTGTGCTGTTGGGTGCTCCCAGAGACACCTTTTTGGCT
TCTCTCAGAGGGACGATATGAAGAAGCACAAAAAATAGTTGACATCATGGCCAAGTGGAACAGGGCAA
GCTCCTGTAAACTGTCAGAACTTTTATCACTGGACCTACAAGGTCCTGTTAGTAATAGCCCCACTGAA
GTTCAGAAGCACAACCTATCATATCTGTTTTATAACTGGAGCATTACGAAAAGGACACTTACCGTTTG
GCTAATCTGGTTCACTGGAAGTTTGGGATTCTACTCGTTTTCCTTGAATTCTGTTAACTTAGGAGGCA
ATGAATACTTAAACCTCTTCCTCCTGGGTGTAGTGGAAATTCCCGCCTACACCTTCGTGTGCATCGCC
ACGGACAAGGTCGGGAGGAGAACAGTCCTGGCCTACTCTCTTTTCTGCAGTGCACTGGCCTGTGGTGT
CGTTATGGTGATCCCCCAGATCGCTGGCTGTGGGAAGCGGCAGCATGGTGTGTCGCCTGGCCAGCATC
CTGGCGCCGTTCTCTGTGGACCTCAGCAGCATTTGGATCTTCATACCACAGTTGTTTGTTGGGACTAT
GGCCCTCCTGAGTGGAGTGTTAACACTAAAGCTTCCAGAAACCCTTGGGAAACGGCTAGCAACTACTT
GGGAGGAGGCTGCAAAACTGGAGTCAGAGAA**TGA**AAGCAAGTCAAGCAAATTACTTCTCACAACTAAT
AATAGTGGGCTGGAAAAAACGGAAGCGATTACCCCCAGGGATTCTGGTCTTGGTGAATAAATGTGC

NOV50a, CG55988-03 Protein Sequence	SEQ ID NO: 776	509 aa	MW at 57300.6kD

MGSRHFEGIYDHVGHFGRFQRVLYFICAFQNISCGIHYLASVFMGVTPHHVCRPPGNVSQVVFHNHSN
WSLEDTGALLSSGQKDYVTVQLQNGEIWELSRCSRNKRENTSSLGYEYTGSKKEFPCVDGYIYDQNTW
KSTAVTQWNLVCDRKWLAMLIQPLFMFGVLLGSVTFGYFSDRLGRRVVLWATSSSMFLFGIAAAFAVD
YYTFMAARFFLAMVASGYLVVGFVYVMEFIGMKSRTWASVHLHSFFAVGTLLVALTGYLVRTWWLYQM
ILSTVTVPFILCCWVLPETPFWLLSEGRYEEAQKIVDIMAKWNRASSCKLSELLSLDLQGPVSNSPTE
VQKHNLSYLFYNWSITKRTLTVWLIWFTGSLGFYSFSLNSVNLGGNEYLNLFLLGVVEIPAYTFVCIA
TDKVGRRTVLAYSLFCSALACGVVMVIPQIAGCGKRQHGVSPGQHPGAVLCGPQQHLDLHTTVVCWDY
GPPEWSVNTKASRNPWETASNYLGGGCKTGVRE

NOV50b, CG55988-04 DNA Sequence	SEQ ID NO: 777	1599 bp
	ORF Start: ATG at 1	ORF Stop: TAA at 1546

ATGGGGTCCCGCCACTTCGAGGGGATTTATGACCACGTGGGGCACTTCGGCAGATTCCAGAGAGTCCT
CTATTTCATATGTGCCTTCCAGAACATCTCTTGTGGTATTCACTACTTGGCTTCTGTGTTCATGGGAG
TCACCCCTCATCATGTCTGCAGGCCCCCAGGCAATGTGAGTCAGGTTGTTTTCCATAATCACTCTAAT
TGGAGTTTGGAGGACACCGGGGCCCTGTTGTCTTCAGGCCAGAAAGATTATGTTACGGTGCAGTTGCA
GAATGGTGAGATCTGGGAGCTCTCAAGGTGTAGCAGGAATAAGAGGGAGAACACATCGAGTTTGGGCT
ATGAATACACTGGCAGTAAGAAAGAGTTTCCTTGTGTGGATGGCTACATATATGACCAGAACACATGG
AAAAGCACTGCGGTGACCCAGTGGAACCTGGTCTGTGACCGAAAATGGCTTGCAATGCTGATCCAGCC
CCTATTTATGTTTGGAGTCCTACTGGGATCGGTGACTTTTGGCTACTTTTCTGACAGGCTAGGACGCC
GGGTGGTCTTGTGGGCCACAAGCAGTAGCATGTTTTTGTTTGGAATAGCAGCGGCGTTTGCAGTTGAT
TATTACACCTTCATGGCTGCTCGCTTTTTTCTTGCCATGGTTGCAAGTGGCTATCTTGTGGTGGGGTT
TGTCTATGTGATGGAATTCATTGGCATGAAGTCTCGGACATGGGCGTCTGTCCATTTGCATTCCTTTT
TTGCAGTTGGAACCCTGCTGGTGGCTTTGACAGGATACTTGGTCAGGACCTGGTGGCTTTACCAGATG
ATCCTCTCCACAGTGACTGTCCCCTTTATCCTGTGCTGTTGGGTGCTCCCAGAGACACCTTTTTGGCT
TCTCTCAGAGGGACGATATGAAGAAGCACAAAAAATAGTTGACATCATGGCCAAGTGGAACAGGGCAA
GCTCCTGTAAACTGTCAGAACTTTTATCACTGGACCTACAAGGTCCTGTTAGTAATAGCCCCACTGAA

GTTCAGAAGCACAACCTATCATATCTGTTTTATAACTGGAGCATTACGAAAAGGACACTTACCGTTTG
GCTAATCTGGTTCACTGGAAGTTTGGGATTCTACTCGTTTTCCTTGAATTCTGTTAACTTAGGAGGCA
ATGAATACTTAAACCTCTTCCTCCTGGGTGTAGTGGAAATTCCCGCCTACACCTTCGTGTGCATCGCC
ACGGACAAGGTCGGGAGGAGAACAGTCCTGGCCTACTCTCTTTTCTGCAGTGCACTGGCCTGTGGTGT
CGTTATGGTGATCCCCCAGAAACATTATATTTTGGGTGTGGTGACAGCTATGGTTGGAAAATTTGCCA
TCGGGGCAGCATTTGGTCTCATTTATCTTTATACAGCTGAGCTGTATCCAACCATTGTAAGATCGCTG
GCTGTGGGAAGCGGCAGCATGGTGTGTCGCCTGGCCAGCATCCTGGCGCCGTTCTCTGTGGACCTCAG
CAGCATTTGGATCTTCATACCACAGCTCCTAGGACAGCACCTTCAGGAG**TAA**TAATGCAGGCAGTCGC
TGGCAATTAAGGAGGCGCTCCATGGTGGGTCATGA

NOV50b, CG55988-04 Protein Sequence	SEQ ID NO: 778	515 aa	MW at 58030.0kD

MGSRHFEGIYDHVGHFGRFQRVLYFICAFQNISCGIHYLASVFMGVTPHHVCRPPGNVSQVVFHNHSN
WSLEDTGALLSSGQKDYVTVQLQNGEIWELSRCSRNKRENTSSLGYEYTGSKKEFPCVDGYIYDQNTW
KSTAVTQWNLVCDRKWLAMLIQPLFMFGVLLGSVTFGYFSDRLGRRVVLWATSSSMFLFGIAAAFAVD
YYTFMAARFFLAMVASGYLVVGFVYVMEFIGMKSRTWASVHLHSFFAVGTLLVALTGYLVRTWWLYQM
ILSTVTVPFILCCWVLPETPFWLLSEGRYEEAQKIVDIMAKWNRASSCKLSELLSLDLQGPVSNSPTE
VQKHNLSYLFYNWSITKRTLTVWLIWFTGSLGFYSFSLNSVNLGGNEYLNLFLLGVVEIPAYTFVCIA
TDKVGRRTVLAYSLFCSALACGVVMVIPQKHYILGVVTAMVGKFAIGAAFGLIYLYTAELYPTIVRSL
AVGSGSMVCRLASILAPFSVDLSSIWIFIPQLLGQHLQE

NOV50c, CG55988-01 DNA Sequence	SEQ ID NO: 779	2069 bp
	ORF Start: ATG at 279	ORF Stop: TAA at 1881

GCTTCTAGGCCTTCTCAGTAGATGGAGCTAAGTAATATATGTATATATACTAACCCACAGATATAAAT
ATGTCTATAATTATTTCTATATTTATCCATTCGTGTATATGTTAAGATAAACATGATGGAGACCCTTC
AAATTTGCTTATGTTCTTTTTCAGCCTATAGACCAGATATAATAATTAGCTTTTCTTCTCTTGCAGAT
TCCAGAGAGTCCTCTATTTCATATGTGCCTTCCAGAACATCTCTTGTGGTATTCACTACTTGGCTTCT
GTGTTC**ATG**GGAGTCACCCCTCATCATGTCTGCAGGCCCCCAGGCAATGTGAGTCAGGTTGTTTTCCA
TAATCACTCTAATTGGAGTTTGGAGGACACCGGGGCCCTGTTGTCTTCAGGCCAGAAAGATTATGTTA
CGGTGCAGTTGCAGAATGGTGAGATCTGGGAGCTCTCAAGGTGTAGCAGGAATAAGAGGGAGAACACA
TCGAGTTTGGGCTATGAATACACTGGCAGTAAGAAAGAGTTTCCTTGTGTGGATGGCTACATATATGA
CCAGAACACATGGAAAAGCACTGCGGTGACCCAGTGGAACCTGGTCTGTGACCGAAAATGGCTTGCAA
TGCTGATCCAGCCCCTATTTATGTTTGGAGTCCTACTGGGATCGGTGACTTTTGGCTACTTTTCTGAC
AGGCTAGGACGCCGGGTGGTCTTGTGGGCCACAAGCAGTAGCATGTTTTTGTTTGGAATAGCAGCGGC
GTTTGCAGTTGATTATTACACCTTCATGGCTGCTCGCTTTTTTCTTGCCATGGTTGCAAGTGGCTATC
TTGTGGTGGGGTTTGTCTATGTGATGGAATTCATTGGCATGAAGTCTCGGACATGGGCGTCTGTCCAT
TTGCATTCCTTTTTTGCAGTTGGAACCCTGCTGGTGGCTTTGACAGGATACTTGGTCAGGACCTGGTG
GCTTTACCAGATGATCCTCTCCACAGTGACTGTCCCCTTTATCCTGTGCTGTTGGGTGCTCCCAGAGA
CACCTTTTTGGCTTCTCTCAGAGGGACGATATGAAGAAGCACAAAAAATAGTTGACATCATGGCCAAG
TGGAACAGGGCAAGCTCCTGTAAACTGTCAGAACTTTTATCACTGGACCTACAAGGTCCTGTTAGTAA
TAGCCCCACTGAAGTTCAGAAGCACAACCTATCATATCTGTTTTATAACTGGAGCATTACGAAAAGGA
CACTTACCGTTTGGCTAATCTGGTTCACTGGAAGTTTGGGATTCTACTCGTTTTCCTTGAATTCTGTT
AACTTAGGAGGCAATGAATACTTAAACCTCTTCCTCCTGGGTGTAGTGGAAATTCCCGCCTACACCTT
CGTGTGCATCGCCATGGACAAGGTCGGGAGGAGAACAGTCCTGGCCTACTCTCTTTTCTGCAGTGCAC
TGGCCTGTGGTGTCGTTATGGTGATCCCCCAGAAACATTATATTTTGGGTGTGGTGACAGCTATGGTT
GGAAAATTTGCCATCGGGGCAGCATTTGGCCTCATTTATCTTTATACAGCTGAGCTGTATCCAACCAT
TGTAAGATCGCTGGCTGTGGGAAGCGGCAGCATGGTGTGTCGCCTGGCCAGCATCCTGGCGCCGTTCT
CTGTGGACCTCAGCAGCATTTGGATCTTCATACCACAGTTGTTTGTTGGGACTATGGCCCTCCTGAGT
GGAGTGTTAACACTAAAGCTTCCAGAAACCCTTGGGAAACGGCTAGCAACTACTTGGGAGGAGGCTGC
AAAACTGGAGTCAGAGAATGAAAGCAAGTCAAGCAAATTACTTCTCACAACTAATAATAGTGGGCTGG
AAAAAACGGAAGCGATTACCCCCAGGGATTCTGGTCTTGGTGAA**TAA**ATGTGCCATGCCTGCTGTCTA
GCACCTGAAATATTATTTACCCTAATGCCTTTGTATTAGAGGAATCTTATTCTCATCTCCCATATGTT
GTTTGTATGTCTTTTTAATAAATTTTGTAAGAAAATTTTAAAGCAAATATGTTATAAAAGAAATAAAA
ACTAAGATGAAAATTCTCAGTTTTAAAAA

NOV50c, CG55988-01 Protein Sequence	SEQ ID NO: 780	534 aa	MW at 59606.7kD

MGVTPHHVCRPPGNVSQVVFHNHSNWSLEDTGALLSSGQKDYVTVQLQNGEIWELSRCSRNKRENTSS
LGYEYTGSKKEFPCVDGYIYDQNTWKSTAVTQWNLVCDRKWLAMLIQPLFMFGVLLGSVTFGYFSDRL

GRRVVLWATSSSMFLFGIAAAFAVDYYTFMAARFFLAMVASGYLVVGFVYVMEFIGMKSRTWASVHLH
SFFAVGTLLVALTGYLVRTWWLYQMILSTVTVPFILCCWVLPETPFWLLSEGRYEEAQKIVDIMAKWN
RASSCKLSELLSLDLQGPVSNSPTEVQKHNLSYLFYNWSITKRTLTVWLIWFTGSLGFYSFSLNSVNL
GGNEYLNLFLLGVVEIPAYTFVCIAMDKVGRRTVLAYSLFCSALACGVVMVIPQKHYILGVVTAMVGK
FAIGAAFGLIYLYTAELYPTIVRSLAVGSGSMVCRLASILAPFSVDLSSIWIFIPQLFVGTMALLSGV
LTLKLPETLGKRLATTWEEAAKLESENESKSSKLLLLTTNNSGLEKTEAITPRDSGLGE

NOV50d, CG55988-02 DNA Sequence	SEQ ID NO: 781	1666 bp
	ORF Start: ATG at 76	ORF Stop: TAA at 1654

TTCCAGAGAGTCCTCTATTTCATATGTGCCTTCCAGAACATCTCTTGTGGTATTCACTACTTGGCTTC
TGTGTTCATGGGAGTCACCCCTCATCATGTCTGCAGGCCCCCAGGCAATGTGAGTCAGGTTGTTTTCC
ATAATCACTCTAATTGGAGTTTGGAGGACACCGGGGCCCTGTTGTCTTCAGGCCAGAAAGATTATGTT
ACGGTGCAGTTGCAGAATGGTGAGATCTGGGAGCTCTCAAGGTGTAGCAGGAATAAGAGGGAGAACAC
ATCGAGTTTGGGCTATGAATACACTGGCAGTAAGAAAGAGTTTCCTTGTGTGGATGGCTACATATATG
ACCAGAACACATGGAAAAGCACTGCGGTGACCCAGTGGAACCTGGTCTGTGACCGAAAATGGCTTGCA
ATGCTGATCCAGCCCCTATTTATGTTTGGAGTCCTACTGGGATCGGTGACTTTTGGCTACTTTTCTGA
CAGGCTTTTTTGCCTATATGTGATTTGCAATGGGGTCAGACTCCTCAATAGTTATAAATGTGACCTTG
AATATAAATCCCTATTATTTGTTTTTCAGGTTGCAAGTGGCTATCTTGTGGTGGGGTTTGTCTATGTG
ATGGAATTCATTGGCATGAAGTCTCGGACATGGGCGTCTGTCCATTTGCATTCCTTTTTTGCAGTTGG
AACCCTGCTGGTGGCTTTGACAGGATACTTGGTCAGGACCTGGTGGCTTTACCAGATGATCCTCTCCA
CAGTGACTGTCCCCTTTATCCTGTGCTGTTGGGTGCTCCCAGAGACACCTTTTTGGCTTCTCTCAGAG
GGACGATATGAAGAAGCACAAAAAATAGTTGACATCATGGCCAAGTGGAACAGGGCAAGCTCCTGTAA
ACTGTCAGAACTTTTATCACTGGACCTACAAGGTCCTGTTAGTAATAGCCCCACTGAAGTTCAGAAGC
ACAACCTATCATATCTGTTTTATAACTGGAGCATTACGAAAAGGACACTTACCGTTTGGCTAATCTGG
TTCACTGGAAGTTTGGGATTCTACTCGTTTTCCTTGAATTCTGTTAACTTAGGAGGCAATGAATACTT
AAACCTCTTCCTCACAGGTGTAGTGGAAATTCCCGCCTACACCTTCGTGTGCATCGCCATGGACAAGG
TCGGGAGGAGAACAGTCCTGGCCTACTCTCTTTTCTGCAGTGCACTGGCCTGTGGTGTCGTTATGGTG
ATCCCCCAGGTGAGTTATCTTCTGGGTGTGGTGACAGCTATGGTTGGAAAATTTGCCATCGGGGCAGC
ATTTGGCCTCATTTATCTTTATACAGCTGAGCTGTATCCAACCATTGTAAGGTCGCTGGCTGTGGGAA
GCGGCAGCATGGTGTGTCGCCTGGCCAGCATCCTGGCGCCGTTCTCTGTGGACCTCAGCAGCATTTGG
ATCTTCATACCACAGTTGTTTGTTGGGACTATGGCCCTCCTGAGTGGAGTGTTAACACTAAAGCTTCC
AGAAACCCTTGGGAAACGGCTAGCAACTACTTGGGAGGAGGCTGCAAAACTGGAGTCAGAGAATGAAA
GCAAGTCAAGCAAATTACTTCTCACAACTAATAATAGTGGGCTGGAAAAAACGGAAGCGATTACCCCC
AGGGATTCTGGTCTTGGTGAATAAATGTGCCATG

NOV50d, CG55988-02 Protein Sequence	SEQ ID NO: 782	526 aa	MW at 58820.8kD

MGVTPHHVCRPPGNVSQVVFHNHSNWSLEDTGALLSSGQKDYVTVQLQNGEIWELSRCSRNKRENTSS
LGYEYTGSKKEFPCVDGYIYDQNTWKSTAVTQWNLVCDRKWLAMLIQPLFMFGVLLGSVTFGYFSDRL
FCLYVICNGVRLLNSYKCDLEYKSLLFVFQVASGYLVVGFVYVMEFIGMKSRTWASVHLHSFFAVGTL
LVALTGYLVRTWWLYQMILSTVTVPFILCCWVLPETPFWLLSEGRYEEAQKIVDIMAKWNRASSCKLS
ELLSLDLQGPVSNSPTEVQKHNLSYLFYNWSITKRTLTVWLIWFTGSLGFYSFSLNSVNLGGNEYLNL
FLTGVVEIPAYTFVCIAMDKVGRRTVLAYSLFCSALACGVVMVIPQVSYLLGVVTAMVGKFAIGAAFG
LIYLYTAELYPTIVRSLAVGSGSMVCRLASILAPFSVDLSSIWIFIPQLFVGTMALLSGVLTLKLPET
LGKRLATTWEEAAKLESENESKSSKLLLLTTNNSGLEKTEAITPRDSGLGE

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 50B.

Table 50B. Comparison of the NOV50 protein sequences.

```
NOV50a  MGSRHFEGIYDHVGHFGRFQRVLYFICAFQNISCGIHYLASVFMGVTPHHVCRPPGNVSQ
NOV50b  MGSRHFEGIYDHVGHFGRFQRVLYFICAFQNISCGIHYLASVFMGVTPHHVCRPPGNVSQ
NOV50c  ------------------------------------------MGVTPHHVCRPPGNVSQ
NOV50d  ------------------------------------------MGVTPHHVCRPPGNVSQ

NOV50a  VVFHNHSNWSLEDTGALLSSGQKDYVTVQLQNGEIWELSRCSRNKRENTSSLGYEYTGSK
```

```
NOV50b  VVFHNHSNWSLEDTGALLSSGQKDYVTVQLQNGEIWELSRCSRNKRENTSSLGYEYTGSK
NOV50c  VVFHNHSNWSLEDTGALLSSGQKDYVTVQLQNGEIWELSRCSRNKRENTSSLGYEYTGSK
NOV50d  VVFHNHSNWSLEDTGALLSSGQKDYVTVQLQNGEIWELSRCSRNKRENTSSLGYEYTGSK

NOV50a  KEFPCVDGYIYDQNTWKSTAVTQWNLVCDRKWLAMLIQPLFMFGVLLGSVTFGYFSDRLG
NOV50b  KEFPCVDGYIYDQNTWKSTAVTQWNLVCDRKWLAMLIQPLFMFGVLLGSVTFGYFSDRLG
NOV50c  KEFPCVDGYIYDQNTWKSTAVTQWNLVCDRKWLAMLIQPLFMFGVLLGSVTFGYFSDRLG
NOV50d  KEFPCVDGYIYDQNTWKSTAVTQWNLVCDRKWLAMLIQPLFMFGVLLGSVTFGYFSDRLF

NOV50a  RRVVLWATSSSMFLFGIAAAFAVDYYTFMAARFFLAMVASGYLVVGFVYVMEFIGMKSRT
NOV50b  RRVVLWATSSSMFLFGIAAAFAVDYYTFMAARFFLAMVASGYLVVGFVYVMEFIGMKSRT
NOV50c  RRVVLWATSSSMFLFGIAAAFAVDYYTFMAARFFLAMVASGYLVVGFVYVMEFIGMKSRT
NOV50d  CLYVICN--------GVRLLNSYKCDLEYKSLLFVFQVASGYLVVGFVYVMEFIGMKSRT

NOV50a  WASVHLHSFFAVGTLLVALTGYLVRTWWLYQMILSTVTVPFILCCWVLPETPFWLLSEGR
NOV50b  WASVHLHSFFAVGTLLVALTGYLVRTWWLYQMILSTVTVPFILCCWVLPETPFWLLSEGR
NOV50c  WASVHLHSFFAVGTLLVALTGYLVRTWWLYQMILSTVTVPFILCCWVLPETPFWLLSEGR
NOV50d  WASVHLHSFFAVGTLLVALTGYLVRTWWLYQMILSTVTVPFILCCWVLPETPFWLLSEGR

NOV50a  YEEAQKIVDIMAKWNRASSCKLSELLSLDLQGPVSNSPTEVQKHNLSYLFYNWSITKRTL
NOV50b  YEEAQKIVDIMAKWNRASSCKLSELLSLDLQGPVSNSPTEVQKHNLSYLFYNWSITKRTL
NOV50c  YEEAQKIVDIMAKWNRASSCKLSELLSLDLQGPVSNSPTEVQKHNLSYLFYNWSITKRTL
NOV50d  YEEAQKIVDIMAKWNRASSCKLSELLSLDLQGPVSNSPTEVQKHNLSYLFYNWSITKRTL

NOV50a  TVWLIWFTGSLGFYSFSLNSVNLGGNEYLNLFLLGVVEIPAYTFVCIATDKVGRRTVLAY
NOV50b  TVWLIWFTGSLGFYSFSLNSVNLGGNEYLNLFLLGVVEIPAYTFVCIATDKVGRRTVLAY
NOV50c  TVWLIWFTGSLGFYSFSLNSVNLGGNEYLNLFLLGVVEIPAYTFVCIAMDKVGRRTVLAY
NOV50d  TVWLIWFTGSLGFYSFSLNSVNLGGNEYLNLFLTGVVEIPAYTFVCIAMDKVGRRTVLAY

NOV50a  SLFCSALACGVVMVIPQI--------AGCGKRQHGVSPG--------QHP--------G-
NOV50b  SLFCSALACGVVMVIPQKHYILGVVTAMVGKFAIGAAFGLIYLYTAELYPTIVRSLAVGS
NOV50c  SLFCSALACGVVMVIPQKHYILGVVTAMVGKFAIGAAFGLIYLYTAELYPTIVRSLAVGS
NOV50d  SLFCSALACGVVMVIPQVSYLLGVVTAMVGKFAIGAAFGLIYLYTAELYPTIVRSLAVGS

NOV50a  -AVLCGPQQHLDLHTTVVC--WDYGP------------------PE---------WSVNT
NOV50b  GSMVCRLASILAPFSVDLSSIWIFIPQLLGQHLQE------------------------
NOV50c  GSMVCRLASILAPFSVDLSSIWIFIPQLFVGTMALLSGVLTLKLPETLGKRLATTWEEAA
NOV50d  GSMVCRLASILAPFSVDLSSIWIFIPQLFVGTMALLSGVLTLKLPETLGKRLATTWEEAA

NOV50a  KASRNPWETASNYLGGGCKTGVRE-------------
NOV50b  ------------------------------------
NOV50c  KLESENESKSSKLLLTTNNSGLEKTEAITPRDSGLGE
NOV50d  KLESENESKSSKLLLTTNNSGLEKTEAITPRDSGLGE

NOV50a  (SEQ ID NO:  776)
NOV50b  (SEQ ID NO:  778)
NOV50c  (SEQ ID NO:  780)
NOV50d  (SEQ ID NO:  782)
```

Further analysis of the NOV50a protein yielded the following properties shown in Table 50C.

Table 50C. Protein Sequence Properties NOV50a	
SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 1;  neg.chg 2
      H-region:  length 6;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -1.51
      possible cleavage site: between 34 and 35

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   7
      INTEGRAL    Likelihood = -3.03   Transmembrane  154 - 170
      INTEGRAL    Likelihood = -1.12   Transmembrane  183 - 199
      INTEGRAL    Likelihood = -4.57   Transmembrane  214 - 230
      INTEGRAL    Likelihood = -3.08   Transmembrane  248 - 264
      INTEGRAL    Likelihood = -5.89   Transmembrane  272 - 288
      INTEGRAL    Likelihood = -4.51   Transmembrane  391 - 407
      INTEGRAL    Likelihood = -5.10   Transmembrane  419 - 435
      PERIPHERAL  Likelihood =  2.81 (at 360)
      ALOM score:  -5.89  (number of TMSs: 7)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 161
      Charge difference:  0.0   C( 2.0) - N( 2.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:        1       Hyd Moment(75):  7.38
      Hyd Moment(95):  9.73     G content:       2
      D/E content:     2        S/T content:     1
      Score: -6.19

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 14  SRH|FE

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.5%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none
```

```
ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: GSRH

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:
      Leucine zipper pattern (PS00029):  *** found ***
         LVCDRKWLAMLIQPLFMFGVLL at 146
 none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

---------------------------

Final Results (k = 9/23):

       77.8 %: endoplasmic reticulum
       22.2 %: mitochondrial

>> prediction for CG55988-03 is end (k=9)
```

A search of the NOV50a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 50D.

Table 50D. Geneseq Results for NOV50a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV50a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABP74100	Human TRICH SEQ ID NO 5 - Homo sapiens, 577 aa. [WO200246415-A2, 13-JUN-2002]	1..437 1..437	436/437 (99%) 436/437 (99%)	0.0
AAM00930	Human bone marrow protein, SEQ ID NO: 406 - Homo sapiens, 584 aa. [WO200153453-A2, 26-JUL-2001]	1..437 7..443	436/437 (99%) 436/437 (99%)	0.0
AAM78367	Human protein SEQ ID NO 1029 - Homo sapiens, 577 aa. [WO200157190-A2, 09-AUG-2001]	1..437 1..437	436/437 (99%) 436/437 (99%)	0.0
AAM79351	Human protein SEQ ID NO 2997 - Homo sapiens, 585 aa. [WO200157190-A2, 09-AUG-2001]	1..437 7..444	432/438 (98%) 432/438 (98%)	0.0
AAB43038	Human ORFX ORF2802 polypeptide sequence SEQ ID NO:5604 - Homo sapiens, 560 aa. [WO200058473-A2, 05-OCT-2000]	18..437 1..420	419/420 (99%) 419/420 (99%)	0.0

In a BLAST search of public sequence databases, the NOV50a protein was found to have homology to the proteins shown in the BLASTP data in Table 50E.

Table 50E. Public BLASTP Results for NOV50a				
Protein Accession Number	**Protein/Organism/Length**	**NOV50a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96RU0	Organic cation transporter OKB1 - Homo sapiens (Human), 577 aa.	1..437 1..437	437/437 (100%) 437/437 (100%)	0.0
AAH47565	Organic cation transporter OKB1 - Homo sapiens (Human), 577 aa.	1..437 1..437	436/437 (99%) 436/437 (99%)	0.0
Q8IZD5	Putative transmembrane transporter FLIPT 2 - Homo sapiens (Human), 577 aa.	1..437 1..437	436/437 (99%) 436/437 (99%)	0.0
O14567	WUGSC:RG331P03.1 protein - Homo sapiens (Human), 456 aa (fragment).	18..437 1..420	419/420 (99%) 419/420 (99%)	0.0
Q8IUG8	Carnitine transporter 2 - Homo sapiens (Human), 543 aa.	18..437 16..403	387/420 (92%) 387/420 (92%)	0.0

PFam analysis indicates that the NOV50a protein contains the domains shown in the Table 50F.

Table 50F. Domain Analysis of NOV50a			
Pfam Domain	**NOV50a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
sugar_tr	120..499	67/493 (14%) 242/493 (49%)	7.8e-05

Example 51.

The NOV51 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 51A.

Table 51A. NOV51 Sequence Analysis		
NOV51a, CG56071-01 DNA Sequence	SEQ ID NO: 783	3092 bp
	ORF Start: ATG at 1	ORF Stop: TAG at 3073
ATGGAGCCCTCCAGAGCGCTTCTCGGCTGCCTAGCGAGCGCCGCCGCTGCCGCCCCGCCGGGGGAGGA TGGAGCAGGGGCCGGGGCCGAGGAGGAGGAGGAGGAGGAGGAGGCGGCGGCGGCGGTGGGCCCCG GGGAGCTGGGCTGCGACGCGCCGCTGCCCTACTGGACGGCCGTGTTCGAGTACGAGGCGGCGGGCGAG GACGAGCTGACCCTGCGGCTGGGCGACGTGGTGGAGGTGCTGTCCAAGGACTCGCAGGTGTCCGGCGA		

CGAGGGCTGGTGGACCGGGCAGCTGAACCAGCGGGTGGGCATCTTCCCCAGCAACTACGTGACCCCGC
GCAGCGCCTTCTCCAGCCGCTGCCAGCCCGGCGGCGAAATTGATTTTGCGGAGCTCACCTTGGAAGAG
ATTATTGGCATCGGGGGCTTTGGGAAGGTCTATCGTGCTTTCTGGATAGGGGATGAGGTTGCTGTGAA
AGCAGCTCGCCACGACCCTGATGAGGACATCAGCCAGACCATAGAGAATGTTCGCCAAGAGGCCAAGC
TCTTCGCCATGCTGAAGCACCCCAACATCATTGCCCTAAGAGGGGTATGTCTGAAGGAGCCCAACCTC
TGCTTGGTCATGGAGTTTGCTCGTGGAGGACCTTTGAATAGAGTGTTATCTGGGAAAAGGATTCCCCC
AGACATCCTGGTGAATTGGGCTGTGCAGATTGCCAGAGGGATGAACTACTTACATGATGAGGCAATTG
TTCCCATCATCCACCGCGACCTTAAGTCCAGCAACGTATTGATCCTCCAGAAGGTGGAGAATGGAGAC
CTGAGCAACAAGATTCTGAAGATCACTGATTTTGGCCTGGCTCGGGAATGGCACCGAACCACCAAGAT
GAGTGCGGCAGGGACGTATGCTTGGATGGCACCCGAAGTCATCCGGGCCTCCATGTTTTCCAAAGGCA
GTGATGTGTGGAGCTATGGGGTGCTACTTTGGGAGTTGCTGACTGGTGAGGTGCCCTTTCGAGGCATT
GATGGCTTAGCAGTCGCTTATGGAGTGGCCATGAACAAACTCGCCCTTCCTATTCCTTCTACGTGCCC
AGAACCTTTTGCCAAACTCATGGAAGACTGCTGGAATCCTGATCCCCACTCACGACCATCTTTCACGA
ATATCCTGGACCAGCTAACCACCATAGAGGAGTCTGGTTTCTTTGAAATGCCCAAGGACTCCTTCCAC
TGCCTGCAGGACAACTGGAAACACGAGATTCAGGAGATGTTTGACCAACTCAGGGCCAAAGAAAAGGA
ACTTCGCACCTGGGAGGAGGAGCTGACGCGGGCTGCACTGCAGCAGAAGAACCAGGAGGAACTGCTGC
GGCGTCGGGAGCAGGAGCTGGCCGAGCGGGAGATTGACATCCTGGAACGGGAGCTCAACATCATCATC
CACCAGCTGTGCCAGGAGAAGCCCCGGGTGAAGAAACGCAAGGGCAAGTTCAGGAAGAGCCGGCTGAA
GCTCAAGGATGGCAACCGCATCAGCCTCCCTTCTGGTTTCCAGCACAAGTTCACGGTGCAGGCCTCCC
CTACCATGGATAAAAGGAAGAGTCTTATCAACAGCCGCTCCAGTCCTCCTGCAAGCCCCACCATCATT
CCTCGCCTTCGAGCCATCCAGTGTGAGACTGTTTCCCAAATTAGCTGGGGCCAGAACACACAGGGGCA
CCTGTCCGAAAGCAGCAAAACCTGGGGCAGGAGCTCAGTCGTCCCAAAGGAGGAAGGGGAGGAGGAGG
AGAAGAGGGCCCCAAAGAAGAAGGGACGGACGTGGGGGCCAGGGACGCTTGGTCAGAAGGAGCTTGCC
TCGGGAGATGAACTCAAGTCCCTGGTAGATGGATATAAGCAGTGGTCGTCCAGTGCCCCCAACCTGGT
GAAGGGCCCAAGGAGTACCCCGGCCCTGCCAGGGTTCACCAGCCTTATGGAGATGGAGGATGAGGACA
GTGAAGGCCCAGGGAGTGGAGAGAGTCGCCTACAGCATTCACCCAGCCAGTCCTACCTCTGTATCCCA
TTCCCTCGTGGAGAGCCCACCCCAGTCAACTCGGCCACGAGTACCCCTCAGCTGACGCCAACCAACAG
CCTCAAGCGGGGCGGTGCCCACCACCGCCGCTGCGAGGTGGCTCTGCTCGGCTGTGGGGCTGTTCTGG
CAGCCACAGGCCTAGGGTTTGACTTGCTGGAAGCTGGCAAGTGCCAGCTGCTTCCCCTGGAGGAGCCT
GAGCCACCAGCCCGGGAGGAGAAGAAAAGACGGGAGGGTCTTTTTCAGAGGTCCAGCCGTCCTCGTCG
GAGCACCAGCCCCCCATCCCGAAAGCTTTTCAAGAAGGAGGAGCCCATGCTGTTGCTAGGAGACCCCT
CTGCCTCCCTGACGCTGCTCTCCCTCTCCTCCATCTCCGAGTGCAACTCCACACGCTCCCTGCTGCAG
TCCGACAGCGATGAAATTGTCGTGTATGAGATGCCAGTCAGCCCAGTCGAGGCCCCTCCCCTGAGTCC
ATGTACCCACAACCCCCTGGTCAATGTCCGAGTAGAGCGCTTCAAACGAGATCCTAACCAATCTCTGA
CTCCCACCCATGTCACCCTCACCACCCCCTCGCAGCCCAGCAGTCACCGGCGGACTCCTTCTGATGGG
GCCCTTAAGCCAGAGACTCTCCTAGCCAGCAGGAGCCCCAGTCCCAGCCGAGACCCAGGTGAATTCCC
CCGTCTCCCTGACCCCAATGTGGTCTTCCCCCCAACCCCAAGGCGCTGGAACACTCAGCAGGACTCTA
CCTTGGAGAGACCCAAGACTCTGGAGTTTCTGCCTCGGCCGCGTCCTTCTGCCAACCGGCAACGGCTG
GACCCTTGGTGGTTTGTGTCCCCCAGCCATGCCCGCAGCACCTCCCCAGCCAACAGCTCCAGCACAGA
GACGCCCGGGCCGCTGCCCCCGACTGAGCGGACGCTCCTGGACCTGGATGCAGAGGGGCAGAGTCAGG
ACAGCACCGTGCCGCTGTGCAGAGCGGAACTGAACACACACAGGCCTGCCCCTTATGAGATCCAGCAG
GAGTTCTGGTCT**TAG**CACGAAAAGGATTGGGG

NOV51a, CG56071-01 Protein Sequence	SEQ ID NO: 784	1024 aa	MW at 113682.1kD

MEPSRALLGCLASAAAAAPPGEDGAGAGAEEEEEEEEEAAAAVGPGELGCDAPLPYWTAVFEYEAAGE
DELTLRLGDVVEVLSKDSQVSGDEGWWTGQLNQRVGIFPSNYVTPRSAFSSRCQPGGEIDFAELTLEE
IIGIGGFGKVYRAFWIGDEVAVKAARHDPDEDISQTIENVRQEAKLFAMLKHPNIIALRGVCLKEPNL
CLVMEFARGGPLNRVLSGKRIPPDILVNWAVQIARGMNYLHDEAIVPIIHRDLKSSNVLILQKVENGD
LSNKILKITDFGLAREWHRTTKMSAAGTYAWMAPEVIRASMFSKGSDVWSYGVLLWELLTGEVPFRGI
DGLAVAYGVAMNKLALPIPSTCPEPFAKLMEDCWNPDPHSRPSFTNILDQLTTIEESGFFEMPKDSFH
CLQDNWKHEIQEMFDQLRAKEKELRTWEEELTRAALQQKNQEELLRRREQELAEREIDILERELNIII
HQLCQEKPRVKKRKGKFRKSRLKLKDGNRISLPSGFQHKFTVQASPTMDKRKSLINSRSSPPASPTII
PRLRAIQCETVSQISWGQNTQGHLSESSKTWGRSSVVPKEEGEEEEKRAPKKKGRTWGPGTLGQKELA
SGDELKSLVDGYKQWSSSAPNLVKGPRSTPALPGFTSLMEMEDEDSEGPGSGESRLQHSPSQSYLCIP
FPRGEPTPVNSATSTPQLTPTNSLKRGGAHHRRCEVALLGCGAVLAATGLGFDLLEAGKCQLLPLEEP
EPPAREEKKRREGLFQRSSRPRRSTSPPSRKLFKKEEPMLLLGDPSASLTLLSLSSISECNSTRSLLQ
SDSDEIVVYEMPVSPVEAPPLSPCTHNPLVNVRVERFKRDPNQSLTPTHVTLTTPSQPSSHRRTPSDG
ALKPETLLASRSPSPSRDPGEFPRLPDPNVVFPPTPRRWNTQQDSTLERPKTLEFLPRPRPSANRQRL

DPWWFVSPSHARSTSPANSSSTETPGPLPPTERTLLDLDAEGQSQDSTVPLCRAELNTHRPAPYEIQQ
EFWS

NOV51b, 274082270 DNA Sequence	SEQ ID NO: 785	807 bp
	ORF Start: at 1	ORF Stop: end of sequence

ACCGGATCCCTCACCTTGGAAGAGATTATTGGCATCGGGGGCTTTGGGAAGGTCTATCGTGCTTTCTG
GATAGGGGATGAGGTTGCTGTGAAAGCAGCTCGCCACGACCCTGATGAGGACATCAGCCAGACCATAG
AGAATGTTCGCCAAGAGGCCAAGCTCTTCGCCATGCTGAAGCACCCCAACATCATTGCCCTAAGAGGG
GTATGTCTGAAGGAGCCCAACCTCTGCTTGGTCATGGAGTTTGCTCGTGGAGGACCTTTGAATAGAGT
GTTATCTGGGAAAAGGATTCCCCCAGACATCCTGGTGAATTGGGCTGTGCAGATTGCCAGAGGGATGA
ACTACTTACATGATGAGGCAATTGTTCCCATCATCCACCGCGACCTTAAGTCCAGCAACATATTGATC
CTCCAGAAGGTGGAGAATGGAGACCTGAGCAACAAGATTCTGAAGATCACTGATTTTGGCCTGGCTCG
GGAATGGCACCGAACCACCAAGATGAGTGCGGCAGGGACGTATGCTTGGATGGCACCCGAAGTCATCC
GGGCCTCCATGTTTTCCAAAGGCAGTGATGTGTGGAGCTATGGGGTGCTACTTTGGGAGTTGCTGACT
GGTGAGGTGCCCTTTCGAGGCATTGATGGCTTAGCAGTCGCTTATGGAGTGGCCATGAACAAACTCGC
CCTTCCTATTCCTTCTACGTGCCCAGAACCTTTTGCCAAACTCATGGAAGACTGCTGGAATCCTGATC
CCCACTCACGACCATCTTTCACGAATATCCTGGACCAGCTAACCACCATACTCGAGGGC

NOV51b, 274082270 Protein Sequence	SEQ ID NO: 786	269 aa	MW at 29861.3kD

TGSLTLEEIIGIGGFGKVYRAFWIGDEVAVKAARHDPDEDISQTIENVRQEAKLFAMLKHPNIIALRG
VCLKEPNLCLVMEFARGGPLNRVLSGKRIPPDILVNWAVQIARGMNYLHDEAIVPIIHRDLKSSNILI
LQKVENGDLSNKILKITDFGLAREWHRTTKMSAAGTYAWMAPEVIRASMFSKGSDVWSYGVLLWELLT
GEVPFRGIDGLAVAYGVAMNKLALPIPSTCPEPFAKLMEDCWNPDPHSRPSFTNILDQLTTILEG

NOV51c, SNP13376041 of CG56071-01, DNA Sequence	SEQ ID NO: 787	3092 bp
	ORF Start: ATG at 1	ORF Stop: TAG at 3073
	SNP Pos: 2712	SNP Change: T to C

ATGGAGCCCTCCAGAGCGCTTCTCGGCTGCCTAGCGAGCGCCGCCGCTGCCGCCCCGCCGGGGGAGGA
TGGAGCAGGGGCCGGGGCCGAGGAGGAGGAGGAGGAGGAGGAGGCGGCGGCGGCGGTGGGCCCCG
GGGAGCTGGGCTGCGACGCGCCGCTGCCCTACTGGACGGCCGTGTTCGAGTACGAGGCGGCGGGCGAG
GACGAGCTGACCCTGCGGCTGGGCGACGTGGTGGAGGTGCTGTCCAAGGACTCGCAGGTGTCCGGCGA
CGAGGGCTGGTGGACCGGGCAGCTGAACCAGCGGGTGGGCATCTTCCCCAGCAACTACGTGACCCCGC
GCAGCGCCTTCTCCAGCCGCTGCCAGCCCGGCGGCGAAATTGATTTTGCGGAGCTCACCTTGGAAGAG
ATTATTGGCATCGGGGGCTTTGGGAAGGTCTATCGTGCTTTCTGGATAGGGGATGAGGTTGCTGTGAA
AGCAGCTCGCCACGACCCTGATGAGGACATCAGCCAGACCATAGAGAATGTTCGCCAAGAGGCCAAGC
TCTTCGCCATGCTGAAGCACCCCAACATCATTGCCCTAAGAGGGGTATGTCTGAAGGAGCCCAACCTC
TGCTTGGTCATGGAGTTTGCTCGTGGAGGACCTTTGAATAGAGTGTTATCTGGGAAAAGGATTCCCCC
AGACATCCTGGTGAATTGGGCTGTGCAGATTGCCAGAGGGATGAACTACTTACATGATGAGGCAATTG
TTCCCATCATCCACCGCGACCTTAAGTCCAGCAACGTATTGATCCTCCAGAAGGTGGAGAATGGAGAC
CTGAGCAACAAGATTCTGAAGATCACTGATTTTGGCCTGGCTCGGGAATGGCACCGAACCACCAAGAT
GAGTGCGGCAGGGACGTATGCTTGGATGGCACCCGAAGTCATCCGGGCCTCCATGTTTTCCAAAGGCA
GTGATGTGTGGAGCTATGGGGTGCTACTTTGGGAGTTGCTGACTGGTGAGGTGCCCTTTCGAGGCATT
GATGGCTTAGCAGTCGCTTATGGAGTGGCCATGAACAAACTCGCCCTTCCTATTCCTTCTACGTGCCC
AGAACCTTTTGCCAAACTCATGGAAGACTGCTGGAATCCTGATCCCCACTCACGACCATCTTTCACGA
ATATCCTGGACCAGCTAACCACCATAGAGGAGTCTGGTTTCTTTGAAATGCCCAAGGACTCCTTCCAC
TGCCTGCAGGACAACTGGAAACACGAGATTCAGGAGATGTTTGACCAACTCAGGGCCAAAGAAAAGGA
ACTTCGCACCTGGGAGGAGGAGCTGACGCGGGCTGCACTGCAGCAGAAGAACCAGGAGGAACTGCTGC
GGCGTCGGGAGCAGGAGCTGGCCGAGCGGGAGATTGACATCCTGGAACGGGAGCTCAACATCATCATC
CACCAGCTGTGCCAGGAGAAGCCCCGGGTGAAGAAACGCAAGGGCAAGTTCAGGAAGAGCCGGCTGAA
GCTCAAGGATGGCAACCGCATCAGCCTCCCTTCTGGTTTCCAGCACAAGTTCACGGTGCAGGCCTCCC
CTACCATGGATAAAAGGAAGAGTCTTATCAACAGCCGCTCCAGTCCTCCTGCAAGCCCCACCATCATT
CCTCGCCTTCGAGCCATCCAGTGTGAGACTGTTTCCCAAATTAGCTGGGGCCAGAACACACAGGGGCA
CCTGTCCGAAAGCAGCAAAACCTGGGGCAGGAGCTCAGTCGTCCCAAAGGAGGAAGGGGAGGAGGAGG
AGAAGAGGGCCCCAAAGAAGAAGGGACGGACGTGGGGGCCAGGGACGCTTGGTCAGAAGGAGCTTGCC
TCGGGAGATGAACTCAAGTCCCTGGTAGATGGATATAAGCAGTGGTCGTCCAGTGCCCCCAACCTGGT
GAAGGGCCCAAGGAGTACCCCGGCCCTGCCAGGGTTCACCAGCCTTATGGAGATGGAGGATGAGGACA
GTGAAGGCCCAGGGAGTGGAGAGAGTCGCCTACAGCATTCACCCAGCCAGTCCTACCTCTGTATCCCA

TTCCCTCGTGGAGAGCCCACCCCAGTCAACTCGGCCACGAGTACCCCTCAGCTGACGCCAACCAACAG
CCTCAAGCGGGGCGGTGCCCACCACCGCCGCTGCGAGGTGGCTCTGCTCGGCTGTGGGGCTGTTCTGG
CAGCCACAGGCCTAGGGTTTGACTTGCTGGAAGCTGGCAAGTGCCAGCTGCTTCCCCTGGAGGAGCCT
GAGCCACCAGCCCGGGAGGAGAAGAAAAGACGGGAGGGTCTTTTTCAGAGGTCCAGCCGTCCTCGTCG
GAGCACCAGCCCCCCATCCCGAAAGCTTTTCAAGAAGGAGGAGCCCATGCTGTTGCTAGGAGACCCCT
CTGCCTCCCTGACGCTGCTCTCCCTCTCCTCCATCTCCGAGTGCAACTCCACACGCTCCCTGCTGCAG
TCCGACAGCGATGAAATTGTCGTGTATGAGATGCCAGTCAGCCCAGTCGAGGCCCCTCCCCTGAGTCC
ATGTACCCACAACCCCCTGGTCAATGTCCGAGTAGAGCGCTTCAAACGAGATCCTAACCAATCTCTGA
CTCCCACCCATGTCACCCTCACCACCCCCTCGCAGCCCAGCAGTCACCGGCGGACTCCTTCTGATGGG
GCCCTTAAGCCAGAGACTCTCCTAGCCAGCAGGAGCCCCAGTCCCAGCCGAGACCCAGGCGAATTCCC
CCGTCTCCCTGACCCCAATGTGGTCTTCCCCCCAACCCCAAGGCGCTGGAACACTCAGCAGGACTCTA
CCTTGGAGAGACCCAAGACTCTGGAGTTTCTGCCTCGGCCGCGTCCTTCTGCCAACCGGCAACGGCTG
GACCCTTGGTGGTTTGTGTCCCCCAGCCATGCCCGCAGCACCTCCCCAGCCAACAGCTCCAGCACAGA
GACGCCCGGGCCGCTGCCCCCGACTGAGCGGACGCTCCTGGACCTGGATGCAGAGGGGCAGAGTCAGG
ACAGCACCGTGCCGCTGTGCAGAGCGGAACTGAACACACACAGGCCTGCCCCTTATGAGATCCAGCAG
GAGTTCTGGTCTTAGCACGAAAAGGATTGGGG

NOV51c, SNP13376041 of CG56071-01, Protein Sequence	SEQ ID NO: 788	1024 aa	MW at 113682.1kD
	SNP Pos: 904		SNP Change: Gly to Gly

MEPSRALLGCLASAAAAAPPGEDGAGAGAEEEEEEEEEAAAAVGPGELGCDAPLPYWTAVFEYEAAGE
DELTLRLGDVVEVLSKDSQVSGDEGWWTGQLNQRVGIFPSNYVTPRSAFSSRCQPGGEIDFAELTLEE
IIGIGGFGKVYRAFWIGDEVAVKAARHDPDEDISQTIENVRQEAKLFAMLKHPNIIALRGVCLKEPNL
CLVMEFARGGPLNRVLSGKRIPPDILVNWAVQIARGMNYLHDEAIVPIIHRDLKSSNVLILQKVENGD
LSNKILKITDFGLAREWHRTTKMSAAGTYAWMAPEVIRASMFSKGSDVWSYGVLLWELLTGEVPFRGI
DGLAVAYGVAMNKLALPIPSTCPEPFAKLMEDCWNPDPHSRPSFTNILDQLTTIEESGFFEMPKDSFH
CLQDNWKHEIQEMFDQLRAKEKELRTWEEELTRAALQQKNQEELLRRREQELAEREIDILERELNIII
HQLCQEKPRVKKRKGKFRKSRLKLKDGNRISLPSGFQHKFTVQASPTMDKRKSLINSRSSPPASPTII
PRLRAIQCETVSQISWGQNTQGHLSESSKTWGRSSVVPKEEGEEEEKRAPKKKGRTWGPGTLGQKELA
SGDELKSLVDGYKQWSSSAPNLVKGPRSTPALPGFTSLMEMEDEDSEGPGSGESRLQHSPSQSYLCIP
FPRGEPTPVNSATSTPQLTPTNSLKRGGAHHRRCEVALLGCGAVLAATGLGFDLLEAGKCQLLPLEEP
EPPAREEKKRREGLFQRSSRPRRSTSPPSRKLFKKEEPMLLLGDPSASLTLLSLSSISECNSTRSLLQ
SDSDEIVVYEMPVSPVEAPPLSPCTHNPLVNVRVERFKRDPNQSLTPTHVTLTTPSQPSSHRRTPSDG
ALKPETLLASRSPSPSRDPGEFPRLPDPNVVFPPTPRRWNTQQDSTLERPKTLEFLPRPRPSANRQRL
DPWWFVSPSHARSTSPANSSSTETPGPLPPTERTLLDLDAEGQSQDSTVPLCRAELNTHRPAPYEIQQ
EFWS

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 51B.

Table 51B. Comparison of the NOV51 protein sequences.

```
NOV51a  MEPSRALLGCLASAAAAAPPGEDGAGAGAEEEEEEEEEAAAAVGPGELGCDAPLPYWTAV
NOV51b  ------------------------------------------------------------

NOV51a  FEYEAAGEDELTLRLGDVVEVLSKDSQVSGDEGWWTGQLNQRVGIFPSNYVTPRSAFSSR
NOV51b  ------------------------------------------------------------

NOV51a  CQPGGEIDFAELTLEEIIGIGGFGKVYRAFWIGDEVAVKAARHDPDEDISQTIENVRQEA
NOV51b  --------TGSLTLEEIIGIGGFGKVYRAFWIGDEVAVKAARHDPDEDISQTIENVRQEA

NOV51a  KLFAMLKHPNIIALRGVCLKEPNLCLVMEFARGGPLNRVLSGKRIPPDILVNWAVQIARG
NOV51b  KLFAMLKHPNIIALRGVCLKEPNLCLVMEFARGGPLNRVLSGKRIPPDILVNWAVQIARG

NOV51a  MNYLHDEAIVPIIHRDLKSSNVLILQKVENGDLSNKILKITDFGLAREWHRTTKMSAAGT
NOV51b  MNYLHDEAIVPIIHRDLKSSNILILQKVENGDLSNKILKITDFGLAREWHRTTKMSAAGT
```

```
NOV51a  YAWMAPEVIRASMFSKGSDVWSYGVLLWELLTGEVPFRGIDGLAVAYGVAMNKLALPIPS
NOV51b  YAWMAPEVIRASMFSKGSDVWSYGVLLWELLTGEVPFRGIDGLAVAYGVAMNKLALPIPS

NOV51a  TCPEPFAKLMEDCWNPDPHSRPSFTNILDQLTTIEESGFFEMPKDSFHCLQDNWKHEIQE
NOV51b  TCPEPFAKLMEDCWNPDPHSRPSFTNILDQLTTILEG----------------------

NOV51a  MFDQLRAKEKELRTWEEELTRAALQQKNQEELLRRREQELAEREIDILERELNIIIHQLC
NOV51b  ------------------------------------------------------------

NOV51a  QEKPRVKKRKGKFRKSRLKLKDGNRISLPSGFQHKFTVQASPTMDKRKSLINSRSSPPAS
NOV51b  ------------------------------------------------------------

NOV51a  PTIIPRLRAIQCETVSQISWGQNTQGHLSESSKTWGRSSVVPKEEGEEEEKRAPKKKGRT
NOV51b  ------------------------------------------------------------

NOV51a  WGPGTLGQKELASGDELKSLVDGYKQWSSSAPNLVKGPRSTPALPGFTSLMEMEDEDSEG
NOV51b  ------------------------------------------------------------

NOV51a  PGSGESRLQHSPSQSYLCIPFPRGEPTPVNSATSTPQLTPTNSLKRGGAHHRRCEVALLG
NOV51b  ------------------------------------------------------------

NOV51a  CGAVLAATGLGFDLLEAGKCQLLPLEEPEPPAREEKKRREGLFQRSSRPRRSTSPPSRKL
NOV51b  ------------------------------------------------------------

NOV51a  FKKEEPMLLLGDPSASLTLLSLSSISECNSTRSLLQSDSDEIVVYEMPVSPVEAPPLSPC
NOV51b  ------------------------------------------------------------

NOV51a  THNPLVNVRVERFKRDPNQSLTPTHVTLTTPSQPSSHRRTPSDGALKPETLLASRSPSPS
NOV51b  ------------------------------------------------------------

NOV51a  RDPGEFPRLPDPNVVFPPTPRRWNTQQDSTLERPKTLEFLPRPRPSANRQRLDPWWFVSP
NOV51b  ------------------------------------------------------------

NOV51a  SHARSTSPANSSSTETPGPLPPTERTLLDLDAEGQSQDSTVPLCRAELNTHRPAPYEIQQ
NOV51b  ------------------------------------------------------------

NOV51a  EFWS
NOV51b  ----

NOV51a  (SEQ ID NO:  784)
NOV51b  (SEQ ID NO:  786)
```

Further analysis of the NOV51a protein yielded the following properties shown in Table 51C.

Table 51C. Protein Sequence Properties NOV51a

SignalP analysis:	Cleavage site between residues 18 and 19
PSORT II analysis:	

```
PSG:  a new signal peptide prediction method
      N-region:  length 5;  pos.chg 1;  neg.chg 1
      H-region:  length 16;  peak value   7.95
```

```
      PSG score:   3.55

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   0.45
      possible cleavage site: between 17 and 18

>>> Seems to have a cleavable signal peptide (1 to 17)

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 18
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -3.66   Transmembrane  716 - 732
      PERIPHERAL  Likelihood =  2.44 (at 342)
      ALOM score:  -3.66  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 8
      Charge difference: -10.0   C(-9.0) - N( 1.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 1a  (cytoplasmic tail 733 to 1024)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  4.60
      Hyd Moment(95):  9.12    G content:       2
      D/E content:     2       S/T content:     2
      Score: -6.34

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 15  SRA|LL

NUCDISC: discrimination of nuclear localization signals
      pat4: KKRK (5) at  487
      pat4: PKKK (4) at  594
      pat4: KKRR (5) at  756
      pat4: RPRR (4) at  768
      pat7: PRVKKRK (5) at  484
      pat7: PTMDKRK (3) at  522
      pat7: PKKKGRT (5) at  594
      bipartite: KKRREGLFQRSSRPRRS at  756
      content of basic residues:  12.1%
      NLS Score:  2.27

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: EPSR

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none
```

```
VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: too long tail

Dileucine motif in the tail: found
      LL at 734
      LL at 742
      LL at 788
      LL at 789
      LL at 799
      LL at 814
      LL at 891

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 89

COIL: Lupas's algorithm to detect coiled-coil regions
 409 L  0.51
 410 Q  0.87
 411 D  0.99
 412 N  0.99
 413 W  0.99
 414 K  1.00
 415 H  1.00
 416 E  1.00
 417 I  1.00
 418 Q  1.00
 419 E  1.00
 420 M  1.00
 421 F  1.00
 422 D  1.00
 423 Q  1.00
 424 L  1.00
 425 R  1.00
```

426	A	1.00
427	K	1.00
428	E	1.00
429	K	1.00
430	E	1.00
431	L	1.00
432	R	1.00
433	T	1.00
434	W	1.00
435	E	1.00
436	E	1.00
437	E	1.00
438	L	1.00
439	T	1.00
440	R	1.00
441	A	1.00
442	A	0.97
443	L	0.97
444	Q	0.96
445	Q	0.96
446	K	0.96
447	N	0.96
448	Q	0.96
449	E	0.96
450	E	0.96
451	L	0.81
452	L	0.81
453	R	0.81
454	R	0.81
455	R	0.71
456	E	0.71
457	Q	0.71
458	E	0.71
459	L	0.71
460	A	0.71
461	E	0.71
462	R	0.64
463	E	0.64
464	I	0.64
465	D	0.64
466	I	0.64
467	L	0.64
468	E	0.64
469	R	0.64
470	E	0.64
471	L	0.64
472	N	0.64
473	I	0.64
474	I	0.64
475	I	0.64
476	H	0.64
477	Q	0.64
478	L	0.64
479	C	0.64

```
480 Q  0.64
481 E  0.64
482 K  0.64
     total: 74 residues

--------------------------

Final Results (k = 9/23):

       44.4 %: extracellular, including cell wall
       22.2 %: Golgi
       22.2 %: endoplasmic reticulum
       11.1 %: plasma membrane

>> prediction for CG56071-01 is exc (k=9)
```

A search of the NOV51a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 51D.

5

Table 51D. Geneseq Results for NOV51a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV51a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABB98408	Human NOV7, a mixed lineage kinase 2-like protein - Homo sapiens, 1024 aa. [WO200255704-A2, 18-JUL-2002]	1..1024 1..1024	1024/1024 (100%) 1024/1024 (100%)	0.0
AAE21717	Human PKIN-12 protein - Homo sapiens, 1097 aa. [WO200218557-A2, 07-MAR-2002]	1..1024 1..1097	1003/1111 (90%) 1006/1111 (90%)	0.0
AAE11775	Human kinase (PKIN)-9 protein - Homo sapiens, 1046 aa. [WO200181555-A2, 01-NOV-2001]	1..1024 1..1046	994/1072 (92%) 997/1072 (92%)	0.0
ABP61000	Novel human protein. SEQ ID 87 - Homo sapiens, 1021 aa. [WO200250105-A1, 27-JUN-2002]	24..1024 19..1021	540/1061 (50%) 682/1061 (63%)	0.0
ABB80923	Novel human protein (NHP) kinase - Homo sapiens, 1036 aa. [WO200255685-A2, 18-JUL-2002]	24..1024 19..1036	537/1076 (49%) 684/1076 (62%)	0.0

In a BLAST search of public sequence databases, the NOV51a protein was found to have homology to the proteins shown in the BLASTP data in Table 51E.

Table 51E. Public BLASTP Results for NOV51a				
Protein Accession Number	**Protein/Organism/Length**	**NOV51a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9H2N5	Mixed lineage kinase MLK1 - Homo sapiens (Human), 1066 aa (fragment).	53..1024 1..1066	952/1080 (88%) 955/1080 (88%)	0.0
Q8K2L2	Similar to mitogen-activated protein kinase kinase kinase 9 - Mus musculus (Mouse), 732 aa (fragment).	341..1024 1..732	634/746 (84%) 649/746 (86%)	0.0
Q8BIG8	Mixed lineage kinase MLK1 homolog - Mus musculus (Mouse), 608 aa.	1..616 1..607	572/628 (91%) 581/628 (92%)	0.0
Q02779	Mitogen-activated protein kinase kinase kinase 10 (EC 2.7.1.37) (Mixed lineage kinase 2) (Protein kinase MST) - Homo sapiens (Human), 954 aa.	35..1006 2..950	579/988 (58%) 699/988 (70%)	0.0
Q8WWN1	Mixed lineage kinase 4beta - Homo sapiens (Human), 1036 aa.	24..1024 19..1036	537/1076 (49%) 680/1076 (62%)	0.0

PFam analysis indicates that the NOV51a protein contains the domains shown in the Table 51F.

Table 51F. Domain Analysis of NOV51a			
Pfam Domain	**NOV51a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
SH3	55..114	26/63 (41%) 50/63 (79%)	4.7e-15
pkinase	132..393	100/307 (33%) 216/307 (70%)	7e-95

Example 52.

The NOV52 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 52A.

5

Table 52A. NOV52 Sequence Analysis

NOV52a, CG56142-01 DNA Sequence	SEQ ID NO: 789		866 bp
	ORF Start: ATG at 19		ORF Stop: TGA at 820
CCCAGCCTTGAAGACAGA**ATG**AGAGGGGTTTCCTGTCTCCAGGTCCTGCTCCTTCTGGTGCTGGCCTG CGGGCAGCCCCGCATGTCCAGTCGGATCGTTGGGGGCCGGGATGGCCGGGACGGAGAGTGGCCGTGGC AGGCGAGCATCCAGCATCGTGGGGCACACGTGTGCGGGGGGTCGCTCATCGCCCCCCAGTGGGTGCTG ACAGCGGCGCACTGCTTCCCCAGGGCACTGCCAGCTGAGTACCGCGTGCGCCTGGGGGCGCTGCGTCT GGGCTCCACCTCGCCCCGCACGCTCTCGGTGCCCGTGCGACGGGTGCTGCTGCCCCCGGACTACTCCG AGGACGGGGCCCGCGGCGACCTGGCACTGCTGCAGCTGCGTCGCCCGGTGCCCCTGAGCGCTCGCGTC CAACCCGTCTGCCTGCCCGTGCCCGGCGCCCGCCCGCCGCCCGGCACACCATGCCGGGTCACCGGCTG GGGCAGCCTCCGCCCAGGAGTGCCCCTCCCAGAGTGGCGACCGCTACAAGGAGTAAGGGTGCCGCTGC TGGACTCGCGCACCTGCGACGGCCTCTACCACGTGGGCGCGGACGTGCCCCAGGCTGAGCGCATTGTG CTGCCTGGGAGTCTGTGTGCCGGCTACCCCCAGGGCCACAAGGACGCCTGCCAGGTGTGCACCCAGCC TCCCCAGCCTCCGGAGTCCCCTCCCTGTGCCCAGCACCCTCCCTCCCTGAACTCCAGGACCCAGGACA TCCCAACTCAGGCTCAGGATCCTGGCCTCCAACCTAGAGGCACCACGCCAGGGGTCTGGAACCCTGAG AACT**GA**AGTCCTGGGAGGGCTGGGACTTAGGCTCCTCTTTCTCCTGCAGG			
NOV52a, CG56142-01 Protein Sequence	SEQ ID NO: 790	267 aa	MW at 28699.8kD
MRGVSCLQVLLLLVLACGQPRMSSRIVGGRDGRDGEWPWQASIQHRGAHVCGGSLIAPQWVLTAAHCF PRALPAEYRVRLGALRLGSTSPRTLSVPVRRVLLPPDYSEDGARGDLALLQLRRPVPLSARVQPVCLP VPGARPPPGTPCRVTGWGSLRPGVPLPEWRPLQGVRVPLLDSRTCDGLYHVGADVPQAERIVLPGSLC AGYPQGHKDACQVCTQPPQPPESPPCAQHPPSLNSRTQDIPTQAQDPGLQPRGTTPGVWNPEN			
NOV52b, CG56142-04 DNA Sequence	SEQ ID NO: 791		638 bp
	ORF Start: ATG at 14		ORF Stop: TGA at 578
CCTTGAAGACAGA**ATG**AGAGGGGTTTCCTGTCTCCAGGTCCTGCTCCTTCTGGTGCTGGGAGCTGCTG GGACTCAGGGAAGGAAGTCTGCAGCCTGCGGGCAGCCCCGCATGTCCAGTCGGATCGTTGGGGGCCGG GATGGCCGGGACGGAGAGTGGCCGTGGCAGGCGAGCATCCAGCATCGTGGGGCACACGTGTGCGGGGG GTCGCTCATCGCCCCCCAGTGGGTGCTGACAGCGGCGCACTGCTTCCCCAGTGCCCCTCCAGAGTGGC GACCGCTACAAGGAGTAAGGGTGCCGCTGCTGGACTCGCGCACCTGCGACGGCCTCTACCACGTGGGC GCGGACGTGCCCCAGGCTGAGCGCATTGTGCTGCCTGGGAGTCTGTGTGCCGGCTACCCCCAGGGCCA CAAGGACGCCTGCCAGGTGTGCACCCAGCCTCCCCAGCCTCCGGAGTCCCCTCCCTGTGCCCAGCACC CTCCCTCCCTGAACTCCAGGACCCAGGACATCCCAACTCAGGCTCAGGATCCTGGCCTCCAACCTAGA GGCACCACGCCAGGGGTCTGGAACCCTGAGAAC**TGA**AGTCCTGGGAGGGCTGGGACTTAGGCTCCTCT TTCTCCTGGAGGGTGATTCTGGGGGA			
NOV52b, CG56142-04 Protein Sequence	SEQ ID NO: 792	188 aa	MW at 19994.5kD
MRGVSCLQVLLLLVLGAAGTQGRKSAACGQPRMSSRIVGGRDGRDGEWPWQASIQHRGAHVCGGSLIA PQWVLTAAHCFPSAPPEWRPLQGVRVPLLDSRTCDGLYHVGADVPQAERIVLPGSLCAGYPQGHKDAC QVCTQPPQPPESPPCAQHPPSLNSRTQDIPTQAQDPGLQPRGTTPGVWNPEN			
NOV52c, 276873337 DNA Sequence	SEQ ID NO: 793		1165 bp
	ORF Start: at 606		ORF Stop: TGA at 948
CACCAAGCTTTGGGACTACGAGACGCCCAAGGTGATCGTGGTGAGGAACCGGCGCCTGGGGGTCCTGT ACCGCGCCGTGCAGCTGCTCATCCTGCTCTACTTCGTGTGGTACGTATTCATCGTGCAGAAAAGCTAC CAGGAGAGCGAGACGGGCCCCGAGAGCTCCATCATCACCAAGGTCAAGGGGATCACCACGTCCGAGCA CAAAGTGTGGGACGTGGAGGAGTACGTGAAGCCCCCCGAGGGGGGCAGCGTGTTCAGCATCATCACCA			

GGGTCGAGGCCACCCACTCCCAGACCCAGGGAACCTGCCCCGAGAGCATAAGGGTCCACAACGCCACC
TGCCTCTCCGACGCCGACTGCGTGGCTGGGGAGCTGGACATGCTGGGAAACGGCCTGAGGACCGGGCG
CTGTGTGCCCTATTACCAGGGGCCCTCCAAGACCTGCGAGGTGTTCGGCTGGTGCCCGGTGGAAGATG
GGGCCTCTGTCAGCCAATTTCTGGGTACGATGGCCCCAAATTTCACCATCCTCATCAAGAACAGCATC
CACTACCCCAAATTCCACTTCTCCAAGGGCAACATCGCCGACCGCACAGACGGGTACCTGAAGCGCTG
CACGTTCCACGAGGCCTCCGACCTCTACTGCCCCATCTTCAAGCTGGGCTTTATCGTGGAGAAGGCTG
GGGAGAGCTTCACAGAGCTCGCACACAAGGGTGGTGTCATCGGGGTCATTATCAACTGGGACTGTGAC
CTGGACCTGCCTGCATCGGAGTGCAACCCCAAGTACTCCTTCCGGAGGCTTGACCCCAAGCACGTGCC
TGCCTCGTCAGGCTACAACTTCAGGTTTGCCAAATACTACAAGATCAATGGCACCACCACCCGCACGC
TCATCAAGGCCTACGGGATCCACATTGACGTCATTGTGCATGGACAGGTCGGGAAGTTCAGCCTGATT
CCCACCATTATTAATCTGGCCACAGCTCTGACTTCCGTCGGGGTGGTAAGGAACCCTCTCTGGGGTCC
CAGCGGGTGCGGGGGGTCCACCAGGCCCTTACACACCGGTCTCTGCTGGCCCCAGGGCTCCTTCCTGT
GCGACTGGATCTTGCTAACATTCATGAACAAAAACAAGGTCTACAGCCATAAGAAATTTGACAAGGTG
GTCGACGGC

NOV52c, 276873337 Protein Sequence	SEQ ID NO: 794	114 aa	MW at 11702.8kD

SAARSTRPPTSTAPSSSWALSWRRLGRASQSSHTRVVSSGSLSTGTVTWTCLHRSATPSTPSGGLTPS
TCLPRQATTSGLPNTTRSMAPPPARSSRPTGSTLTSLCMDRSGSSA

NOV52d, 276863970 DNA Sequence	SEQ ID NO: 795	1164 bp
	ORF Start: at 605	ORF Stop: TGA at 947

CACAAGCTTTGGGACTACGAGACGCCCAAGGTGATCGTGGTGAGGAACCGGCGCCTGGGGGTCCTGTA
CCGCGCCGTGCAGCTGCTCATCCTGCTCTACTTCGTGTGGTACGTATTCATCGTGCAGAAAAGCTACC
AGGAGAGCGAGACGGGCCCCGAGAGCTCCATCATCACCAAGGTCAAGGGGATCACCACGTCCGAGCAC
AAAGTGTGGGACGTGGAGGAGTACGTGAAGCCCCCCGAGGGGGGCAGCGTGTTCAGCATCATCACCAG
GGTCGAGGCCACCCACTCCCAGACCCAGGGAACCTGCCCCGAGAGCATAAGGGTCCACAACGCCACCT
GCCTCTCCGACGCCGACTGCGTGGCTGGGGAGCTGGACATGCTGGGAAACGGCCTGAGGACCGGGCGC
TGTGTGCCCTATTACCAGGGGCCCTCCAAGACCTGCGAGGTGTTCGGCTGGTGCCCGGTGGAAGATGG
GGCCTCTGTCAGCCAATTTCTGGGTACGATGGCCCCAAATTTCACCATCCTCATCAAGAACAGCATCC
ACTACCCCAAATTCCACTTCTCCAAGGGCAACATCGCCGACCGCACAGACGGGTACCTGAAGCGCTGC
ACGTTCCACGAGGCCTCCGACCTCTACTGCCCCATCTTCAAGCTGGGCTTTATCGTGGAGAAGGCTGG
GGAGAGCTTCACAGAGCTCGCACACAAGGGTGGTGTCATCGGGGTCATTATCAACTGGGACTGTGACC
TGGACCTGCCTGCATCGGAGTGCAACCCCAAGTACTCCTTCCGGAGGCTTGACCCCAAGCACGTGCCT
GCCTCGTCAGGCTACAACTTCAGGTTTGCCAAATACTACAAGATCAATGGCACCACCACCCGCACGCT
CATCAAGGCCTACGGGATCCGCATTGACGTCATTGTGCATGGACAGGTCGGGAAGTTCAGCCTGATTC
CCACCATTATTAATCTGGCCACAGCTCTGACTTCCGTCGGGGTGGTAAGGAACCCTCTCTGGGGTCCC
AGCGGGTGCGGGGGGTCCACCAGGCCCTTACACACCGGTCTCTGCTGGCCCCAGGGCTCCTTCCTGTG
CGACTGGATCTTGCTAACATTCATGAACAAAAACAAGGTCTACAGCCATAAGAAATTTGACAAGGTGG
TCGACGGC

NOV52d, 276863970 Protein Sequence	SEQ ID NO: 796	114 aa	MW at 11672.8kD

SAARSTRPPTSTAPSSSWALSWRRLGRASQSSHTRVVSSGSLSTGTVTWTCLHRSATPSTPSGGLTPS
TCLPRQATTSGLPNTTRSMAPPPARSSRPTGSALTSLCMDRSGSSA

NOV52e, 276863992 DNA Sequence	SEQ ID NO: 797	1087 bp
	ORF Start: at 606	ORF Stop: TGA at 948

CACCAAGCTTTGGGACTACGAGACGCCCAAGGTGATCGTGGTGAGGAACCGGCGCCTGGGGGTCCTGT
ACCGCGCCGTGCAGCTGCTCATCCTGCTCTACTTCGTGTGGTACGTATTCATCGTGCAGAAAAGCTAC
CAGGAGAGCGAGACGGGCCCCGAGAGCTCCATCATCACCAAGGTCAAGGGGATCACCACGTCCGAGCA
CAAAGTGTGGGACGTGGAGGAGTACGTGAAGCCCCCCGAGGGGGGCAGCGTGTTCAGCATCATCACCA
GGGTCGAGGCCACCCACTCCCAGACCCAGGGAACCTGCCCCGAGAGCATAAGGGTCCACAACGCCACC
TGCCTCTCCGACGCCGACTGCGTGGCTGGGGAGCTGGACATGCTGGGAAACGGCCTGAGGACCGGGCG
CTGTGTGCCCTATTACCAGGGGCCCTCCAAGACCTGCGAGGTGTTCGGCTGGTGCCCGGTGGAAGATG
GGGCCTCTGTCAGCCAATTTCTGGGTACGATGGCCCCAAATTTCACCATCCTCATCAAGAACAGCATC
CACTACCCCAAATTCCACTTCTCCAAGGGCAACATCGCCGACCGCACAGACGGGTACCTGAAGCGCTG
CACGTTCCACGAGGCCTCCGACCTCTACTGCCCCATCTTCAAGCTGGGCTTTATCGTGGAGAAGGCTG

GGGAGAGCTTCACAGAGCTCGCACACAAGGGTGGTGTCATCGGGGTCATTATCAACTGGGACTGTGAC
CTGGACCTGCCTGCATCGGAGTGCAACCCCAAGTACTCCTTCCGGAGGCTTGACCCCAAGCACGTGCC
TGCCTCGTCAGGCTACAACTTCAGGTTTGCCAAATACTACAAGATCAATGGCACCACCACCCGCACGC
TCATCAAGGCCTACGGGATCCGCATTGACGTCATTGTGCATGGACAGGCCGGGAAGTTCAGCCTGATT
CCCACCATTATTAATCTGGCCACAGCTCTGACTTCCGTCGGGGTGGGCTCCTTCCTGTGCGACTGGAT
CTTGCTAACATTCATGAACAAAAACAAGGTCTACAGCCATAAGAAATTTGACAAGGTGGTCGACGGC

NOV52e, 276863992 Protein Sequence	SEQ ID NO: 798	114 aa	MW at 11682.8kD

SAARSTRPPTSTAPSSSWALSWRRLGRASQSSHTRVVSSGSLSTGTVTWTCLHRSATPSTPSGGLTPS
TCLPRQATTSGLPNTTRSMAPPPARSSRPTGSALTSLCMDRPGSSA

NOV52f, 276873330 DNA Sequence	SEQ ID NO: 799	1087 bp
	ORF Start: at 606	ORF Stop: TGA at 948

CACCAAGCTTTGGGACTACGAGACGCCCAAGGTGATCGTGGTGAGGAACCGGCGCCTGGGGGTCCTGT
ACCGCGCCGTGCAGCTGCTCATCCTGCTCTACTTCGTGTGGTACGTATTCATCGTGCAGAAAAGCTAC
CAGGAGAGCGAGACGGGCCCCGAGAGCTCCATCATCACCAAGGTCAAGGGGATCACCACGTCCGAGCA
CAAAGTGTGGGACGTGGAGGAGTACGTGAAGCCCCCCGAGGGGGGCAGCGTGTTCAGCATCATCACCA
GGGTCGAGGCCACCCACTCCCAGACCCAGGGAACCTGCCCCGAGAGCATAAGGGTCCACAACGCCACC
TGCCTCTCCGACGCCGACTGCGTGGCTGGGGAGCTGGACATGCTGGGAAACGGCCTGAGGACCGGGCG
CTGTGTGCCCTATTACCAGGGGCCCTCCAAGACCTGCGAGGTGTTCGGCTGGTGCCCGGTGGAAGATG
GGGCCTCTGTCAGCCAATTTCTGGGTACGATGGCCCCAAATTTCACCATCCTCATCAAGAACAGCATC
CACTACCCCAAATTCCACTTCTCCAAGGGCAACATCGCCGACCGCACAGACGGGTACCTGAAGCGCTG
CACGTTCCACGAGGCCTCCGACCTCTACTGCCCCATCTTCAAGCTGGGCTTTATCGTGGAGAAGGCTG
GGGAGAGCTTCACAGAGCTCGCACACAAGGGTGGTGTCATCGGGGTCATTATCAACTGGGACTGTGAC
CTGGACCTGCCTGCATCGGAGTGCAACCCCAAGTACTCCTTCCGGAGGCTTGACCCCAAGCACGTGCC
TGCCTCGTCAGGCTACAACTTCAGGTTTGCCAAATACTACAAGATCAATGGCACCACCACCCGCACGC
TCATCAAGGCCTACGGGATCCGCATTGACGTCATTGTGCATGGACAGGCCGGGAAGTTCATCCTGATT
CCCACCATTATTAATCTGGCCACAGCTCTGACTTCCGTCGGGGTGGGCTCCTTCCTGTGCGACTGGAT
CTTGCTAACATTCATGAACAAAAACAAGGTCTACAGCCATAAGAAATTTGACAAGGTGGTCGACGGC

NOV52f, 276873330 Protein Sequence	SEQ ID NO: 800	114 aa	MW at 11698.8kD

SAARSTRPPTSTAPSSSWALSWRRLGRASQSSHTRVVSSGSLSTGTVTWTCLHRSATPSTPSGGLTPS
TCLPRQATTSGLPNTTRSMAPPPARSSRPTGSALTSLCMDRPGSSS

NOV52g, CG56142-02 DNA Sequence	SEQ ID NO: 801	1020 bp
	ORF Start: ATG at 91	ORF Stop: TAA at 931

AGGACTCTCCTCTCTTCTCCCTGCTGGCTCCAGACCAGAGTCCAAGCCCTAGGCAGTGCCACCCTTAC
CCAGCCCAGCCTTGAAGACAGAATGAGAGGGGTTTCCTGTCTCCAGGTCCTGCTCCTTCTGGTGCTGG
GAGCTGCTGGGACTCAGGGAAGGAAGTCTGCAGCCTGCGGGCAGCCCCGCATGTCCAGTCGGATCGTT
GGGGGCCGGGATGGCCGGGACGGAGAGTGGCCGTGGCAGGCGAGCATCCAGCATCGTGGGGCACACGT
GTGCGGGGGGTCGCTCATCGCCCCCCAGTGGGTGCTGACAGCGGCGCACTGCTTCCCCAGGAGGGCAC
TGCCAGCTGAGTACCGCGTGCGCCTGGGGGCGCTGCGTCTGGGCTCCACCTCGCCCCGCACGCTCTCG
GTGCCCGTGCGACGGGTGCTGCTGCCCCCGGACTACTCCGAGGACGGGGCCCGCGGCGACCTGGCACT
GCTGCAGCTGCGTCGCCCGGTGCCCCTGAGCGCTCGCGTCCAACCCGTCTGCCTGCCCGTGCCCGGCG
CCCGCCCGCCGCCCGGCACACCATGCCGGGTCACCGGCTGGGGCAGCCTCCGCCCAGGAGTGCCCCTC
CCAGAGTGGCGACCGCTACAAGGAGTAAGGGTGCCGCTGCTGGACTCGCGCACCTGCGACGGCCTCTA
CCACGTGGGCGCGGACGTGCCCCAGGCTGAGCGCATTGTGCTGCCTGGGAGTCTGTGTGCCGGCTACC
CCCAGGGCCACAAGGACGCCTGCCAGGGTGATTCTGGGGGACCTCTGACCTGCCTGCAGTCTGGGAGC
TGGGTCCTGGTGGGCGTGGTGAGCTGGGGCAAGGGTTGTGCCCTGCCCAACCGTCCAGGGGTCTACAC
CAGTGTGGCCACATATAGCCCCTGGATTCAGGCTCGCGTCAGCTTCTAATGCTAGCCGGTGAGGCTGA
CCTGGAGCCAGCTGCTGGGGTCCCTCAGCCTCCTGGTTCATCCAGGCACCTGCCTATACCCCACATCC

NOV52g, CG56142-02 Protein Sequence	SEQ ID NO: 802	280 aa	MW at 29786.2kD

MRGVSCLQVLLLLVLGAAGTQGRKSAACGQPRMSSRIVGGRDGRDGEWPWQASIQHRGAHVCGGSLIA
PQWVLTAAHCFPRRALPAEYRVRLGALRLGSTSPRTLSVPVRRVLLPPDYSEDGARGDLALLQLRRPV

PLSARVQPVCLPVPGARPPPGTPCRVTGWGSLRPGVPLPEWRPLQGVRVPLLDSRTCDGLYHVGADVP
QAERIVLPGSLCAGYPQGHKDACQGDSGGPLTCLQSGSWVLVGVVSWGKGCALPNRPGVYTSVATYSP
WIQARVSF

NOV52h, CG56142-03 DNA Sequence	SEQ ID NO: 803	1606 bp
	ORF Start: ATG at 69	ORF Stop: TGA at 921

CCACGCGTCCGACCAGAGTCCAAGCCCTAGGCAGTGCCACCCTTACCCAGCCCAGCCTTGAAGACAGA
ATGAGAGGGGTTTCCTGTCTCCAGGTCCTGCTCCTTCTGGTGCTGGGAGCTGCTGGGACTCAGGGAAG
GAAGTCTGCAGCCTGCGGGCAGCCCCGCATGTCCAGTCGGATCGTTGGGGGCCGGGATGGCCGGGACG
GAGAGTGGCCGTGGCAGGCGAGCATCCAGCATCCTGGGGCACACGTGTGCGGGGGGTCGCTCATCGCC
CCCCAGTGGGTGCTGACAGCGGCGCACTGCTTCCCCAGGAGGGCACTGCCAGCTGAGTACCGCGTGCG
CCTGGGGGCGCTGCGTCTGGGCTCCACCTCGCCCCGCACGCTCTCGGTGCCCGTGCGACGGGTGCTGC
TGCCCCCGGACTACTCCGAGGACGGGGCCCGCGGCGACCTGGCACTGCTGCAGCTGCGTCGCCCGGTG
CCCCTGAGCGCTCGCGTCCAACCCGTCTGCCTGCCCGTGCCCGGCGCCCGCCCGCCGCCCGGCACACC
ATGCCGGGTCACCGGCTGGGGCAGCCTCCGCCCAGGAGTGCCCCTCCCAGAGTGGCGACCGCTACAAG
GAGTAAGGGTGCCGCTGCTGGACTCGCGCACCTGCGACGGCCTCTACCACGTGGGCGCGGACGTGCCC
CAGGCTGAGCGCATTGTGCTGCCTGGGAGTCTGTGTGCCGGCTACCCCCAGGGCCACAAGGACGCCTG
CCAGGGTGATTCTGGGGGACCTCTGACCTGCCTGCAGTCTGGGAGCTGGGTCCTGGTGGGCGTGGTGA
GCTGGGGCAAGGGTTGTGCCCTGCCCAACCGTCCAGGGGTCTACACCAGTGTGGCCACATATAGCCCC
TGGATTCAGGCTCGCGTCACTTCTAATGCTAGCCGGTGAGGCTGACCTGGAGCCAGCTGCTGGGGTCC
CTCAGCCTCCTGGTTCATCCAGGCACCTGCCTATACCCCACATCCCTTCTGCCTCGAGGCCAAGATGC
CTAAAAAAGCTAAAGGCCACCCCACCCCCCACCCACCACCTTCTGGCTCCTCTCCTCTTTGGGGATCA
CCAGCTCTGACTCCACCAACCCTCATCCAGGAATCTGCCATGAGTCCCAGGGAGTCACACTCCCCACT
CCCTTCCTGGCTTGTATTTACTTTTCTTGGCCCTGGCCAGGGCTGGGCGCAAGGCACGCAGTGATGGG
CAAACCAATTGCTGCCCATCTGGCCTGTGTGCCCATCTTTTTCTGGAGAAAGTCAGATTCACAGCATG
ACAGAGATTTGACACCAGGGAGATCCTCCATAGCTGGCTTTGAGGACACGGGGACCACAGCCATGAGC
GGCCTCTAAGAGCTGAGAGACAGCCGGCAGGGAATCGGAACCCTCAGACCCACAGCCGCAAGGCACTG
GATTCTGGCAGCACCCTGAAGGAGCTGGGAAGTAAGTTCTTCCCCAGCCTCCAGATAAGAGCCCCGCC
GGCCAATCCCTTCATTTCAACCTAAAGAGACCCTAAGCAGAGAACCTAGCTGAGCCACTCCTGACCTA
CAAAGTTGTGACTTAATAAATGTGTGCTTTAAGCTGCCAAAA

NOV52h, CG56142-03 Protein Sequence	SEQ ID NO: 804	284 aa	MW at 30109.5kD

MRGVSCLQVLLLLVLGAAGTQGRKSAACGQPRMSSRIVGGRDGRDGEWPWQASIQHPGAHVCGGSLIA
PQWVLTAAHCFPRRALPAEYRVRLGALRLGSTSPRTLSVPVRRVLLPPDYSEDGARGDLALLQLRRPV
PLSARVQPVCLPVPGARPPPGTPCRVTGWGSLRPGVPLPEWRPLQGVRVPLLDSRTCDGLYHVGADVP
QAERIVLPGSLCAGYPQGHKDACQGDSGGPLTCLQSGSWVLVGVVSWGKGCALPNRPGVYTSVATYSP
WIQARVTSNASR

NOV52i, CG56142-05 DNA Sequence	SEQ ID NO: 805	762 bp
	ORF Start: at 7	ORF Stop: at 757

AGATCTCAGCCCCGCATGTCCAGTCGGATCGTTGGGGGCCGGGATGGCCGGGACGGAGAGTGGCCGTG
GCAGGCGAGCATCCAGCATCGTGGGGCACACGTGTGCGGGGGGTCGCTCATCGCCCCCCAGTGGGTGC
TGACAGCGGCGCACTGCTTCCCCAGGAGGGCACTGCCAGCTGAGTACCGCGTGCGCCTGGGGGCGCTG
CGTCTGGGCTCCACCTCGCCCCGCACGCTCTCGGTGCCCGTGCGACGGGTGCTGCTGCCCCCGGACTA
CTCCGAGGACGGGGCCCGCGGCGACCTGGCACTGCTGCAGCTGCGTCGCCCGGTGCCCCTGAGCGCTC
GCGTCCAACCCGTCTGCCTGCCCGTGCCCGCCGCCCGCCCGCCGCCCGGCACACCATGCCGGGTCACC
GGCTGGGGCAGCCTCCGCCCAGGAGTGCCCCTCCCAGAGGGGCGACCGCTACAAGGAGTAAGGGTGCC
GCTGCTGGACTCGCGCACCTGCGACGGCCTCTACCACGTGGGCGCGGACGTGCCCCAGGCTGAGCGCA
TTGTGCTGCCTGGGAGTCTGTGTGCCGGCTACCCCCAGGTCCACAAGGACGCCTGCCAGGTGTGCACC
CAGCCTCCCCAGCCTCCGGAGTCCCCTCCCTGTGCCCAGCTCCCTCCCTCCCTGAACTCCAGGACCCA
GGACATCCCAACTCAGGCTCAGGATCCTGGCCTCCAACCTAGAGGCACCACGCCAGGGGTCTGGAACC
CTGAGAACCTCGAG

NOV52i, CG56142-05 Protein Sequence	SEQ ID NO: 806	250 aa	MW at 26888.5kD

QPRMSSRIVGGRDGRDGEWPWQASIQHRGAHVCGGSLIAPQWVLTAAHCFPRRALPAEYRVRLGALRL
GSTSPRTLSVPVRRVLLPPDYSEDGARGDLALLQLRRPVPLSARVQPVCLPVPAARPPPGTPCRVTGW

GSLRPGVPLPEGRPLQGVRVPLLDSRTCDGLYHVGADVPQAERIVLPGSLCAGYPQVHKDACQVCTQP PQPPESPPCAQLPPSLNSRTQDIPTQAQDPGLQPRGTTPGVWNPEN		
NOV52j, CG56142-06 DNA Sequence	SEQ ID NO: 807	762 bp
	ORF Start: at 7	ORF Stop: at 757
AGATCTCAGCCCCGCATGTCCAGTCGGATCGTTGGGGGCCGGGATGGCCGGGACGGAGAGTGGCCGTG GCAGGCGAGCATCCAGCATCGTGGGGCACACGTGTGCGGGGGGTCGCTCATCGCCCCCCAGTGGGTGC TGACAGCGGCGCACTGCTTCCCCAGGAGGGCACTGCCAGCTGAGTACCGCGTGCGCCTGGGGGCGCTG CGTCTGGGCTCCACCTCGCCCCGCACGCTCTCGGTGCCCGTGCGACGGGTGCTGCTGCCCCCGGACTA CTCCGAGGACGGGGCCCGCGGCGACCTGGCACTGCTGCAGCTGCGTCGCCCGGTGCCCCTGAGCGCTC GCGTCCAACCCGTCTGCCTGCCCGTGCCCGGCGCCCGCCCGCCGCCCGGCACACCATGCCGGGTCACC GGCTGGGGCAGCCTCCGCCCAGGAGTGCCCCTCCCAGAGTGGCGACCGCTACAAGGAGTAAGGGTGCC GCTGCTGGACTCGCGCACCTGCGACGGCCTCTACCACGTGGGCGCGGACGTGCCCCAGGCTGAGCGCA TTGTGCTGCCTGGGAGTCTGTGTGCCGGCTACCCCCAGGGCCACAAGGACGCCTGCCAGGTGTGCACC CAGCCTCCCCAGCCTCCGGAGTCCCCTCCCTGTGCCCAGCTCCCTCCCTCCCTGAACTCCAGGACCCA GGACATCCCAACTCAGGCTCAGGATCCTGGCCTCCAACCTAGAGGCACCACGCCAGGGGTCTGGAACC CTGAGAACCTCGAG		
NOV52j, CG56142-06 Protein Sequence	SEQ ID NO: 808	250 aa	MW at 26961.6kD
QPRMSSRIVGGRDGRDGEWPWQASIQHRGAHVCGGSLIAPQWVLTAAHCFPRRALPAEYRVRLGALRL GSTSPRTLSVPVRRVLLPPDYSEDGARGDLALLQLRRPVPLSARVQPVCLPVPGARPPPGTPCRVTGW GSLRPGVPLPEWRPLQGVRVPLLDSRTCDGLYHVGADVPQAERIVLPGSLCAGYPQGHKDACQVCTQP PQPPESPPCAQLPPSLNSRTQDIPTQAQDPGLQPRGTTPGVWNPEN			

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 52B.

Table 52B. Comparison of the NOV52 protein sequences.

```
NOV52a  MRGVSCLQVLLLLVL----------ACGQPRMSSRIVGGRDGRDGEWPWQASIQHRGAH
NOV52b  MRGVSCLQVLLLLVLGAAGTQGRKSAACGQPRMSSRIVGGRDGRDGEWPWQASIQHRGAH
NOV52c  ------------------------------------------------------------
NOV52d  ------------------------------------------------------------
NOV52e  ------------------------------------------------------------
NOV52f  ------------------------------------------------------------
NOV52g  MRGVSCLQVLLLLVLGAAGTQGRKSAACGQPRMSSRIVGGRDGRDGEWPWQASIQHRGAH
NOV52h  MRGVSCLQVLLLLVLGAAGTQGRKSAACGQPRMSSRIVGGRDGRDGEWPWQASIQHPGAH
NOV52i  ----------------------------QPRMSSRIVGGRDGRDGEWPWQASIQHRGAH
NOV52j  ----------------------------QPRMSSRIVGGRDGRDGEWPWQASIQHRGAH

NOV52a  VCGGSLIAPQWVLTAAHCFPR-ALPAEYRVRLGALRLGSTSPRTLSVPVRRVLLPPDYSE
NOV52b  VCGGSLIAPQWVLTAAHCFPS------------------------------------------
NOV52c  ------------------------------------------------------------
NOV52d  ------------------------------------------------------------
NOV52e  ------------------------------------------------------------
NOV52f  ------------------------------------------------------------
NOV52g  VCGGSLIAPQWVLTAAHCFPRRALPAEYRVRLGALRLGSTSPRTLSVPVRRVLLPPDYSE
NOV52h  VCGGSLIAPQWVLTAAHCFPRRALPAEYRVRLGALRLGSTSPRTLSVPVRRVLLPPDYSE
NOV52i  VCGGSLIAPQWVLTAAHCFPRRALPAEYRVRLGALRLGSTSPRTLSVPVRRVLLPPDYSE
NOV52j  VCGGSLIAPQWVLTAAHCFPRRALPAEYRVRLGALRLGSTSPRTLSVPVRRVLLPPDYSE

NOV52a  DGARGDLALLQLRRPVPLSARVQPVCLPVPGARPPPGTPCRVTGWGSLRPGVPLPEWRPL
NOV52b  ----------------------------------------AP------------PEWRPL
NOV52c  --------------------------SAARSTRPPTST-APSSSWA--------LSWRRL
```

```
NOV52d  ------------------------SAARSTRPPTST-APSSSWA--------LSWRRL
NOV52e  ------------------------SAARSTRPPTST-APSSSWA--------LSWRRL
NOV52f  ------------------------SAARSTRPPTST-APSSSWA--------LSWRRL
NOV52g  DGARGDLALLQLRRPVPLSARVQPVCLPVPGARPPPGTPCRVTGWGSLRPGVPLPEWRPL
NOV52h  DGARGDLALLQLRRPVPLSARVQPVCLPVPGARPPPGTPCRVTGWGSLRPGVPLPEWRPL
NOV52i  DGARGDLALLQLRRPVPLSARVQPVCLPVPAARPPPGTPCRVTGWGSLRPGVPLPEGRPL
NOV52j  DGARGDLALLQLRRPVPLSARVQPVCLPVPGARPPPGTPCRVTGWGSLRPGVPLPEWRPL

NOV52a  QGVRVPLLDSRTCDGLYHVG-ADVPQAERIVLPGSLCAGYPQGHKDACQVCTQ-PPQPPE
NOV52b  QGVRVPLLDSRTCDGLYHVG-ADVPQAERIVLPGSLCAGYPQGHKDACQVCTQ-PPQPPE
NOV52c  GRASQSSHTRVVSSGSLSTGTVTWTCLHRSATPSTPSGGLTPSTCLPRQATTSGLPNTTR
NOV52d  GRASQSSHTRVVSSGSLSTGTVTWTCLHRSATPSTPSGGLTPSTCLPRQATTSGLPNTTR
NOV52e  GRASQSSHTRVVSSGSLSTGTVTWTCLHRSATPSTPSGGLTPSTCLPRQATTSGLPNTTR
NOV52f  GRASQSSHTRVVSSGSLSTGTVTWTCLHRSATPSTPSGGLTPSTCLPRQATTSGLPNTTR
NOV52g  QGVRVPLLDSRTCDGLYHVG-ADVPQAERIVLPGSLCAGYPQGHKDACQGDSGGPLTCLQ
NOV52h  QGVRVPLLDSRTCDGLYHVG-ADVPQAERIVLPGSLCAGYPQGHKDACQGDSGGPLTCLQ
NOV52i  QGVRVPLLDSRTCDGLYHVG-ADVPQAERIVLPGSLCAGYPQVHKDACQVCTQ-PPQPPE
NOV52j  QGVRVPLLDSRTCDGLYHVG-ADVPQAERIVLPGSLCAGYPQGHKDACQVCTQ-PPQPPE

NOV52a  SPPCAQHPPSLNSRTQDIPTQAQDPGLQPRGTTPGVWNPEN-------
NOV52b  SPPCAQHPPSLNSRTQDIPTQAQDPGLQPRGTTPGVWNPEN-------
NOV52c  SM--APP-PARSSRPTGSTLTSL-CMDRSGSSA---------------
NOV52d  SM--APP-PARSSRPTGSALTSL-CMDRSGSSA---------------
NOV52e  SM--APP-PARSSRPTGSALTSL-CMDRPGSSA---------------
NOV52f  SM--APP-PARSSRPTGSALTSL-CMDRPGSSS---------------
NOV52g  SG--SWVLVGVVSWGKGCALPNR-PGVYTSVATYSPWIQARVSF----
NOV52h  SG--SWVLVGVVSWGKGCALPNR-PGVYTSVATYSPWIQARVTSNASR
NOV52i  SPPCAQLPPSLNSRTQDIPTQAQDPGLQPRGTTPGVWNPEN-------
NOV52j  SPPCAQLPPSLNSRTQDIPTQAQDPGLQPRGTTPGVWNPEN-------

NOV52a  (SEQ ID NO:  790)
NOV52b  (SEQ ID NO:  792)
NOV52c  (SEQ ID NO:  794)
NOV52d  (SEQ ID NO:  796)
NOV52e  (SEQ ID NO:  798)
NOV52f  (SEQ ID NO:  800)
NOV52g  (SEQ ID NO:  802)
NOV52h  (SEQ ID NO:  804)
NOV52i  (SEQ ID NO:  806)
NOV52j  (SEQ ID NO:  808)
```

Further analysis of the NOV52a protein yielded the following properties shown in Table 52C.

Table 52C. Protein Sequence Properties NOV52a

SignalP analysis:	Cleavage site between residues 19 and 20
PSORT II analysis:	

```
PSG:  a new signal peptide prediction method
      N-region:  length 2;  pos.chg 1;  neg.chg 0
      H-region:  length 18;  peak value  10.88
      PSG score:   6.48
```

```
GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   1.06
      possible cleavage site: between 18 and 19

>>> Seems to have a cleavable signal peptide (1 to 18)

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 19
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  5.30 (at 48)
      ALOM score:   5.30  (number of TMSs: 0)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 9
      Charge difference:  0.0   C( 2.0) - N( 2.0)
      N >= C: N-terminal side will be inside

MITDISC: discrimination of mitochondrial targeting seq
      R content:        4       Hyd Moment(75):  6.32
      Hyd Moment(95):  5.75     G content:       4
      D/E content:      1       S/T content:     3
      Score: -2.36

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 35  SRI|VG

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  10.1%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: RGVS

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none
```

```
NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 76.7

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: extracellular, including cell wall
       33.3 %: mitochondrial
       11.1 %: nuclear

>> prediction for CG56142-01 is exc (k=9)
```

A search of the NOV52a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 52D.

Table 52D. Geneseq Results for NOV52a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV52a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU74742	Human protease PRTS-2 protein sequence - Homo sapiens, 365 aa. [WO200198468-A2, 27-DEC-2001]	1..267 87..365	267/279 (95%) 267/279 (95%)	e-159
ABP61012	Novel human protein. SEQ ID 99 - Homo sapiens, 279 aa. [WO200250105-A1, 27-JUN-2002]	1..267 1..279	267/279 (95%) 267/279 (95%)	e-159
AAE00290	Human plasminogen-like protein from clone HEOQB08 - Homo sapiens, 279 aa. [WO200124815-A1, 12-APR-2001]	1..267 1..279	267/279 (95%) 267/279 (95%)	e-159
ABB07286	Human prostasin-like serine protease - Homo sapiens, 272 aa. [WO200198466-A2, 27-DEC-2001]	1..216 1..217	216/217 (99%) 216/217 (99%)	e-127
ABP61010	Novel human protein. SEQ ID 97 - Homo sapiens, 280 aa. [WO200250105-A1, 27-JUN-2002]	1..216 1..228	216/228 (94%) 216/228 (94%)	e-125

In a BLAST search of public sequence databases, the NOV52a protein was found to have homology to the proteins shown in the BLASTP data in Table 52E.

Table 52E. Public BLASTP Results for NOV52a				
Protein Accession Number	**Protein/Organism/Length**	**NOV52a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8N171	Similar to protease, serine, 8 (Prostasin) - Homo sapiens (Human), 327 aa (fragment).	1..216 48..275	216/228 (94%) 216/228 (94%)	e-124
Q8NF86	Serine protease EOS - Homo sapiens (Human), 284 aa.	1..216 1..228	215/228 (94%) 215/228 (94%)	e-123
AAP20885	Tryptase-6 - Mus musculus (Mouse), 277 aa.	1..216 1..225	175/225 (77%) 186/225 (81%)	3e-98
Q9ER10	Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4) - Mus musculus (Mouse), 306 aa.	17..215 41..238	91/202 (45%) 125/202 (61%)	1e-46
Q9ES87	Prostasin precursor (EC 3.4.21.-) - Rattus norvegicus (Rat), 342 aa.	16..216 36..235	93/202 (46%) 125/202 (61%)	1e-46

PFam analysis indicates that the NOV52a protein contains the domains shown in the Table 52F.

Table 52F. Domain Analysis of NOV52a			
Pfam Domain	**NOV52a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
trypsin	26..241	83/269 (31%) 165/269 (61%)	1e-38

Example 53.

The NOV53 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 53A.

Table 53A. NOV53 Sequence Analysis		
NOV53a, CG56144-01 DNA Sequence	SEQ ID NO: 809 ORF Start: ATG at 90	1151 bp ORF Stop: TAG at 1056

TAATCTTTGCAGGTGGGATAGCACAGGTTGAACTCTAATCATATATACTGTAGAAGGTATATATAGAA
GGTGAAGAAGCCCTGTAAAAA**ATG**ACAAGGAGATTTCCAGGAGCCATGCTTCCCTCTAATATCACCTC
AACACATCCAGCTGTCTTTTTGTTGGTAGGAATTCCTGGTTTGGAACACCTGCATGCCTGGATCTCCA
TCCCCTTCTGCTTTGCTTATACTCTGGCCCTGCTAGGCAACTGTACCCTTCTCTTCATTATCCAGGCT
GATGCAGCCCTCCATGAACCCATGTACCTCTTTCTGGCCATGTTGGCAACCATTGACTTGGTTCTTTC

TTCTACAACGCTGCCCAAAATGCTTGCCATATTCTGGTTCAGGGATCAGGAGATCAACTTCTTTGCCT
GTCTGGTCCAGATGTTCTTCCTTCACTCCTTCTCCATCATGGAGTCAGCAGTGCTGCTGGCCATGGCC
TTTGACCGCTATGTGGCCATCTGCAAGCCATTGCACTACACGACGGTCCTGACTGGGTCCCTCATCAC
CAAGATTGGCATGGCTGCTGTGGCCCGGGCTGTGACACTAATGACTCCACTCCCCTTCCTGCTCAGAC
GCTTCCACTACTGCCGAGGCCCAGTGATTGCCCATTGCTACTGTGAACACATGGCTGTGGTAAGGCTG
GCGTGTGGGGACACTAGCTTCAACAATATCTATGGCATTGCTGTGGCCATGTTTAGTGTGGTGTTGGA
CCTGCTCTTTGTTATCCTGTCTTATGTCTTCATCCTTCAGGCAGTTCTCCAGCTTGCCTCTCAGGAGG
CCCGCTACAAAGCATTTGGGACATGTGTGTCTCACATAGGTGCCATCCTGTCCACCTACACTCCAGTA
GTCATCTCTTCAGTCATGCACCGTGTAGCCCGCCATGCTGCCCCTCGTGTCCACATACTCCTTGCTAT
TTTCTATCTCCTTTTCCCACCCATGGTCAATCCTATCATATATGGAGTCAAGACCAAGCAGATTCGTG
AGTATGTGCTCAGTCTATTCCAGAGAAAGAACATGTAGATGGATAGTTCTCTTTTTTTATCCCACTTG
CCAAGTAATGAGAATGCTGGATTGGGGTTGAGGGGAAAAATCTAAATAGGAAAATTGCAGAGT

NOV53a, CG56144-01 Protein Sequence	SEQ ID NO: 810	322 aa	MW at 36121.9kD

MTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLFIIQADAALHEP
MYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSFSIMESAVLLAMAFDRYVAI
CKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFHYCRGPVIAHCYCEHMAVVRLACGDTSF
NNIYGIAVAMFSVVLDLLFVILSYVFILQAVLQLASQEARYKAFGTCVSHIGAILSTYTPVVISSVMH
RVARHAAPRVHILLAIFYLLFPPMVNPIIYGVKTKQIREYVLSLFQRKNM

NOV53b, 170645965 DNA Sequence	SEQ ID NO: 811	981 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCACCATGACAAGGAGATTTCCAGGAGCCATGCTTCCCTCTAATATCACCTCAACACATCCAGC
TGTCTTTTTGTTGGTAGGAATTCCTGGTTTGGAACACCTGCATGCCTGGATCTCCATCCCCTTCTGCT
TTGCTTATACTCTGGCCCTGCTAGGCAACTGTACCCTTCTCTTCATTATCCAGGCTGATGCAGCCCTC
CATGAACCCATGTACCTCTTTCTGGCCATGTTGGCAACCATTGACTTGGTTCTTTCTTCTACAACGCT
GCCCAAAATGCTTGCCATATTCTGGTTCAGGGATCAGGAGATCAACTTCTTTGCCTGTCTGGTCCAGA
TGTTCTTCCTTCACTCCTTCTCCATCATGGAGTCAGCAGTGCTGCTGGCCATGGCCTTTGACCGCTAT
GTGGCCATCTGCAAGCCATTGCACTACACGACGGTCCTGACTGGGTCCCTCATCACCAAGATTGGCAT
GGCTGCTGTGGCCCGGGCTGTGACACTAATGACTCCACTCCCCTTCCTGCTCAGACGCTTCCACTACT
GCCGAGGCCCAGTGATTGCCCATTGCTACTGTGAACACATGGCTGTGGTAAGGCTGGCGTGTGGGGAC
ACTAGCTTCAACAATATCTATGGCATTGCTGTGGCCATGTTTATTGTGGTGTTGGACCTGCTCTTTGT
TATCCTGTCTTATGTCTTCATCCTTCAGGCAGTTCTCCAGCTTGCCTCTCAGGAGGCCCGCTACAAGG
CATTTGGGACATGTGTGTCTCACATAGGTGCCATCCTGTCCACCTACACTCCAGTAGTCATCTCTTCA
GTCATGCACCGTGTAGCCCGCCATGCTGCCCCTCGTGTCCACATACTCCTTGCTATTTTCTATCTCCT
TTTCCCACCCATGGTCAATCCTATCATATATGGAGTCAAGACCAAGCAGATTCGTGAGTATGTGCTCA
GTCTATTCCAGAGAAAGAACATGCTCGAG

NOV53b, 170645965 Protein Sequence	SEQ ID NO: 812	327 aa	MW at 36635.5kD

GSTMTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLFIIQADAAL
HEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSFSIMESAVLLAMAFDRY
VAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFHYCRGPVIAHCYCEHMAVVRLACGD
TSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQLASQEARYKAFGTCVSHIGAILSTYTPVVISS
VMHRVARHAAPRVHILLAIFYLLFPPMVNPIIYGVKTKQIREYVLSLFQRKNMLE

NOV53c, 168869277 DNA Sequence	SEQ ID NO: 813	981 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCACCATGACAAGGAGATTTCCAGGAGCCATGCTTCCCTCTAATATCACCTCAACACATCCAGC
TGTCTTTTTGTTGGTAGGAATTCCTGGTTTGGAACACCTGCATGCCTGGATCTCCATCCCCTTCTGCT
TTGCTTATACTCTGGCCCTGCTAGGCAACTGTACCCTTCTCTTCATTATCCAGGCTGATGCAGCCCTC
CATGAACCCATGTACCTCTTTCTGGCCATGTTGGCAACCATTGACTTGGTTCTTTCTTCTACAACGCT
GCCCAAAATGCTTGCCATATTCTGGTTCAGGGATCAGGAGATCAACTTCTTTGCCTGTCTGGTCCAGA
TGTTCTTCCTTCACTCCTTCTCCATCATGGAGTCAGCAGTGCTGCTGGCCATGGCCTTTGACCGCTAT
GTGGCCATCTGCAAGCCATTGCACTACACGACGGTCCTGACTGGGTCCCTCATCACCAAGATTGGCAT
GGCTGCTGTGGCCCGGGCTGTGACACTAATGACTCCACTCCCCTTCCTGCTCAGACGCTTCCACTACT
GCCGAGGCCCAGTGATTGCCCATTGCTACTGTGAACACATGGCTGTGGTAAGGCTGGCGTGTGGGGAC

ACTAGCTTCAACAATATCTATGGCATTGCTGTGGCCATGTTTATTGTGGTGTTGGACCTGCTCTTTGT
TATCCTGTCTTATGTCTTCATCCTTCAGGCAGTTCTCCAGCTTGCCTCTCAGGAGGCCCGCTACAAGG
CATTTGGGACATGTGCGTCTCACATAGGTGCCATCCTGTCCACCTACACTCCAGTAGTCATCTCTTCA
GTCATGCACCGTGTAGCCCGCCATGCTGCCCCTCGTGTCCACATACTCCTTGCTATTTTCTATCTCCT
TTTCCCACCCATGGTCAATCCTATCATATATGGAGTCAAGACCAAGCAGATTCGTGAGTATGTGCTCA
GTCTATTCCAGAGAAAGAACATGCTCGAG

NOV53c, 168869277 Protein Sequence	SEQ ID NO: 814	327 aa	MW at 36607.4kD

GSTMTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLFIIQADAAL
HEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSFSIMESAVLLAMAFDRY
VAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFHYCRGPVIAHCYCEHMAVVRLACGD
TSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQLASQEARYKAFGTCASHIGAILSTYTPVVISS
VMHRVARHAAPRVHILLAIFYLLFPPMVNPIIYGVKTKQIREYVLSLFQRKNMLE

NOV53d, 170645981 DNA Sequence	SEQ ID NO: 815	981 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCACCAAGACAAGGAGATTTCCAGGAGCCATGCTTCCCTCTAATATCACCTCAACACATCCAGC
TGTCTTTTTGTTGGTAGGAATTCCTGGTTTGGAACACCTGCATGCCTGGATCTCCATCCCCTTCTGCT
TTGCTTATACTCTGGCCCTGCTAGGCAACTGTACCCTTCTCTTCATTATCCAGGCTGATGCAGCCCTC
CATGAACCCATGTACCTCTTTCTGGCCATGTTGGCAACCATTGACTTGGTTCTTTCTTCTACAACGCT
GCCCAAAATGCTTGCCATATTCTGGTTCAGGGATCAGGAGATCAACTTCTTTGCCTGTCTGGTCCAGA
TGTTCTTCCTTCACTCCTTCTCCATCATGGAGTCAGCAGTGCTGCTGGCCATGGCCTTTGACCGCTAT
GTGGCCATCTGCAAGCCATTGCACTACACGACGGTCCTGACTGGGTCCCTCATCACCAAGATTGGCAT
GGCTGCTGTGGCCCGGGCTGTGACACTAATGACTCCACTCCCCTTCCTGCTCAGACGCTTCCACTACT
GCCGAGACCCAGTGATTGCCCATTGCTACTGTGAACACATGGCTGTGGTAAGGCTGGCGTGTGGGGAC
ACTAGCTTCAACAATATCTATGGCATTGCTGTGGCCATGTTTATTGTGGTGTTGGACCTGCTCTTTGT
TATCCTGTCTTATGTCTTCATCCTTCAGGCAGTTCTCCAGCTTGCCTCTCAGGAGGCCCGCTACAAGG
CATTTGGGACATGTGTGTCTCACATAGGTACCATCCTGTCCACCTACACTCCAGTAGTCATCTCTTCA
GTCATGCACCGTGTAGCCCGCCATGCTGCCCCTCGTGTCCACATACTCCTTGCTATTTTCTATCTCCT
TTTCCCACCCATGGTCAATCCTATCATATATGGAGTCAAGACCAAGCAGATTCGTGAGTATGTGCTCA
GTCTATTCCAGAGAAAGAACATGCTCGAG

NOV53d, 170645981 Protein Sequence	SEQ ID NO: 816	327 aa	MW at 36720.5kD

GSTKTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLFIIQADAAL
HEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSFSIMESAVLLAMAFDRY
VAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFHYCRDPVIAHCYCEHMAVVRLACGD
TSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQLASQEARYKAFGTCVSHIGTILSTYTPVVISS
VMHRVARHAAPRVHILLAIFYLLFPPMVNPIIYGVKTKQIREYVLSLFQRKNMLE

NOV53e, 168869262 DNA Sequence	SEQ ID NO: 817	981 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCACCATGACAAGGAGATTTCCAGGAGCCATGCTTCCCTCTAATATCACCTCAACACATCCAGC
TGTCTTTTTGTTGGTAGGAATTCCTGGTTTGGAACACCTGCATGCCTGGATCTCCATCCCCTTCTGCT
TTGCTTATACTCTGGCCCTGCTAGGCAACTGTACCCTTCTCTTCATTATCCGGGCTGATGCAGCCCTC
CATGAACCCATGTACCTCTTTCTGGCCATGTTGGCAACCATTGACTTGGTTCTTTCTTCTACAACGCT
GCCCAAAATGCTTGCCATATTCTGGTTCAGGGATCAGGAGATCAACTTCTTTGCCTGTCTGGTCCAGA
TGTTCTTCCTTCACTCCTTCTCCATCATGGAGTCAGCAGTGCTGCTGGCCATGGCCTTTGACCGCTAT
GTGGCCATCTGCAAGCCATTGCACTACACGACGGTCCTGACTGGGTCCCTCATCACCAAGATTGGCAT
GGCTGCTGTGGCCTGGGCTGTGACACTAATGACTCCACTCCCCTTCCTGCTCAGACGCTTCCACTACT
GCCGAGGCCCAGTGATTGCCCATTGCTACTGTGAACACATGGCTGTGGTAAGGCTGGCGTGTGGGGAC
ACTAGCTTCAACAATATCTATGGCATTGCTGTGGCCATGTTTATTGTGGTGTTGGACCTGCTCTTTGT
TATCCTGTCTTATGTCTTCATCCTTCAGGCAGTTCTCCAGCTTGCCTCTCAGGAGGCCCGCTACAAGG
CATTTGGGACATGTGTGTCTCACATAGGTGCCATCCTGTCCACCTACACTCCAGTAGTCATCTCTTCA
GTCATGCACCGTGTAGCCCGCCATGCTGCCCCTCGTGTCCACATACTCCTTGCTATTTTCTATCTCCT
TTTCCCACCCATGGTCAATCCTATCATATATGGAGTCAAGACCAAGCAGATTCGTGAGTATGTGCTCA
GTCTATTCCAGAGAAAGAACATGCTCGAG

NOV53e, 168869262 Protein Sequence	SEQ ID NO: 818	327 aa	MW at 36693.5kD

GSTMTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLFIIRADAAL
HEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSFSIMESAVLLAMAFDRY
VAICKPLHYTTVLTGSLITKIGMAAVAWAVTLMTPLPFLLRRFHYCRGPVIAHCYCEHMAVVRLACGD
TSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQLASQEARYKAFGTCVSHIGAILSTYTPVVISS
VMHRVARHAAPRVHILLAIFYLLFPPMVNPIIYGVKTKQIREYVLSLFQRKNMLE

NOV53f, 168869254 DNA Sequence	SEQ ID NO: 819	981 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCACCATGACAAGGAGATTTCCAGGAGCCATGCTTCCCTCTAATATCACCTCAACACATCCAGC
TGTCTTTTTGTTGGTAGGAATTCCTGGTTTGGAACACCTGCATGCCTGGATCTCCATCCCCCTCTGCT
TTGCTTATACTCTGGCCCTGCTAGGCAACTGTACCCTTCTCTTCATTATCCAGGCTGATGCAGCCCTC
CATGAACCCATGTACCTCTTTCTGGCCATGTTGGCAACCATTGACTTGGTTCTTTCTTCTACAACGCT
GCCCAAAATGCTTGCCATATTCTGGTTCAGGGATCAGGAGATCAACTTCTTTGCCTGTCTGGTCCAGA
TGTTCTTCCTTCACTCCTTCTCCATCATGGAGTCAGCAGTGCTGCTGGCCATGGCCTTTGACCGCTAT
GTGGCCATCTGCAAGCCATTGCACTACACGACGGTCCTGACTGGGTCCCTCATCACCAAGATTGGCAT
GGCTGCTGTGGCCCGGGCTGTGACACTAATGACTCCACTCCCCTTCCTGCTCAGACGCTTCCACTACT
GCCGAGGCCCAGTGATTGCCCATTGCTACTGTGAACACATGGCTGTGGTAAGGCTGGCGTGTGGGGAC
ACTAGCTTCAACAATATCTATGGCATTGCTGTGGCCATGTTTATTGTGGTGTTGGACCTGCTCTTTGT
TATCCTGTCTTATGTCTTCATCCTTCAGGCAGTTCTCCAGCTTGCCTCTCAGGAGGCCCGCTACAAGG
CATTTGGGACATGTGTGTCTCACATAGGTGCCATCCTGTCCACCTACACTCCAGTAGTCATCTCTTCA
GTCATGCACCGTGTAGCCCGCCATGCTGCCCCTCGTGTCCACATACTCCTTGCTATTTTCTATCTCCT
TTTCCCACCCATGGTCAATCCTATCATATATGGAGTCAAGACCAAGCAGATTCGTGAGTATGTGCTCA
GTCTATTCCAGAGAAAGAACATGCTCGAG

NOV53f, 168869254 Protein Sequence	SEQ ID NO: 820	327 aa	MW at 36601.4kD

GSTMTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPLCFAYTLALLGNCTLLFIIQADAAL
HEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSFSIMESAVLLAMAFDRY
VAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFHYCRGPVIAHCYCEHMAVVRLACGD
TSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQLASQEARYKAFGTCVSHIGAILSTYTPVVISS
VMHRVARHAAPRVHILLAIFYLLFPPMVNPIIYGVKTKQIREYVLSLFQRKNMLE

NOV53g, CG56144-02 DNA Sequence	SEQ ID NO: 821	966 bp
	ORF Start: ATG at 1	ORF Stop: end of sequence

ATGACAAGGAGATTTCCAGGAGCCATGCTTCCCTCTAATATCACCTCAACACATCCAGCTGTCTTTTT
GTTGGTAGGAATTCCTGGTTTGGAACACCTGCATGCCTGGATCTCCATCCCCTTCTGCTTTGCTTATA
CTCTGGCCCTGCTAGGCAACTGTACCCTTCTCTTCATTATCCAGGCTGATGCAGCCCTCCATGAACCC
ATGTACCTCTTTCTGGCCATGTTGGCAACCATTGACTTGGTTCTTTCTTCTACAACGCTGCCCAAAAT
GCTTGCCATATTCTGGTTCAGGGATCAGGAGATCAACTTCTTTGCCTGTCTGGTCCAGATGTTCTTCC
TTCACTCCTTCTCCATCATGGAGTCAGCAGTGCTGCTGGCCATGGCCTTTGACCGCTATGTGGCCATC
TGCAAGCCATTGCACTACACGACGGTCCTGACTGGGTCCCTCATCACCAAGATTGGCATGGCTGCTGT
GGCCCGGGCTGTGACACTAATGACTCCACTCCCCTTCCTGCTCAGACGCTTCCACTACTGCCGAGGCC
CAGTGATTGCCCATTGCTACTGTGAACACATGGCTGTGGTAAGGCTGGCGTGTGGGGACACTAGCTTC
AACAATATCTATGGCATTGCTGTGGCCATGTTTATTGTGGTGTTGGACCTGCTCTTTGTTATCCTGTC
TTATGTCTTCATCCTTCAGGCAGTTCTCCAGCTTGCCTCTCAGGAGGCCCGCTACAAGGCATTTGGGA
CATGTGTGTCTCACATAGGTGCCATCCTGTCCACCTACACTCCAGTAGTCATCTCTTCAGTCATGCAC
CGTGTAGCCCGCCATGCTGCCCCTCGTGTCCACATACTTCTTGCTATTTTCTATCTCCTTTTCCCACC
CATGGTCAATCCTATCATATATGGAGTCAAGACCAAGCAGATTCGTGAGTATGTGCTCAGTCTATTCC
AGAGAAAGAACATG

NOV53g, CG56144-02 Protein Sequence	SEQ ID NO: 822	322 aa	MW at 36148.0kD

MTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLFIIQADAALHEP
MYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSFSIMESAVLLAMAFDRYVAI
CKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFHYCRGPVIAHCYCEHMAVVRLACGDTSF
NNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQLASQEARYKAFGTCVSHIGAILSTYTPVVISSVMH

RVARHAAPRVHILLAIFYLLFPPMVNPIIYGVKTKQIREYVLSLFQRKNM			
NOV53h, CG56144-03 DNA Sequence	SEQ ID NO: 823	777 bp	
	ORF Start: at 1	ORF Stop: end of sequence	
GCCCTCCATGAACCCATGTACCTCTTTCTGGCCATGTTGGCAACCATTGACTTGGTTCTTTCTTCTAC AACGCTGCCCAAAATGCTTGCCATATTCTGGTTCAGGGATCAGGAGATCAACTTCTTTGCCTGTCTGG TCCAGATGTTCTTCCTTCACTCCTTCTCCATCATGGAGTCAGCAGTGCTGCTGGCCATGGCCTTTGAC CGCTATGTGGCCATCTGCAAGCCATTGCACTACACGACGGTCCTGACTGGGTCCCTCATCACCAAGAT TGGCATGGCTGCTGTGGCCTGGGCTGTGACACTAATGACTCCACTCCCCTTCCTGCTCAGACGCTTCC ACTACTGCCGAGGCCCAGTGATTGCCCATTGCTACTGTGAACACATGGCTGTGGTAAGGCTGGCGTGT GGGGACACTAGCTTCAACAATATCTATGGCATTGCTGTGGCCATGTCTATTGTGGTGTTGGACCTGCT CTTTGTTATCCTGTCTTATGTCTTCATCCTTCAGGCAGTTCTCCAGCTTGCCTCTCAGGAGGCCCGCT ACAAGGCATTTGGGACATGTGTGTCTCACATAGGTGCCATCCTGTCCACCTACACTCCAGTAGTCATC TCTTCAGTCATGCACCGTGTAGCCCGCCATGCTGCCCCTCGTGTCCACATACTCCTTGCTATTTTCTA TCTCCTTTTCCCACCCATGGTCAATCCTATCATATATGGAGTCAAGACCAAGCAGATTCGTGAGTATG TGCTCAGTCTATTCCAGAGAAAGAACATG			
NOV53h, CG56144-03 Protein Sequence	SEQ ID NO: 824	259 aa	MW at 29252.8kD
ALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSFSIMESAVLLAMAFD RYVAICKPLHYTTVLTGSLITKIGMAAVAWAVTLMTPLPFLLRRFHYCRGPVIAHCYCEHMAVVRLAC GDTSFNNIYGIAVAMSIVVLDLLFVILSYVFILQAVLQLASQEARYKAFGTCVSHIGAILSTYTPVVI SSVMHRVARHAAPRVHILLAIFYLLFPPMVNPIIYGVKTKQIREYVLSLFQRKNM			
NOV53i, CG56144-04 DNA Sequence	SEQ ID NO: 825	777 bp	
	ORF Start: at 1	ORF Stop: end of sequence	
GCCCTCCATGAACCCATGTACCTCTTTCTGGCCATGTTGGCAACCATTGACTTGGTTCTTTCTTCTAC AACGCTGCCCAAAATGCTTGCCATATTCTGGTTCAGGGATCAGGAGATCAACTTCTTTGCCTGTCTGG TCCAGATGTTCTTCCTTCACTCCTTCTCCATCATGGAGTCAGCAGTGCTGCTGGCCATGGCCTTTGAC CGCTATGTGGCCATCTGCAAGCCATTGCACTACACGACGGTCCTGACTGGGTCCCTCATCACCAAGAT TGGCATGGCTGCTGTGGCCTGGGCTGTGACACTAATGACTCCACTCCCCTTCCTGCTCAGACGCTTCC ACTACTGCCGAGGCCCAGTGATTGCCCATTGCTACTGTGAACACATGGCTGTGGTAAGGCTGGCGTGT GGGGACACTAGCTTCAACAATATCTATGGCATTGCTGTGGCCATGTCTATTGTGGTGTTGGACCTGCT CTTTGTTATCCTGTCTTATGTCTTCATCCTTCAGGCAGTTCTCCAGCTTGCCTCTCAGGAGGCCCGCT ACAAGGCATTTGGGACATGTGTGTCTCACATAGGTGCCATCCTGTCCACCTACACTCCAGTAGTCATC TCTTCAGTCATGCACCGTGTAGCCCGCCATGCTGCCCCTCGTGTCCACATACTCCTTGCTATTTTCTA TCTCCTTTTCCCACCCATGGTCAATCCTATCATATATGGAGTCAAGACCAAGCAGATTCGTGAGTATG TGCTCAGTCTATTCCAGAGAAAGAACATG			
NOV53i, CG56144-04 Protein Sequence	SEQ ID NO: 826	259 aa	MW at 29252.8kD
ALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSFSIMESAVLLAMAFD RYVAICKPLHYTTVLTGSLITKIGMAAVAWAVTLMTPLPFLLRRFHYCRGPVIAHCYCEHMAVVRLAC GDTSFNNIYGIAVAMSIVVLDLLFVILSYVFILQAVLQLASQEARYKAFGTCVSHIGAILSTYTPVVI SSVMHRVARHAAPRVHILLAIFYLLFPPMVNPIIYGVKTKQIREYVLSLFQRKNM			
NOV53j, CG56144-05 DNA Sequence	SEQ ID NO: 827	789 bp	
	ORF Start: at 7	ORF Stop: at 784	
GGATCCGCCCTCCATGAACCCATGTACCTCTTTCTGGCCATGTTGGCAACCATTGACTTGGTTCTTTC TTCTACAACGCTGCCCAAAATGCTTGCCATATTCTGGTTCAGGGATCAGGAGATCAACTTCTTTGCCT GTCTGGTCCAGATGTTCTTCCTTCACTCCTTCTCCATCATGGAGTCAGCAGTGCTGCTGGCCATGGCC TTTGACCGCTATGTGGCCATCTGCAAGCCATTGCACTACACGACGGTCCTGACTGGGTCCCTCATCAC CAAGATTGGCATGGCTGCTGTGGCCCGGGCTGTGACACTAATGACTCCACTCCCCTTCCTGCTCAGAC GCTTCCACTACTGCCGAGGCCCAGTGATTGCCCATTGCTACTGTGAACACATGGCTGTGGTAAGGCTG GCGTGTGGGGACACTAGCTTCAACAATATCTATGGCATTGCTGTGGCCATGTTTATTGTGGTGTTGGA CCTGCTCTTTGTTATCCTGTCTTATGTCTTCATCCTTCAGGCAGTTCTCCAGCTTGCCTCTCAGGAGG CCCGCTACAAGGCATTTGGGACATGTGTGTCTCACATAGGTGCCATCCTGTCCACCTACACTCCAGTA GTCATCTCTTCAGTCATGCACCGTGTAGCCCGCCATGCTGCCCCTCGTGTCCACATACTCCTTGCTAT			

TTTCTATCTCCTTTTCCCACCCATGGTCAATCCTATCATATATGGAGTCAAGACCAAGCAGATTCGTG AGTATGTGCTCAGTCTATTCCAGAGAAAGAACATGCTCGAG

NOV53j, CG56144-05 Protein Sequence	SEQ ID NO: 828	259 aa	MW at 29282.9kD

ALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSFSIMESAVLLAMAFD RYVAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFHYCRGPVIAHCYCEHMAVVRLAC GDTSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQLASQEARYKAFGTCVSHIGAILSTYTPVVI SSVMHRVARHAAPRVHILLAIFYLLFPPMVNPIIYGVKTKQIREYVLSLFQRKNM

NOV53k, CG56144-06 DNA Sequence	SEQ ID NO: 829	789 bp
	ORF Start: at 7	ORF Stop: at 784

GGATCCGCCCTCCATGAACCCATGTACCTCTTTCTGGCCATGTTGGCAACCATTGACTTGGTTCTTTC TTCTACAACGCTGCCCAAAATGCTTGCCATATTCTGGTTCAGGGATCAGGAGATCAACTTCTTTGCCT GTCTGGTCCAGATGTTCTTCCTTCACTCCTTCTCCATCATGGAGTCAGCAGTGCTGCTGGCCATGGCC TTTGACCGCTATGTGGCCATCTGCAAGCCATTGCACTACACGACGGTCCTGACTGGGTCCCTCATCAC CAAGATTGGCATGGCTGCTGTGGCCCGGGCTGTGACACTAATGACTCCACTCCCCTTCCTGCTCAGAC GCTTCCACTACTGCCGAGGCCCAGTGATTGCCCATTGCTACTGTGAACACATGGCTGTGGTAAGGCTG GCGTGTGGGGACACTAGCTTCAACAATATCTATGGCATTGCTGTGGCCATGTTTATTGTGGTGTTGGA CCTGCTCTTTGTTATCCTGTCTTATGTCTTCATCCTTCAGGCAGTTCTCCAGCTTGCCTCTCAGGAGG CCCGCTACAAGGCATTTGGGACATGTGTGTCTCACATAGGTGCCATCCTGTCCACCTACACTCCAGTA GTCATCTCTTCAGTCATGCACCGTGTAGCCCGCCATGCTGCCCCTCGTGTCCACATACTCCTTGCTAT TTTCTATCTCCTTTTCCCACCCATGGTCAATCCTATCATATATGGAGTCAAGACCAAGCAGATTCGTG AGTATGTGCTCAGTCTATTCCAGAGAAAGAACATGCTCGAG

NOV53k, CG56144-06 Protein Sequence	SEQ ID NO: 830	259 aa	MW at 29282.9kD

ALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSFSIMESAVLLAMAFD RYVAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFHYCRGPVIAHCYCEHMAVVRLAC GDTSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQLASQEARYKAFGTCVSHIGAILSTYTPVVI SSVMHRVARHAAPRVHILLAIFYLLFPPMVNPIIYGVKTKQIREYVLSLFQRKNM

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 53B.

Table 53B. Comparison of the NOV53 protein sequences.

```
NOV53a  ---MTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLF
NOV53b  GSTMTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLF
NOV53c  GSTMTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLF
NOV53d  GSTKTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLF
NOV53e  GSTMTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLF
NOV53f  GSTMTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPLCFAYTLALLGNCTLLF
NOV53g  ---MTRRFPGAMLPSNITSTHPAVFLLVGIPGLEHLHAWISIPFCFAYTLALLGNCTLLF
NOV53h  ------------------------------------------------------------
NOV53i  ------------------------------------------------------------
NOV53j  ------------------------------------------------------------
NOV53k  ------------------------------------------------------------

NOV53a  IIQADAALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSF
NOV53b  IIQADAALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSF
NOV53c  IIQADAALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSF
NOV53d  IIQADAALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSF
NOV53e  IIRADAALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSF
NOV53f  IIQADAALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSF
```

```
NOV53g  IIQADAALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSF
NOV53h  ------ALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSF
NOV53i  ------ALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSF
NOV53j  ------ALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSF
NOV53k  ------ALHEPMYLFLAMLATIDLVLSSTTLPKMLAIFWFRDQEINFFACLVQMFFLHSF

NOV53a  SIMESAVLLAMAFDRYVAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFH
NOV53b  SIMESAVLLAMAFDRYVAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFH
NOV53c  SIMESAVLLAMAFDRYVAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFH
NOV53d  SIMESAVLLAMAFDRYVAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFH
NOV53e  SIMESAVLLAMAFDRYVAICKPLHYTTVLTGSLITKIGMAAVAWAVTLMTPLPFLLRRFH
NOV53f  SIMESAVLLAMAFDRYVAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFH
NOV53g  SIMESAVLLAMAFDRYVAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFH
NOV53h  SIMESAVLLAMAFDRYVAICKPLHYTTVLTGSLITKIGMAAVAWAVTLMTPLPFLLRRFH
NOV53i  SIMESAVLLAMAFDRYVAICKPLHYTTVLTGSLITKIGMAAVAWAVTLMTPLPFLLRRFH
NOV53j  SIMESAVLLAMAFDRYVAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFH
NOV53k  SIMESAVLLAMAFDRYVAICKPLHYTTVLTGSLITKIGMAAVARAVTLMTPLPFLLRRFH

NOV53a  YCRGPVIAHCYCEHMAVVRLACGDTSFNNIYGIAVAMFSVVLDLLFVILSYVFILQAVLQ
NOV53b  YCRGPVIAHCYCEHMAVVRLACGDTSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQ
NOV53c  YCRGPVIAHCYCEHMAVVRLACGDTSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQ
NOV53d  YCRDPVIAHCYCEHMAVVRLACGDTSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQ
NOV53e  YCRGPVIAHCYCEHMAVVRLACGDTSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQ
NOV53f  YCRGPVIAHCYCEHMAVVRLACGDTSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQ
NOV53g  YCRGPVIAHCYCEHMAVVRLACGDTSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQ
NOV53h  YCRGPVIAHCYCEHMAVVRLACGDTSFNNIYGIAVAMSIVVLDLLFVILSYVFILQAVLQ
NOV53i  YCRGPVIAHCYCEHMAVVRLACGDTSFNNIYGIAVAMSIVVLDLLFVILSYVFILQAVLQ
NOV53j  YCRGPVIAHCYCEHMAVVRLACGDTSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQ
NOV53k  YCRGPVIAHCYCEHMAVVRLACGDTSFNNIYGIAVAMFIVVLDLLFVILSYVFILQAVLQ

NOV53a  LASQEARYKAFGTCVSHIGAILSTYTPVVISSVMHRVARHAAPRVHILLAIFYLLFPPMV
NOV53b  LASQEARYKAFGTCVSHIGAILSTYTPVVISSVMHRVARHAAPRVHILLAIFYLLFPPMV
NOV53c  LASQEARYKAFGTCASHIGAILSTYTPVVISSVMHRVARHAAPRVHILLAIFYLLFPPMV
NOV53d  LASQEARYKAFGTCVSHIGTILSTYTPVVISSVMHRVARHAAPRVHILLAIFYLLFPPMV
NOV53e  LASQEARYKAFGTCVSHIGAILSTYTPVVISSVMHRVARHAAPRVHILLAIFYLLFPPMV
NOV53f  LASQEARYKAFGTCVSHIGAILSTYTPVVISSVMHRVARHAAPRVHILLAIFYLLFPPMV
NOV53g  LASQEARYKAFGTCVSHIGAILSTYTPVVISSVMHRVARHAAPRVHILLAIFYLLFPPMV
NOV53h  LASQEARYKAFGTCVSHIGAILSTYTPVVISSVMHRVARHAAPRVHILLAIFYLLFPPMV
NOV53i  LASQEARYKAFGTCVSHIGAILSTYTPVVISSVMHRVARHAAPRVHILLAIFYLLFPPMV
NOV53j  LASQEARYKAFGTCVSHIGAILSTYTPVVISSVMHRVARHAAPRVHILLAIFYLLFPPMV
NOV53k  LASQEARYKAFGTCVSHIGAILSTYTPVVISSVMHRVARHAAPRVHILLAIFYLLFPPMV

NOV53a  NPIIYGVKTKQIREYVLSLFQRKNM--
NOV53b  NPIIYGVKTKQIREYVLSLFQRKNMLE
NOV53c  NPIIYGVKTKQIREYVLSLFQRKNMLE
NOV53d  NPIIYGVKTKQIREYVLSLFQRKNMLE
NOV53e  NPIIYGVKTKQIREYVLSLFQRKNMLE
NOV53f  NPIIYGVKTKQIREYVLSLFQRKNMLE
NOV53g  NPIIYGVKTKQIREYVLSLFQRKNM--
NOV53h  NPIIYGVKTKQIREYVLSLFQRKNM--
NOV53i  NPIIYGVKTKQIREYVLSLFQRKNM--
NOV53j  NPIIYGVKTKQIREYVLSLFQRKNM--
NOV53k  NPIIYGVKTKQIREYVLSLFQRKNM--
```

NOV53a	(SEQ ID NO: 810)
NOV53b	(SEQ ID NO: 812)
NOV53c	(SEQ ID NO: 814)
NOV53d	(SEQ ID NO: 816)
NOV53e	(SEQ ID NO: 818)
NOV53f	(SEQ ID NO: 820)
NOV53g	(SEQ ID NO: 822)
NOV53h	(SEQ ID NO: 824)
NOV53i	(SEQ ID NO: 826)
NOV53j	(SEQ ID NO: 828)
NOV53k	(SEQ ID NO: 830)

Further analysis of the NOV53a protein yielded the following properties shown in Table 53C.

Table 53C. Protein Sequence Properties NOV53a

SignalP analysis:	Cleavage site between residues 64 and 65

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 4;  pos.chg 2;  neg.chg 0
      H-region:  length 26;  peak value  10.58
      PSG score:   6.18

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.34
      possible cleavage site: between 61 and 62

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   7
      INTEGRAL    Likelihood = -3.29   Transmembrane   43 -  59
      INTEGRAL    Likelihood = -2.07   Transmembrane   69 -  85
      INTEGRAL    Likelihood = -1.70   Transmembrane  104 - 120
      INTEGRAL    Likelihood =  0.26   Transmembrane  157 - 173
      INTEGRAL    Likelihood =-11.09   Transmembrane  210 - 226
      INTEGRAL    Likelihood =  0.21   Transmembrane  255 - 271
      INTEGRAL    Likelihood = -3.03   Transmembrane  284 - 300
      PERIPHERAL  Likelihood =  1.54 (at 21)
      ALOM score: -11.09  (number of TMSs: 7)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 50
      Charge difference: -1.5   C(-1.5) - N( 0.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       2       Hyd Moment(75): 12.17
```

```
      Hyd Moment(95): 11.23    G content:       3
      D/E content:     1       S/T content:     5
      Score: -1.65

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 14  RRF|PG

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   6.8%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: TRRF


      KKXX-like motif in the C-terminus: QRKN

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: found
      TLPK at 87

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1
```

```
COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       77.8 %: endoplasmic reticulum
       11.1 %: nuclear
       11.1 %: mitochondrial

>> prediction for CG56144-01 is end (k=9)
```

A search of the NOV53a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 53D.

Table 53D. Geneseq Results for NOV53a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV53a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE11904	Human G-protein coupled receptor 10b (GPCR10b) protein - Homo sapiens, 322 aa. [WO200181378-A2, 01-NOV-2001]	1..322 1..322	318/322 (98%) 321/322 (98%)	0.0
AAE11903	Human G-protein coupled receptor 10a (GPCR10a) protein - Homo sapiens, 321 aa. [WO200181378-A2, 01-NOV-2001]	1..322 1..321	321/322 (99%) 321/322 (99%)	0.0
AAU85178	G-coupled olfactory receptor #39 - Homo sapiens, 314 aa. [WO200198526-A2, 27-DEC-2001]	9..322 1..314	314/314 (100%) 314/314 (100%)	e-180
ABP95675	Human GPCR polypeptide SEQ ID NO 160 - Homo sapiens, 314 aa. [WO200216548-A2, 28-FEB-2002]	9..322 1..314	314/314 (100%) 314/314 (100%)	e-180
AAU24558	Human olfactory receptor AOLFR45 - Homo sapiens, 314 aa. [WO200168805-A2, 20-SEP-2001]	9..322 1..314	314/314 (100%) 314/314 (100%)	e-180

In a BLAST search of public sequence databases, the NOV53a protein was found to have homology to the proteins shown in the BLASTP data in Table 53E.

Table 53E. Public BLASTP Results for NOV53a

Protein Accession Number	Protein/Organism/Length	NOV53a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8NGK4	Seven transmembrane helix receptor - Homo sapiens (Human), 314 aa.	9..322 1..314	311/314 (99%) 312/314 (99%)	e-178
Q8NGK3	Seven transmembrane helix receptor - Homo sapiens (Human), 314 aa.	9..322 1..314	283/314 (90%) 290/314 (92%)	e-161
Q8VH04	Olfactory receptor MOR28-1 - Mus musculus (Mouse), 317 aa.	9..322 1..314	273/314 (86%) 291/314 (91%)	e-157
CAD20423	Sequence 1 from Patent WO0179295 - Homo sapiens (Human), 316 aa.	9..316 1..307	276/308 (89%) 287/308 (92%)	e-154
Q8VG23	Olfactory receptor MOR29-1 - Mus musculus (Mouse), 315 aa.	9..319 1..310	172/311 (55%) 224/311 (71%)	3e-96

PFam analysis indicates that the NOV53a protein contains the domains shown in the Table 53F.

Table 53F. Domain Analysis of NOV53a

Pfam Domain	NOV53a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1	51..302	50/277 (18%) 163/277 (59%)	4.5e-13

Example 54.

The NOV54 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 54A.

Table 54A. NOV54 Sequence Analysis

NOV54a, CG56146-02 DNA Sequence	SEQ ID NO: 831	807 bp
	ORF Start: ATG at 1	ORF Stop: end of sequence

ATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCCTTCTTAGACCT
GTGTCTCATTTCTGACACAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGACTCCATCTCCT
TCCTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGGATCAGAGATTGGCATCCTTACT
GCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTGAGGCTGTCATGAGCAGAGG
GCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTTGTACACAGCTG
GAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATGTCCCTGCCCTA
CTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGTGGCCATTGGGGTCTGTTATGC
ATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTTAAGGATATCAC
AGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCACTGTGTTTCTT
GTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTTCTATTCTAGACTTGCTGGTGTC
TGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAACAAGGACATTA
AATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATGAC

NOV54a, CG56146-02 Protein Sequence	SEQ ID NO: 832	269 aa	MW at 29515.5kD

MILDHRLHMAMYFFLRHLSFLDLCLISDTVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILT
AMSYDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVPAL
LKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTVFL
VTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDD

NOV54b, CG56146-01 DNA Sequence	SEQ ID NO: 833	995 bp
	ORF Start: ATG at 21	ORF Stop: TAA at 966

GGCGCTTATAATTTTGAACT**ATG**ACCAATCAGACACAGATGATGGAATTCTTGCTTGTGAGATTTACT
GAGAATTGGGTGCTCCTGAGGCTGCATGCTTTGCTCTTCTCACTGATCTACCTCACGGCTGTGCTGAT
GAATTTAGTCATCATTCTCCTCATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCC
GACATTTGTCCTTCTTAGACCTGTGTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTC
GCCTCCACTGACTCCATCTCCTTCCTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGG
ATCAGAGATTGGCATCCTTACTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACT
GTGAGGCTGTCATGAGCAGAGGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCC
TTGGGACTCTTGTACACAGCTGGAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTT
CTTCTGCGATGTCCCTGCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTG
TGGCCATTGGGGTCTGTTATGCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTC
TCTGCTGTGTTAAGGATATCACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCT
CATTGTTGTCACTGTGTTTCTTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTT
CTATTCTAGACTTGCTGGTGTCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTAC
TGTCTGAAGAACAAGGACATTAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGT
AATGAAAGATGAC**TAA**AGTTGAAGATGGGAAGTACTTTTTTTG

NOV54b, CG56146-01 Protein Sequence	SEQ ID NO: 834	315 aa	MW at 34934.2kD

MTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAMYFFLRHLSFLD
LCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDRYAAICCPLHCEAVMSR
GLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVPALLKLTCSKEHAIISVSVAIGVCY
AFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLV
SVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDD

NOV54c, 170646057 DNA Sequence	SEQ ID NO: 835	819 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCCTTCTT
AGACCTGTGTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGACTCCA
TCTCCTTCCTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGGATCAGAGATTGGCATC
CTTACTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTGAGGCTGTCATGAG
CAGAGGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTTGTACA

CAGCTGGAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATGTCCCT
GCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGTGGCCATTGGGGTCTG
TTATGCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTTAAGGA
TATCACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCACTGTG
TTTCTTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATACACCTTCTATTCTAGACTTGCT
GGTGTCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAACAAGG
ACATTAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATGACCTC
GAG

NOV54c, 170646057 Protein Sequence	SEQ ID NO: 836	273 aa	MW at 29888.0kD

GSMILDHRLHMAMYFFLRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGI
LTAMSYDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVP
ALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTV
FLVTGAVAYLKPGSDTPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDDL
E

NOV54d, 170646049 DNA Sequence	SEQ ID NO: 837	819 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCCTTCTT
AGACCTGTGTCTCATTTCTGCCGCAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGACTCCA
TCTCCTTCCTGGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGGATCAGAGATTGGCATC
CTTACTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTGAGGCTGTCATGAG
CAGAGGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTTGTACA
CAGCTGGAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATGTCCCT
GCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGTGGCCATTGGGGTCTG
TTATGCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTTAAGGA
TATCACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCACTGTG
TTTCTTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTTCTATTCTAGACTTGCT
GGTGTCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAACAAGG
ACATTAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATGACCTC
GAG

NOV54d, 170646049 Protein Sequence	SEQ ID NO: 838	273 aa	MW at 29827.9kD

GSMILDHRLHMAMYFFLRHLSFLDLCLISAAVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGI
LTAMSYDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVP
ALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTV
FLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDDL
E

NOV54e, 170646053 DNA Sequence	SEQ ID NO: 839	819 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCCTTCTT
AGACCTGTGTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGACTCCA
TCTCCTTCCTGGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCCGGCTGGATCAGAGATTGGCATC
CTTACTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTGAGGCTGTCATGAG
CAGAGGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTTGTACA
CAGCTGGAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATGTCCCT
GCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGGGGCCATTGGGGTCTG
TTATGCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTTAAGGA
TATCACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCACTGTG
TTTCTTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTTCTATTCTAGACTTGCT
GGTGTCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAACAAGG
ACATTAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATGACCTC
GAG

5

NOV54e, 170646053 Protein Sequence	SEQ ID NO: 840	273 aa	MW at 29799.8kD

GSMILDHRLHMAMYFFLRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLPAGSEIGI
LTAMSYDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVP
ALLKLTCSKEHAIISVSGAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTV
FLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDDL
E

NOV54f, 174307717 DNA Sequence	SEQ ID NO: 841	960 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCACCATGACCAATCAGACACAGATGATGGAATTCTTGCTTGTGAGATTTACTGAGAATTGGGT
GCTCCTGAGGCTGCATGCTTTGCTCTTCTCACTGATCTACCTCACGGCTGTGCTGATGAATTTAGTCA
TCATTCTCCTCATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCC
TTCTTAGACCTGTGTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGA
CTCCATCTCCTTCCTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGGATCAGAGATTG
GCATCCCTACTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTAAGGCTGTC
ATGAGCAGAGGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTT
GTACACAGCTGGAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATG
TCCCTGCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGTGGCCATTGGG
GTCTGTTATGCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTT
AAGGATATCACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCA
CTGTGTTTCTTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTTCTATTCTAGAC
TTGCTGGTGTCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAA
CAAGGACATTAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATG
ACCTCGAG

NOV54f, 174307717 Protein Sequence	SEQ ID NO: 842	320 aa	MW at 35404.7kD

GSTMTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAMYFFLRHLS
FLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGIPTAMSYDRYAAICCPLHCKAV
MSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVPALLKLTCSKEHAIISVSVAIG
VCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTVFLVTGAVAYLKPGSDAPSILD
LLVSVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDDLE

NOV54g, 168869383 DNA Sequence	SEQ ID NO: 843	820 bp
	ORF Start: at 296	ORF Stop: end of sequence

GGATCCATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCCTTCTT
AGACCTGTGTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGACTCCA
TCTCCTTCCTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGGATCAGAGATTGGCATC
CTTACTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTGAGGCTGTCATGAG
CAGAGGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTTGTACA
CAGCTGGAACATTCTCTTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATGTCCC
TGCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGTGGCCATTGGGGTCT
GTTATGCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTTAAGG
ATATCACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCACTGT
GTTTCTTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTTCTATTCTAGACTTGC
TGGTGTCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAACAAG
GACATTAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATGACCT
CGAG

NOV54g, 168869383 Protein Sequence	SEQ ID NO: 844	175 aa	MW at 19239.2kD

WLCPGSTEGPWDSCTQLEHSLLNFYGSDELHQFFCDVPALLKLTCSKEHAIISVSVAIGVCYAFSCLV
CIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSV
APPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDDLE

NOV54h, CG56146-03 DNA Sequence	SEQ ID NO: 845	945 bp
	ORF Start: ATG at 1	ORF Stop: end of sequence

ATGACCAATCAGACACAGATGATGGAATTCTTGCTTGTGAGATTTACTGAGAATTGGGTGCTCCTGAG
GCTGCATGCTTTGCTCTTCTCACTGATCTACCTCACGGCTGTGCTGATGAATTTAGTCATCATTCTCC
TCATGATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCCTTCTTAGAC
CTGTGTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGACTCCATCTC
CTTCCTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGGATCAGAGATTGGCATCCTTA
CTGCCATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTGAGGCTGTCATGAGCAGA
GGGCTCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTTGTACACAGC
TGGAACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATGTCCCTGCCC
TACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGTGGCCATTGGGGTCTGTTAT
GCATTTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTTAAGGATATC
ACAGAGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCACTGTGTTTC
TTGTAACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTTCTATTCTAGACTTGCTGGTG
TCTGTGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAACAAGGACAT
TAAATCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATGAC

NOV54h, CG56146-03 Protein Sequence	SEQ ID NO: 846	315 aa	MW at 34934.2kD

MTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAMYFFLRHLSFLD
LCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDRYAAICCPLHCEAVMSR
GLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVPALLKLTCSKEHAIISVSVAIGVCY
AFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLV
SVFYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDD

NOV54i, 262939640 DNA Sequence	SEQ ID NO: 847	952 bp
	ORF Start: at 3	ORF Stop: TAG at 942

CCAATCAGACACAGATGATGGAATTCTTGCTTGTGAGATTTACTGAGAATTGGGTGCTCCTGAGGCTG
CATGCTTTGCTCTTCTCACTGATCTACCTCACGGCTGTGCTGATGAATTTAGTCATCATTCTCCTCAT
GATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCCTTCTTAGACCTGT
GTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGACTCCATCTCCTTC
CTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGGATCAGAGATTGGCATCCTTACTGC
CATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTGAGGCTGTCATGAGCAGAGGGC
TCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTTGTACACAGCTGGA
ACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATGTCCCTGCCCTACT
AAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGTGGCCATTGGGGTCTGTTATGCAT
TTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTTAAGGATATCACAG
AGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCACTGTGTTTCTTGT
AACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTTCTATTCTAGACTTGCTGGTGTCTG
TGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAACAAGGACATTAAA
TCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATGACTAGCAATTGGG

NOV54i, 262939640 Protein Sequence	SEQ ID NO: 848	313 aa	MW at 34701.9kD

NQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAMYFFLRHLSFLDLC
LISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDRYAAICCPLHCEAVMSRGL
CVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVPALLKLTCSKEHAIISVSVAIGVCYAF
SCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSV
FYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDD

NOV54j, CG56146-04 DNA Sequence	SEQ ID NO: 849	60 bp
	ORF Start: ATG at 1	ORF Stop: end of sequence

ATGATGGAATTCTTGCTTGTGAGATTTACTGAGAATTGGGTGCTCCTGAGGCTGCATGCT

NOV54j, CG56146-04 Protein Sequence	SEQ ID NO: 850	20 aa	MW at 2519.0kD
MMEFLLVRFTENWVLLRLHA			

NOV54k, CG56146-05 DNA Sequence	SEQ ID NO: 851	60 bp
	ORF Start: at 1	ORF Stop: end of sequence
GATGTCCCTGCCCTACTAAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGT		

NOV54k, CG56146-05 Protein Sequence	SEQ ID NO: 852	20 aa	MW at 2124.5kD
DVPALLKLTCSKEHAIISVS			

NOV54l, CG56146-06 DNA Sequence	SEQ ID NO: 853	952 bp
	ORF Start: at 3	ORF Stop: TAG at 942

CCAATCAGACACAGATGATGGAATTCTTGCTTGTGAGATTTACTGAGAATTGGGTGCTCCTGAGGCTG
CATGCTTTGCTCTTCTCACTGATCTACCTCACGGCTGTGCTGATGAATTTAGTCATCATTCTCCTCAT
GATTCTGGACCATCGTCTCCACATGGCAATGTACTTTTTCCTCCGACATTTGTCCTTCTTAGACCTGT
GTCTCATTTCTGCCACAGTCCCCAAATCCATCCTCAACTCTGTCGCCTCCACTGACTCCATCTCCTTC
CTGGGGTGTGTGTTGCAGCTCTTCTTGGTGGTACTGCTGGCTGGATCAGAGATTGGCATCCTTACTGC
CATGTCCTATGACCGCTATGCTGCCATCTGCTGCCCCCTACACTGTGAGGCTGTCATGAGCAGAGGGC
TCTGTGTCCAGTTGATGGCTCTGTCCTGGCTCAACAGAGGGGCCTTGGGACTCTTGTACACAGCTGGA
ACATTCTCTCTGAATTTTTATGGCTCTGATGAGCTACATCAGTTCTTCTGCGATGTCCCTGCCCTACT
AAAGCTCACTTGTTCTAAAGAACATGCCATCATTAGTGTCAGTGTGGCCATTGGGGTCTGTTATGCAT
TTTCATGTTTAGTTTGCATTGTAGTTTCCTATGTGTACATTTTCTCTGCTGTGTTAAGGATATCACAG
AGACAGAGACAATCCAAAGCCTTTTCCAACTGTGTGCCTCACCTCATTGTTGTCACTGTGTTTCTTGT
AACAGGTGCTGTTGCTTATTTAAAGCCAGGGTCTGATGCACCTTCTATTCTAGACTTGCTGGTGTCTG
TGTTCTATTCTGTCGCACCTCCAACCTTGAACCCTGTTATCTACTGTCTGAAGAACAAGGACATTAAA
TCCGCTCTGAGTAAAGTCCTGTGGAATGTTAGAAGCAGTGGGGTAATGAAAGATGAC**TAG**CAATTGGG

NOV54l, CG56146-06 Protein Sequence	SEQ ID NO: 854	313 aa	MW at 34701.9kD

NQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAMYFFLRHLSFLDLC
LISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMSYDRYAAICCPLHCEAVMSRGL
CVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFCDVPALLKLTCSKEHAIISVSVAIGVCYAF
SCLVCIVVSYVYIFSAVLRISQRQRQSKAFSNCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSV
FYSVAPPTLNPVIYCLKNKDIKSALSKVLWNVRSSGVMKDD

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 54B.

Table 54B. Comparison of the NOV54 protein sequences.

```
NOV54a  ----------------------------------------------------MILDHRLHMAM
NOV54b  ---MTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAM
NOV54c  ----------------------------------------------------GSMILDHRLHMAM
NOV54d  ----------------------------------------------------GSMILDHRLHMAM
NOV54e  ----------------------------------------------------GSMILDHRLHMAM
NOV54f  GSTMTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAM
NOV54g  ------------------------------------------------------------
NOV54h  ---MTNQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAM
NOV54i  -----NQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAM
NOV54j  ------------------------------------------------------------
NOV54k  ------------------------------------------------------------
NOV54l  -----NQTQMMEFLLVRFTENWVLLRLHALLFSLIYLTAVLMNLVIILLMILDHRLHMAM
```

```
NOV54a  YFFLRHLSFLDLCLISDTVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMS
NOV54b  YFFLRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMS
NOV54c  YFFLRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMS
NOV54d  YFFLRHLSFLDLCLISAAVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMS
NOV54e  YFFLRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLPAGSEIGILTAMS
NOV54f  YFFLRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGIPTAMS
NOV54g  ------------------------------------------------------------
NOV54h  YFFLRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMS
NOV54i  YFFLRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMS
NOV54j  ------------------------------------------------------------
NOV54k  ------------------------------------------------------------
NOV54l  YFFLRHLSFLDLCLISATVPKSILNSVASTDSISFLGCVLQLFLVVLLAGSEIGILTAMS

NOV54a  YDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFC
NOV54b  YDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFC
NOV54c  YDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFC
NOV54d  YDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFC
NOV54e  YDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFC
NOV54f  YDRYAAICCPLHCKAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFC
NOV54g  ------WLCPGSTEGPWDS--CTQLEHS----------------LLNFYGSDELHQFFC
NOV54h  YDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFC
NOV54i  YDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFC
NOV54j  ------------------------------------------------MMEFLLVRFT
NOV54k  ------------------------------------------------------------
NOV54l  YDRYAAICCPLHCEAVMSRGLCVQLMALSWLNRGALGLLYTAGTFSLNFYGSDELHQFFC

NOV54a  DVPALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFS
NOV54b  DVPALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFS
NOV54c  DVPALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFS
NOV54d  DVPALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFS
NOV54e  DVPALLKLTCSKEHAIISVSGAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFS
NOV54f  DVPALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFS
NOV54g  DVPALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFS
NOV54h  DVPALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFS
NOV54i  DVPALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFS
NOV54j  ENWVLLRLHA----------------------------------------------------
NOV54k  DVPALLKLTCSKEHAIISVS----------------------------------------
NOV54l  DVPALLKLTCSKEHAIISVSVAIGVCYAFSCLVCIVVSYVYIFSAVLRISQRQRQSKAFS

NOV54a  NCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKS
NOV54b  NCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKS
NOV54c  NCVPHLIVVTVFLVTGAVAYLKPGSDTPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKS
NOV54d  NCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKS
NOV54e  NCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKS
NOV54f  NCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKS
NOV54g  NCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKS
NOV54h  NCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKS
NOV54i  NCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKS
NOV54j  ------------------------------------------------------------
NOV54k  ------------------------------------------------------------
NOV54l  NCVPHLIVVTVFLVTGAVAYLKPGSDAPSILDLLVSVFYSVAPPTLNPVIYCLKNKDIKS

NOV54a  ALSKVLWNVRSSGVMKDD--
```

```
NOV54b  ALSKVLWNVRSSGVMKDD--
NOV54c  ALSKVLWNVRSSGVMKDDLE
NOV54d  ALSKVLWNVRSSGVMKDDLE
NOV54e  ALSKVLWNVRSSGVMKDDLE
NOV54f  ALSKVLWNVRSSGVMKDDLE
NOV54g  ALSKVLWNVRSSGVMKDDLE
NOV54h  ALSKVLWNVRSSGVMKDD--
NOV54i  ALSKVLWNVRSSGVMKDD--
NOV54j  ----------------------
NOV54k  ----------------------
NOV54l  ALSKVLWNVRSSGVMKDD--

NOV54a  (SEQ ID NO:  832)
NOV54b  (SEQ ID NO:  834)
NOV54c  (SEQ ID NO:  836)
NOV54d  (SEQ ID NO:  838)
NOV54e  (SEQ ID NO:  840)
NOV54f  (SEQ ID NO:  842)
NOV54g  (SEQ ID NO:  844)
NOV54h  (SEQ ID NO:  846)
NOV54i  (SEQ ID NO:  848)
NOV54j  (SEQ ID NO:  850)
NOV54k  (SEQ ID NO:  852)
NOV54l  (SEQ ID NO:  854)
```

Further analysis of the NOV54a protein yielded the following properties shown in Table 54C.

Table 54C. Protein Sequence Properties NOV54a

SignalP analysis:	Cleavage site between residues 70 and 71

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 6;  pos.chg 1;  neg.chg 1
      H-region:  length 9;  peak value   7.79
      PSG score:   3.39

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -2.72
      possible cleavage site: between 59 and 60

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   4
      INTEGRAL    Likelihood = -9.66   Transmembrane   44 -  60
      INTEGRAL    Likelihood = -9.71   Transmembrane  152 - 168
      INTEGRAL    Likelihood = -5.79   Transmembrane  193 - 209
      INTEGRAL    Likelihood =  0.16   Transmembrane  221 - 237
      PERIPHERAL  Likelihood =  1.11 (at 10)
      ALOM score:  -9.71  (number of TMSs: 4)
```

```
MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 51
      Charge difference:  0.0   C(-1.0) - N(-1.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       2       Hyd Moment(75):  6.56
      Hyd Moment(95):  1.99    G content:       0
      D/E content:     2       S/T content:     1
      Score: -5.34

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 26  LRH|LS

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.1%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: VMKD

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none
```

```
checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       33.3 %: mitochondrial

>> prediction for CG56146-02 is end (k=9)
```

A search of the NOV54a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 54D.

5

Table 54D. Geneseq Results for NOV54a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV54a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU11100	Human novel G protein-coupled receptor, NOV7 - Homo sapiens, 315 aa. [WO200177177-A2, 18-OCT-2001]	1..269 47..315	268/269 (99%) 268/269 (99%)	e-152
AAU85251	G-coupled olfactory receptor #112 - Homo sapiens, 315 aa. [WO200198526-A2, 27-DEC-2001]	1..269 47..315	268/269 (99%) 268/269 (99%)	e-152
AAU95586	Human olfactory and pheromone G protein-coupled receptor #73 - Homo sapiens, 315 aa. [WO200224726-A2, 28-MAR-2002]	1..269 47..315	268/269 (99%) 268/269 (99%)	e-152
AAU24631	Human olfactory receptor AOLFR125 - Homo sapiens, 315 aa. [WO200168805-A2, 20-SEP-2001]	1..269 47..315	268/269 (99%) 268/269 (99%)	e-152
ABP95851	Human GPCR polypeptide SEQ ID NO 512 - Homo sapiens, 314 aa. [WO200216548-A2, 28-FEB-2002]	1..267 47..313	266/267 (99%) 266/267 (99%)	e-151

In a BLAST search of public sequence databases, the NOV54a protein was found to have homology to the proteins shown in the BLASTP data in Table 54E.

Table 54E. Public BLASTP Results for NOV54a				
Protein Accession Number	**Protein/Organism/Length**	**NOV54a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8NGZ2	Seven transmembrane helix receptor - Homo sapiens (Human), 315 aa.	1..269 47..315	268/269 (99%) 268/269 (99%)	e-152
Q96R53	Olfactory receptor - Homo sapiens (Human), 217 aa (fragment).	20..236 1..217	216/217 (99%) 216/217 (99%)	e-120
CAD35484	Sequence 23 from Patent WO0208289 - Homo sapiens (Human), 314 aa.	3..257 49..303	134/255 (52%) 181/255 (70%)	6e-77
Q8NHC5	Seven transmembrane helix receptor - Homo sapiens (Human), 309 aa.	3..257 49..302	138/259 (53%) 192/259 (73%)	2e-76
CAD35483	Sequence 21 from Patent WO0208289 - Homo sapiens (Human), 314 aa.	3..257 49..303	134/255 (52%) 180/255 (70%)	3e-76

PFam analysis indicates that the NOV54a protein contains the domains shown in the Table 54F.

Table 54F. Domain Analysis of NOV54a			
Pfam Domain	**NOV54a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	1..242	53/277 (19%) 162/277 (58%)	8.5e-07

Example 55.

The NOV55 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 55A.

Table 55A. NOV55 Sequence Analysis		
NOV55a, CG56258-04 DNA Sequence	SEQ ID NO: 855	2828 bp
	ORF Start: ATG at 63	ORF Stop: TAA at 2826
GTCTCTGGCCTATCAGGAGGACAACTGGTGCTGCAATAGAAGCCAGTGGCTAAGTCTCGTGTATGGCG TGGTTAAGGTTGCAGCCTCTCACCTCTGCCTTCCTCCATTTTGGGCTGGTTACCTTTGTGCTCTTCCT GAATGGTCTTCGAGCAGAGGCTGGTGGCTCAGGGGACGTGCCAAGCACAGGGCAGAACAATGAGTCCT		

GTTCAGGGTCATCGGACTGCAAGGAGGGTGTCATCCTGCCAATCTGGTACCCGGAGAACCCTTCCCTT
GGGGACAAGATTGCCAGGGTCATTGTCTATTTTGTGGCCCTGATATACATGTTCCTTGGGGTGTCCAT
CATTGCTGACCGCTTCATGGCATCTATTGAAGTCATCACCTCTCAAGAGAGGGAGGTGACAATTAAGA
AACCCAATGGAGAAACCAGCACAACCACTATTCGGGTCTGGAATGAAACTGTCTCCAACCTGACCCTT
ATGGCCCTGGGTTCCTCTGCTCCTGAGATACTCCTCTCTTTAATTGAGGTGTGTGGTCATGGGTTCAT
TGCTGGTGATCTGGGACCTTCTACCATTGTAGGGAGTGCAGCCTTCAACATGTTCATCATCATTGGCA
TCTGTGTCTACGTGATCCCAGACGGAGAGACTCGCAAGATCAAACATCTACGAGTCTTCTTCATCACC
GCTGCTTGGAGTATCTTTGCCTACATCTGGCTCTATATGATTCTGGCAGTCTTCTCCCCTGGTGTGGT
CCAGGTTTGGGAAGGCCTCCTCACTCTCTTCTTCTTTCCAGTGTGTGTCCTTCTGGCCTGGGTGGCAG
ATAAACGACTGCTCTTCTACAAATACATGCACAAAAAGTACCGCACAGACAAACACCGAGGAATTATC
ATAGAGACAGAGGGTGACCACCCTAAGGGCATTGAGATGGATGGGAAAATGATGAATTCCCATTTTCT
AGATGGGAACCTGGTGCCCCTGGAAGGGAAGGAAGTGGATGAGTCCCGCAGAGAGATGATCCGGATTC
TCAAGGATCTGAAGCAAAAACACCCAGAGAAGGACTTAGATCAGCTGGTGGAGATGGCCAATTACTAT
GCTCTTTCCCACCAACAGAAGAGCCGCGCCTTCTACCGTATCCAAGCCACTCGTATGATGACTGGTGC
AGGCAATATCCTGAAGAAACATGCAGCAGAACAAGCCAAGAAGGCCTCCAGCATGAGCGAGGTGCACA
CCGATGAGCCTGAGGACTTTATTTCCAAGGTCTTCTTTGACCCATGTTCTTACCAGTGCCTGGAGAAC
TGTGGGGCTGTACTCCTGACAGTGGTGAGGAAAGGGGGAGACATGTCAAAGACCATGTATGTGGACTA
CAAAACAGAGGATGGTTCTGCCAATGCAGGGGCTGACTATGAGTTCACAGAGGGCACGGTGGTTCTGA
AGCCAGGAGAGACCCAGAAGGAGTTCTCCGTGGGCATAATTGATGACGACATTTTTGAGGAGGATGAA
CACTTCTTTGTAAGGTTGAGCAATGTCCGCATAGAGGAGGAGCAGCCAGAGGAGGGGATGCCTCCAGC
AATATTCAACAGTCTTCCCTTGCCTCGGGCTGTCCTAGCCTCCCCTTGTGTGGCCACAGTTACCATCT
TGGATGATGACCATGCAGGCATCTTCACTTTTGAATGTGATACTATTCATGTCAGTGAGAGTATTGGT
GTTATGGAGGTCAAGGTTCTGCGGACATCAGGTGCCCGGGGTACAGTCATCGTCCCCTTTAGGACAGT
AGAAGGGACAGCCAAGGGTGGCGGTGAGGACTTTGAAGACACATATGGGGAGTTGGAATTCAAGAATG
ATGAAACTGTGAAAACCATAAGGGTTAAAATAGTAGATGAGGAGGAATACGAAAGGCAAGAGAATTTC
TTCATTGCCCTTGGTGAACCGAAATGGATGGAACGTGGAATATCAGATGTGACAGACAGGAAGCTGAC
TATGGAAGAAGAGGAGGCCAAGAGGATAGCAGAGATGGGAAAGCCAGTATTGGGTGAACACCCCAAAC
TAGAAGTCATCATTGAAGAGTCCTATGAGTTCAAGACTACGGTGGACAAACTGATCAAGAAGACAAAC
CTGGCCTTGGTTGTGGGGACCCATTCCTGGAGGGACCAGTTCATGGAGGCCATCACCGTCAGTGCAGC
AGGGGATGAGGATGAGGATGAATCCGGGGAGGAGAGGCTGCCCTCCTGCTTTGACTACGTCATGCACT
TCCTGACTGTCTTCTGGAAGGTGCTGTTTGCCTGTGTGCCCCCCACAGAGTACTGCCACGGCTGGGCC
TGCTTCGCCGTCTCCATCCTCATCATTGGCATGCTCACCGCCATCATTGGGGACCTGGCCTCGCACTT
CGGCTGCACCATTGGTCTCAAAGATTCAGTCACAGCTGTTGTTTTCGTGGCATTTGGCACCTCTGTCC
CAGATACGTTTGCCAGCAAAGCTGCTGCCCTCCAGGATGTATATGCAGACGCCTCCATTGGCAACGTG
ACGGGCAGCAACGCCGTCAATGTCTTCCTGGGCATCGGCCTGGCCTGGTCCGTGGCCGCCATCTACTG
GGCTCTGCAGGGACAGGAGTTCCACGTGTCGGCCGGCACACTGGCCTTCTCCGTCACCCTCTTCACCA
TCTTTGCATTTGTCTGCATCAGCGTGCTCTTGTACCGAAGGCGGCCGCACCTGGGAGGGGAGCTTGGT
GGCCCCCGTGGCTGCAAGCTCGCCACAACATGGCTCTTTGTGAGCCTGTGGCTCCTCTACATACTCTT
TGCCACACTAGAGGCCTATTGCTACATCAAGGGGTTC**TAA**

NOV55a, CG56258-04 Protein Sequence	SEQ ID NO: 856	921 aa	MW at 102413.7kD

MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGVILPIWYPENP
SLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGETSTTTIRVWNETVSNL
TLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFIIIGICVYVIPDGETRKIKHLRVFF
ITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRG
IIIETEGDHPKGIEMDGKMMNSHFLDGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMAN
YYALSHQQKSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCL
ENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEE
DEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIHVSES
IGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKTIRVKIVDEEEYERQE
NFFIALGEPKWMERGISDVTDRKLTMEEEEAKRIAEMGKPVLGEHPKLEVIIEESYEFKTTVDKLIKK
TNLALVVGTHSWRDQFMEAITVSAAGDEDEDESGEERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHG
WACFAVSILIIGMLTAIIGDLASHFGCTIGLKDSVTAVVFVAFGTSVPDTFASKAAALQDVYADASIG
NVTGSNAVNVFLGIGLAWSVAAIYWALQGQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHLGGE
LGGPRGCKLATTWLFVSLWLLYILFATLEAYCYIKGF

NOV55b, CG56258-02 DNA Sequence	SEQ ID NO: 857	2840 bp
	ORF Start: ATG at 63	ORF Stop: TAA at 2838

GTCTCTGGCCTATCAGGAGGACAACTGGTGCTGCAATAGAAGCCAGTGGCTAAGTCTCGTGTATGGCG
TGGTTAAGGTTGCAGCCTCTCACCTCTGCCTTCCTCCATTTTGGGCTGGTTACCTTTGTGCTCTTCCT
GAATGGTCTTCGAGCAGAGGCTGGTGGCTCAGGGGACGTGCCAAGCACAGGGCAGAACAATGAGTCCT
GTTCAGGGTCATCGGACTGCAAGGAGGGTGTCATCCTGCCAATCTGGTACCCGGAGAACCCTTCCCTT
GGGGACAAGATTGCCAGGGTCATTGTCTATTTTGTGGCCCTGATATACATGTTCCTTGGGGTGTCCAT
CATTGCTGACCGCTTCATGGCATCTATTGAAGTCATCACCTCTCAAGAGAGGGAGGTGACAATTAAGA
AACCCAATGGAGAAACCAGCACAACCACTATTCGGGTCTGGAATGAAACTGTCTCCAACCTGACCCTT
ATGGCCCTGGGTTCCTCTGCTCCTGAGATACTCCTCTCTTTAATTGAGGTGTGTGGTCATGGGTTCAT
TGCTGGTGATCTGGGACCTTCTACCATTGTAGGGAGTGCAGCCTTCAACATGTTCATCATCATTGGCA
TCTGTGTCTACGTGATCCCAGACGGAGAGACTCGCAAGATCAAGCATCTACGAGTCTTCTTCATCACC
GCTGCTTGGAGTATCTTTGCCTACATCTGGCTCTATATGATTCTGGCAGTCTTCTCCCCTGGTGTGGT
CCAGGTTTGGGAAGGCCTCCTCACTCTCTTCTTCTTTCCAGTGTGTGTCCTTCTGGCCTGGGTGGCAG
ATAAACGACTGCTCTTCTACAAATACATGCACAAAAAGTACCGCACAGACAAACACCGAGGAATTATC
ATAGAGACAGAGGGTGACCACCCTAAGGGCATTGAGATGGATGGGAAAATGATGAATTCCCATTTTCT
AGATGGGAACCTGGTGCCCCTGGAAGGGAAGGAAGTGGATGAGTCCCGCAGAGAGATGATCCGGATTC
TCAAGGATCTGAAGCAAAAACACCCAGAGAAGGACTTAGATCAGCTGGTGGAGATGGCCAATTACTAT
GCTCTTTCCCACCAACAGAAGAGCCGTGCCTTCTACCGTATCCAAGCCACTCGTATGATGACTGGTGC
AGGCAATATCCTGAAGAAACATGCAGCAGAACAAGCCAAGAAGGCCTCCAGCATGAGCGAGGTGCACA
CCGATGAGCCTGAGGACTTTATTTCCAAGGTCTTCTTTGACCCATGTTCTTACCAGTGCCTGGAGAAC
TGTGGGGCTGTACTCCTGACAGTGGTGAGGAAAGGGGGAGACATGTCAAAGACCATGTATGTGGACTA
CAAAACAGAGGATGGTTCTGCCAATGCAGGGGCTGACTATGAGTTCACAGAGGGCACGGTGGTTCTGA
AGCCAGGAGAGACCCAGAAGGAGTTCTCCGTGGGCATAATTGATGACGACATTTTTGAGGAGGATGAA
CACTTCTTTGTAAGGTTGAGCAATGTCCGCATAGAGGAGGAGCAGCCAGAGGAGGGGATGCCTCCAGC
AATATTCAACAGTCTTCCCTTGCCTCGGGCTGTCCTAGCCTCCCCTTGTGTGGCCACAGTTACCATCT
TGGATGATGACCATGCAGGCATCTTCACTTTTGAATGTGATACTATTCATGTCAGTGAGAGTATTGGT
GTTATGGAGGTCAAGGTTCTGCGGACATCAGGTGCCCGGGGTACAGTCATCGTCCCCTTTAGGACAGT
AGAAGGGACAGCCAAGGGTGGCGGTGAGGACTTTGAAGACACATATGGGGAGTTGGAATTCAAGAATG
ATGAAACTGTCAAAACAATTCACATCAAGGTAATTGATGATGAGGCATATGAGAAAAACAAGAATTAC
TTCATTGAGATGATGGGCCCCCGCATGGTGGATATGAGTTTTCAGAAAGCGCTCCTGTTATCTCCAGA
CAGGAAGCTGACTATGGAAGAAGAGGAGGCCAAGAGGATAGCAGAGATGGGAAAGCCAGTATTGGGTG
AACACCCCAAACTAGAAGTCATCATTGAAGAGTCCTATGAGTTCAAGACTACGGTGGACAAACTGATC
AAGAAGACAAACCTGGCCTTGGTTGTGGGGACCCATTCCTGGAGGGACCAGTTCATGGAGGCCATCAC
CGTCAGTGCAGCAGGGGATGAGGATGAGGATGAATCCGGGGAGGAGAGGCTGCCCTCCTGCTTTGACT
ACGTCATGCACTTCCTGACTGTCTTCTGGAAGGTGCTGTTTGCCTGTGTGCCCCCCACAGAGTACTGC
CACGGCTGGGCCTGCTTCGCCGTCTCCATCCTCATCATTGGCATGCTCACCGCCATCATTGGGGACCT
GGCCTCGCACTTCGGCTGCACCATTGGTCTCAAAGATTCAGTCACAGCTGTTGTTTTCGTGGCATTTG
GCACCTCTGTCCCAGATACGTTTGCCAGCAAAGCTGCTGCCCTCCAGGATGTATATGCAGACGCCTCC
ATTGGCAACGTGACGGGCAGCAACGCCGTCAATGTCTTCCTGGGCATCGGCCTGGCCTGGTCCGTGGC
CGCCATCTACTGGGCTCTGCAGGGACAGGAGTTCCACGTGTCGGCCGGCACACTGGCCTTCTCCGTCA
CCCTCTTCACCATCTTTGCATTTGTCTGCATCAGCGTGCTCTTGTACCGAAGGCGGCCGCACCTGGGA
GGGGAGCTTGGTGGCCCCCGTGGCTGCAAGCTCGCCACAACATGGCTCTTTGTGAGCCTGTGGCTCCT
CTACATACTCTTTGCCACACTAGAGGCCTATTGCTACATCAAGGGGTTCTAA

NOV55b, CG56258-02 Protein Sequence	SEQ ID NO: 858	925 aa	MW at 102802.3kD

MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGVILPIWYPENP
SLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGETSTTTIRVWNETVSNL
TLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFIIIGICVYVIPDGETRKIKHLRVFF
ITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRG
IIIETEGDHPKGIEMDGKMMNSHFLDGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMAN
YYALSHQQKSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCL
ENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEE
DEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIHVSES
IGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKTIHIKVIDDEAYEKNK
NYFIEMMGPRMVDMSFQKALLLSPDRKLTMEEEEAKRIAEMGKPVLGEHPKLEVIIEESYEFKTTVDK

LIKKTNLALVVGTHSWRDQFMEAITVSAAGDEDEDESGEERLPSCFDYVMHFLTVFWKVLFACVPPTE
YCHGWACFAVSILIIGMLTAIIGDLASHFGCTIGLKDSVTAVVFVAFGTSVPDTFASKAAALQDVYAD
ASIGNVTGSNAVNVFLGIGLAWSVAAIYWALQGQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPH
LGGELGGPRGCKLATTWLFVSLWLLYILFATLEAYCYIKGF

NOV55c, 258076220 DNA Sequence	SEQ ID NO: 859	2778 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCACCATGGCGTGGTTAAGGTTGCAGCCTCTCACCTCTGCCTTCCTCCATTTTGGGCTGGTTAC
CTTTGTGCTCTTCCTGAATGGTCTTCGAGCAGAGGCTGGTGGCTCAGGGGACGTGCCAAGCACAGGGC
AGAACAATGAGTCCTGTTCAGGGTCATCGGACTGCAAGGAGGGTGTCATCCTGCCAATCTGGTACCCG
GAGAACCCTTCCCTTGGGGACAAGATTGCCAGGGTCATTGTCTATTTTGTGGCCCTGATATACATGTT
CCTTGGGGTGTCCATCATTGCTGACCGCTTCATGGCATCTATTGAAGTCATCACCTCTCAAGAGAGGG
AGGTGACAATTAAGAAACCCAATGGAGAAACCAGCACAACCACTATTCGGGTCTGGAATGAAACTGTC
TCCAACCTGACCCTTATGGCCCTGGGTTCCTCTGCTCCTGAGATACTCCTCTCTTTAATTGAGGTGTG
TGGTCATGGGTTCATTGCTGGTGATCTGGGACCTTCTACCATTGTAGGGAGTGCAGCCTTCAACATGT
TCATCATCATTGGCATCTGTGTCTACGTGATCCCAGACGGAGAGACTCGCAAGATCAAGCATCTACGA
GTCTTCTTCATCACCGCTGCTTGGAGTATCTTTGCCTACATCTGGCTCTATATGATTCTGGCAGTCTT
CTCCCCTGGTGTGGTCCAGGTTTGGGAAGGCCTCCTCACTCTCTTCTTCTTTCCAGTGTGTGTCCTTC
TGGCCTGGGTGGCAGATAAACGACTGCTCTTCTACAAATACATGCACAAAAAGTACCGCACAGACAAA
CACCGAGGAATTATCATAGAGACAGAGGGTGACCACCCTAAGGGCATTGAGATGGATGGGAAAATGAT
GAATTCCCATTTTCTAGATGGGAACCTGGTGCCCCTGGAAGGGAAGGAAGTGGATGAGTCCCGCAGAG
AGATGATCCGGATTCTCAAGGATCTGAAGCAAAAACACCCAGAGAAGGACTTAGATCAGCTGGTGGAG
ATGGCCAATTACTATGCTCTTTCCCACCAACAGAAGAGCCGCGCCTTCTACCGTATCCAAGCCACTCG
TATGATGACTGGTGCAGGCAATATCCTGAAGAAACATGCAGCAGAACAAGCCAAGAAGGCCTCCAGCA
TGAGCGAGGTGCACACCGATGAGCCTGAGGACTTTATTTCCAAGGTCTTCTTTGACCCATGTTCTTAC
CAGTGCCTGGAGAACTGTGGGGCTGTACTCCTGACAGTGGTGAGGAAAGGGGGAGACATGTCAAAGAC
CATGTATGTGGACTACAAAACAGAGGATGGTTCTGCCAATGCAGGGGCTGACTATGAGTTCACAGAGG
GCACGGTGGTTCTGAAGCCAGGAGAGACCCAGAAGGAGTTCTCCGTGGGCATAATTGATGACGACATT
TTTGAGGAGGATGAACACTTCTTTGTAAGGTTGAGCAATGTCCGCATAGAGGAGGAGCAGCCAGAGGA
GGGGATGCCTCCAGCAATATTCAACAGTCTTCCCTTGCCTCGGGCTGTCCTAGCCTCCCCTTGTGTGG
CCACAGTTACCATCTTGGATGATGACCATGCAGGCATCTTCACTTTTGAATGTGATACTATTCATGTC
AGTGAGAGTATTGGTGTTATGGAGGTCAAGGTTCTGCGGACATCAGGTGCCCGGGGTACAGTCATCGT
CCCCTTTAGGACAGTAGAAGGGACAGCCAAGGGTGGCGGTGAGGACTTTGAAGACACATATGGGGAGT
TGGAATTCAAGAATGATGAAACTGTGAAAACCATAAGGGTTAAAATAGTAGATGAGGAGGAATACGAA
AGGCAAGAGAATTTCTTCATTGCCCTTGGTGAACCGAAATGGATGGAACGTGGAATATCAGATGTGAC
AGACAGGAAGCTGACTATGGAAGAAGAGGAGGCCAAGAGGATAGCAGAGATGGGAAAGCCAGTATTGG
GTGAACACCCCAAACTAGAAGTCATCATTGAAGAGTCCTATGAGTTCAAGACTACGGTGGACAAACTG
ATCAAGAAGACAAACCTGGCCTTGGTTGTGGGGACCCATTCCTGGAGGGACCAGTTCATGGAGGCCAT
CACCGTCAGTGCAGCAGGGGATGAGGATGAGGATGAATCCGGGGAGGAGAGGCTGCCCTCCTGCTTTG
ACTACGTCATGCACTTCCTGACTGTCTTCTGGAAGGTGCTGTTTGCCTGTGTGCCCCCCACAGAGTAC
TGCCACGGCTGGGCCTGCTTCGCCGTCTCCATCCTCATCATTGGCATGCTCACCGCCATCATTGGGGA
CCTGGCCTCGCACTTCGGCTGCACCATTGGTCTCAAAGATTCAGTCACAGCTGTTGTTTTCGTGGCAT
TTGGCACCTCTGTCCCAGATACGTTTGCCAGCAAAGCTGCTGCCCTCCAGGATGTATATGCAGACGCC
TCCATTGGCAACGTGACGGGCAGCAACGCCGTCAATGTCTTCCTGGGCATCGGCCTGGCCTGGTCCGT
GGCCGCCATCTACTGGGCTCTGCAGGGACAGGAGTTCCACGTGTCGGCCGGCACACTGGCCTTCTCCG
TCACCCTCTTCACCATCTTTGCATTTGTCTGCATCAGCGTGCTCTTGTACCGAAGGCGGCCGCACCTG
GGAGGGGAGCTTGGTGGCCCCCGTGGCTGCAAGCTCGCCACAACATGGCTCTTTGTGAGCCTGTGGCT
CCTCTACATACTCTTTGCCACACTAGAGGCCTATTGCTACATCAAGGGGTTCCTCGAG

NOV55c, 258076220 Protein Sequence	SEQ ID NO: 860	926 aa	MW at 102901.2kD

GSTMAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGVILPIWYP
ENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGETSTTTIRVWNETV
SNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFIIIGICVYVIPDGETRKIKHLR
VFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFFFPVCVLLAWVADKRLLFYKYMHKKYRTDK
HRGIIIETEGDHPKGIEMDGKMMNSHFLDGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVE
MANYYALSHQQKSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSY
QCLENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDI
FEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIHV

SESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKTIRVKIVDEEEYE
RQENFFIALGEPKWMERGISDVTDRKLTMEEEEAKRIAEMGKPVLGEHPKLEVIIEESYEFKTTVDKL
IKKTNLALVVGTHSWRDQFMEAITVSAAGDEDEDESGEERLPSCFDYVMHFLTVFWKVLFACVPPTEY
CHGWACFAVSILIIGMLTAIIGDLASHFGCTIGLKDSVTAVVFVAFGTSVPDTFASKAAALQDVYADA
SIGNVTGSNAVNVFLGIGLAWSVAAIYWALQGQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHL
GGELGGPRGCKLATTWLFVSLWLLYILFATLEAYCYIKGFLE

NOV55d, 248057963 DNA Sequence	SEQ ID NO: 861	2685 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCGAGGCTGGTGGCTCAGGGGACGTGCCAAGCACAGGGCAGAACAATGAGTCCTGTTCAGGGTC
ATCGGACTGCAAGGAGGGTGTCATCCTGCCAATCTGGTACCCGGAGAACCCTTCCCTTGGGGACAAGA
TTGCCAGGGTCATTGTCTATTTTGTGGCCCTGATATACATGTTCCTTGGGGTGTCCATCATTGCTGAC
CGCTTCATGGCATCTATTGAAGTCATCACCTCTCAAGAGAGGGAGGTGACAATTAAGAAACCCAATGG
AGAAACCAGCACAACCACTATTCGGGTCTGGAATGAAACTGTCTCCAACCTGACCCTTATGGCCCTGG
GTTCCTCTGCTCCTGAGATACTCCTCTCTTTAATTGAGGTGTGTGGTCATGGGTTCATTGCTGGTGAT
CTGGGACCTTCTACCATTGTAGGGAGTGCAGCCTTCAACATGTTCATCATCATTGGCATCTGTGTCTA
CGTGATCCCAGACGGAGAGACTCGCAAGATCAAACATCTACGAGTCTTCTTCATCACCGCTGCTTGGA
GTATCTTTGCCTACATCTGGCTCTATATGATTCTGGCAGTCTTCTCCCCTGGTGTGGTCCAGGTTTGG
GAAGGCCTCCTCACTCTCTTCTTCTTTCCAGTGTGTGTCCTTCTGGCCTGGGTGGCAGATAAACGACT
GCTCTTCTACAAATACATGCACAAAAAGTACCGCACAGACAAACACCGAGGAATTATCATAGAGACAG
AGGGTGACCACCCTAAGGGCATTGAGATGGATGGGAAAATGATGAATTCCCATTTTCTAGATGGGAAC
CTGGTGCCCCTGGAAGGGAAGGAAGTGGATGAGTCCCGCAGAGAGATGATCCGGATTCTCAAGGATCT
GAAGCAAAAACACCCAGAGAAGGACTTAGATCAGCTGGTGGAGATGGCCAATTACTATGCTCTTTCCC
ACCAACAGAAGAGCCGCGCCTTCTACCGTATCCAAGCCACTCGTATGATGACTGGTGCAGGCAATATC
CTGAAGAAACATGCAGCAGAACAAGCCAAGAAGGCCTCCAGCATGAGCGAGGTGCACACCGATGAGCC
TGAGGACTTTATTTCCAAGGTCTTCTTTGACCCATGTTCTTACCAGTGCCTGGAGAACTGTGGGGCTG
TACTCCTGACAGTGGTGAGGAAAGGGGGAGACATGTCAAAGACCATGTATGTGGACTACAAAACAGAG
GATGGTTCTGCCAATGCAGGGGCTGACTATGAGTTCACAGAGGGCACGGTGGTTCTGAAGCCAGGAGA
GACCCAGAAGGAGTTCTCCGTGGGCATAATTGATGACGACATTTTTGAGGAGGATGAACACTTCTTTG
TAAGGTTGAGCAATGTCCGCATAGAGGAGGAGCAGCCAGAGGAGGGGATGCCTCCAGCAATATTCAAC
AGTCTTCCCTTGCCTCGGGCTGTCCTAGCCTCCCCTTGTGTGGCCACAGTTACCATCTTGGATGATGA
CCATGCAGGCATCTTCACTTTTGAATGTGATACTATTCATGTCAGTGAGAGTATTGGTGTTATGGAGG
TCAAGGTTCTGCGGACATCAGGTGCCCGGGGTACAGTCATCGTCCCCTTTAGGACAGTAGAAGGGACA
GCCAAGGGTGGCGGTGAGGACTTTGAAGACACATATGGGGAGTTGGAATTCAAGAATGATGAAACTGT
GAAAACCATAAGGGTTAAAATAGTAGATGAGGAGGAATACGAAAGGCAAGAGAATTTCTTCATTGCCC
TTGGTGAACCGAAATGGATGGAACGTGGAATATCAGATGTGACAGACAGGAAGCTGACTATGGAAGAA
GAGGAGGCCAAGAGGATAGCAGAGATGGGAAAGCCAGTATTGGGTGAACACCCCAAACTAGAAGTCAT
CATTGAAGAGTCCTATGAGTTCAAGACTACGGTGGACAAACTGATCAAGAAGACAAACCTGGCCTTGG
TTGTGGGGACCCATTCCTGGAGGGACCAGTTCATGGAGGCCATCACCGTCAGTGCAGCAGGGGATGAG
GATGAGGATGAATCCGGGGAGGAGAGGCTGCCCTCCTGCTTTGACTACGTCATGCACTTCCTGACTGT
CTTCTGGAAGGTGCTGTTTGCCTGTGTGCCCCCCACAGAGTACTGCCACGGCTGGGCCTGCTTCGCCG
TCTCCATCCTCATCATTGGCATGCTCACCGCCATCATTGGGGACCTGGCCTCGCACTTCGGCTGCACC
ATTGGTCTCAAAGATTCAGTCACAGCTGTTGTTTTCGTGGCATTTGGCACCTCTGTCCCAGATACGTT
TGCCAGCAAAGCTGCTGCCCTCCAGGATGTATATGCAGACGCCTCCATTGGCAACGTGACGGGCAGCA
ACGCCGTCAATGTCTTCCTGGGCATCGGCCTGGCCTGGTCCGTGGCCGCCATCTACTGGGCTCTGCAG
GGACAGGAGTTCCACGTGTCGGCCGGCACACTGGCCTTCTCCGTCACCCTCTTCACCATCTTTGCATT
TGTCTGCATCAGCGTGCTCTTGTACCGAAGGCGGCCGCACCTGGGAGGGGAGCTTGGTGGCCCCCGTG
GCTGCAAGCTCGCCACAACATGGCTCTTTGTGAGCCTGTGGCTCCTCTACATACTCTTTGCCACACTA
GAGGCCTATTGCTACATCAAGGGGTTCCTCGAG

NOV55d, 248057963 Protein Sequence	SEQ ID NO: 862	895 aa	MW at 99385.0kD

GSEAGGSGDVPSTGQNNESCSGSSDCKEGVILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIAD
RFMASIEVITSQEREVTIKKPNGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGD
LGPSTIVGSAAFNMFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVW
EGLLTLFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFLDGN
LVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQATRMMTGAGNI
LKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLTVVRKGGDMSKTMYVDYKTE
DGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFN

SLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGT
AKGGGEDFEDTYGELEFKNDETVKTIRVKIVDEEEYERQENFFIALGEPKWMERGISDVTDRKLTMEE
EEAKRIAEMGKPVLGEHPKLEVIIEESYEFKTTVDKLIKKTNLALVVGTHSWRDQFMEAITVSAAGDE
DEDESGEERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHGWACFAVSILIIGMLTAIIGDLASHFGCT
IGLKDSVTAVVFVAFGTSVPDTFASKAAALQDVYADASIGNVTGSNAVNVFLGIGLAWSVAAIYWALQ
GQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHLGGELGGPRGCKLATTWLFVSLWLLYILFATL
EAYCYIKGFLE

NOV55e, CG56258-01 DNA Sequence	SEQ ID NO: 863	2813 bp
	ORF Start: ATG at 9	ORF Stop: TAG at 2793

TCTCGTGT**ATG**GCGTGGTTAAGGTTGCAGCCTCTCACCTCTGCCTTCCTCCATTTTGGGCTGGTTACC
TTTGTGCTCTTCCTGAATGGTCTTCGAGCAGAGGCTGGTGGCTCAGGGGACGTGCCAAGCACAGGGCA
GAACAATGAGTCCTGTTCAGGGTCATCGGACTGCAAGGAGGGTGTCATCCTGCCAATCTGGTACCCGG
AGAACCCTTCCCTTGGGGACAAGATTGCCAGGGTCATTGTCTATTTTGTGGCCCTGATATACATGTTC
CTTGGGGTGTCCATCATTGCTGACCGCTTCATGGCATCTATTGAAGTCATCACCTCTCAAGAGAGGGA
GGTGACAATTAAGAAACCCAATGGAGAAACCAGCACAACCACTATTCGGGTCTGGAATGAAACTGTCT
CCAACCTGACCCTTATGGCCCTGGGTTCCTCTGCTCCTGAGATACTCCTCTCTTTAATTGAGGTGTGT
GGTCATGGGTTCATTGCTGGTGATCTGGGACCTTCTACCATTGTAGGGAGTGCAGCCTTCAACATGTT
CATCATCATTGGCATCTGTGTCTACGTGATCCCAGACGGAGAGACTCGCAAGATCAAGCATCTACGAG
TCTTCTTCATCACCGCTGCTTGGAGTATCTTTGCCTACATCTGGCTCTATATGATTCTGGCAGTCTTC
TCCCCTGGTGTGGTCCAGGTTTGGGAAGGCCTCCTCACTCTCTTCTTCTTTCCAGTGTGTGTCCTTCT
GGCCTGGGTGGCAGATAAACGACTGCTCTTCTACAAATACATGCACAAAAAGTACCGCACAGACAAAC
ACCGAGGAATTATCATAGAGACAGAGGGTGACCACCCTAAGGGCATTGAGATGGATGGGAAAATGATG
AATTCCCATTTTCTAGATGGGAACCTGGTGCCCCTGGAAGGGAAGGAAGTGGATGAGTCCCGCAGAGA
GATGATCCGGATTCTCAAGGATCTGAAGCAAAAACACCCAGAGAAGGACTTAGATCAGCTGGTGGAGA
TGGCCAATTACTATGCTCTTTCCCACCAACAGAAGAGCCGTGCCTTCTACCGTATCCAAGCCACTCGT
ATGATGACTGGTGCAGGCAATATCCTGAAGAAACATGCAGCAGAACAAGCCAAGAAGGCCTCCAGCAT
GAGCGAGGTGCACACCGATGAGCCTGAGGACTTTATTTCCAAGGTCTTCTTTGACCCATGTTCTTACC
AGTGCCTGGAGAACTGTGGGGCTGTACTCCTGACAGTGGTGAGGAAAGGGGGAGACATGTCAAAGACC
ATGTATGTGGACTACAAAACAGAGGATGGTTCTGCCAATGCAGGGGCTGACTATGAGTTCACAGAGGG
CACGGTGGTTCTGAAGCCAGGAGAGACCCAGAAGGAGTTCTCCGTGGGCATAATTGATGACGACATTT
TTGAGGAGGATGAACACTTCTTTGTAAGGTTGAGCAATGTCCGCATAGAGGAGGAGCAGCCAGAGGAG
GGGATGCCTCCAGCAATATTCAACAGTCTTCCCTTGCCTCGGGCTGTCCTAGCCTCCCCTTGTGTGGC
CACAGTTACCATCTTGGATGATGACCATGCAGGCATCTTCACTTTTGAATGTGATACTATTCATGTCA
GTGAGAGTATTGGTGTTATGGAGGTCAAGGTTCTGCGGACATCAGGTGCCCGGGGTACAGTCATCGTC
CCCTTTAGGACAGTAGAAGGGACAGCCAAGGGTGGCGGTGAGGACTTTGAAGACACATATGGGGAGTT
GGAATTCAAGAATGATGAAACTGTGAAAACTCTTCAGGTGAAGATAGTTGATGACGAGGAATATGAGA
AAAAGGATAATTTCTTCATTGAGCTGGGCCAGCCCCAGTGGCTTAAGCGAGGGATTTCAGCTCTGCTA
CTCAATCAAGGGGATGGGGACAGGAAGCTAACAGCCGAGGAGGAGGAGGCTCGGAGGATAGCAGAGAT
GGGCAAGCCAGTTCTTGGGGAGAACTGCCGGCTGGAGGTCATCATCGAGGAGTCATATGATTTTAAGA
ACACGGTGGATAAACTCATCAAGAAAACGAACTTGGCCTTGGTAATTGGGACCCATTCATGGAGGGAG
CAGTTTTTAGAGGCAATTACGGTGAGCGCAGGGGACGAGGAGGAGGAGGAGGACGGGTCCCGGGAGGA
GCGGCTGCCGTCGTGCTTTGACTACGTGATGCACTTCCTGACGGTGTTCTGGAAGGTGCTCTTCGCCT
GTGTGCCCCCCACCGAGTACTGCCACGGCTGGGCCTGCTTTGGTGTCTCCATCCTGGTCATCGGCCTG
CTCACCGCCCTCATTGGGGACCTCGCCTCCCACTTCGGCTGCACCGTTGGCCTCAAGGACTCTGTCAA
TGCTGTTGTCTTCGTTGCCCTGGGCACCTCCATCCCTGACACGTTCGCCAGCAAGGTGGCGGCGCTGC
AGGACCAGTGCGCCGACGCGTCCATCGGCAACGTGACCGGCTCCAACGCGGTGAACGTGTTCCTTGGC
CTGGGCGTCGCCTGGTCTGTGGCCGCCGTGTACTGGGCGGTGCAGGGCCGCCCCTTCGAGGTGCGCAC
TGGCACGCTGGCCTTCTCCGTCACGCTCTTCACCGTCTTCGCCTTCGTGGGCATTGCCGTGCTGCTGT
ACCGGCGCCGGCCGCACATCGGCGGCGAGCTGGGCGGCCCGCGCGGACCCAAGCTCGCCACCACCGCG
CTCTTCCTGGGCCTCTGGCTCCTGTACATCCTCTTCGCCAGCCTGGAGGCGTACTGCCACATCCGGGG
CTTCT**AG**GGCCTCGCGCAGAGACTC

NOV55e, CG56258-01 Protein Sequence	SEQ ID NO: 864	928 aa	MW at 102900.1kD

MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGVILPIWYPENP
SLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGETSTTTIRVWNETVSNL
TLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFIIIGICVYVIPDGETRKIKHLRVFF
ITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRG

IIIETEGDHPKGIEMDGKMMNSHFLDGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMAN
YYALSHQQKSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCL
ENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEE
DEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIHVSES
IGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKTLQVKIVDDEEYEKKD
NFFIELGQPQWLKRGISALLLNQGDGDRKLTAEEEEARRIAEMGKPVLGENCRLEVIIEESYDFKNTV
DKLIKKTNLALVIGTHSWREQFLEAITVSAGDEEEEEDGSREERLPSCFDYVMHFLTVFWKVLFACVP
PTEYCHGWACFGVSILVIGLLTALIGDLASHFGCTVGLKDSVNAVVFVALGTSIPDTFASKVAALQDQ
CADASIGNVTGSNAVNVFLGLGVAWSVAAVYWAVQGRPFEVRTGTLAFSVTLFTVFAFVGIAVLLYRR
RPHIGGELGGPRGPKLATTALFLGLWLLYILFASLEAYCHIRGF

NOV55f, CG56258-03 DNA Sequence	SEQ ID NO: 865	2685 bp
	ORF Start: at 7	ORF Stop: at 2680

GGATCCGAGGCTGGTGGCTCAGGGGACGTGCCAAGCACAGGGCAGAACAATGAGTCCTGTTCAGGGTC
ATCGGACTGCAAGGAGGGTGTCATCCTGCCAATCTGGTACCCGGAGAACCCTTCCCTTGGGGACAAGA
TTGCCAGGGTCATTGTCTATTTTGTGGCCCTGATATACATGTTCCTTGGGGTGTCCATCATTGCTGAC
CGCTTCATGGCATCTATTGAAGTCATCACCTCTCAAGAGAGGGAGGTGACAATTAAGAAACCCAATGG
AGAAACCAGCACAACCACTATTCGGGTCTGGAATGAAACTGTCTCCAACCTGACCCTTATGGCCCTGG
GTTCCTCTGCTCCTGAGATACTCCTCTCTTTAATTGAGGTGTGTGGTCATGGGTTCATTGCTGGTGAT
CTGGGACCTTCTACCATTGTAGGGAGTGCAGCCTTCAACATGTTCATCATCATTGGCATCTGTGTCTA
CGTGATCCCAGACGGAGAGACTCGCAAGATCAAACATCTACGAGTCTTCTTCATCACCGCTGCTTGGA
GTATCTTTGCCTACATCTGGCTCTATATGATTCTGGCAGTCTTCTCCCCTGGTGTGGTCCAGGTTTGG
GAAGGCCTCCTCACTCTCTTCTTCTTTCCAGTGTGTGTCCTTCTGGCCTGGGTGGCAGATAAACGACT
GCTCTTCTACAAATACATGCACAAAAAGTACCGCACAGACAAACACCGAGGAATTATCATAGAGACAG
AGGGTGACCACCCTAAGGGCATTGAGATGGATGGGAAAATGATGAATTCCCATTTTCTAGATGGGAAC
CTGGTGCCCCTGGAAGGGAAGGAAGTGGATGAGTCCCGCAGAGAGATGATCCGGATTCTCAAGGATCT
GAAGCAAAAACACCCAGAGAAGGACTTAGATCAGCTGGTGGAGATGGCCAATTACTATGCTCTTTCCC
ACCAACAGAAGAGCCGCGCCTTCTACCGTATCCAAGCCACTCGTATGATGACTGGTGCAGGCAATATC
CTGAAGAAACATGCAGCAGAACAAGCCAAGAAGGCCTCCAGCATGAGCGAGGTGCACACCGATGAGCC
TGAGGACTTTATTTCCAAGGTCTTCTTTGACCCATGTTCTTACCAGTGCCTGGAGAACTGTGGGGCTG
TACTCCTGACAGTGGTGAGGAAAGGGGGAGACATGTCAAAGACCATGTATGTGGACTACAAAACAGAG
GATGGTTCTGCCAATGCAGGGGCTGACTATGAGTTCACAGAGGGCACGGTGGTTCTGAAGCCAGGAGA
GACCCAGAAGGAGTTCTCCGTGGGCATAATTGATGACGACATTTTTGAGGAGGATGAACACTTCTTTG
TAAGGTTGAGCAATGTCCGCATAGAGGAGGAGCAGCCAGAGGAGGGGATGCCTCCAGCAATATTCAAC
AGTCTTCCCTTGCCTCGGGCTGTCCTAGCCTCCCCTTGTGTGGCCACAGTTACCATCTTGGATGATGA
CCATGCAGGCATCTTCACTTTTGAATGTGATACTATTCATGTCAGTGAGAGTATTGGTGTTATGGAGG
TCAAGGTTCTGCGGACATCAGGTGCCCGGGGTACAGTCATCGTCCCCTTTAGGACAGTAGAAGGGACA
GCCAAGGGTGGCGGTGAGGACTTTGAAGACACATATGGGGAGTTGGAATTCAAGAATGATGAAACTGT
GAAAACCATAAGGGTTAAAATAGTAGATGAGGAGGAATACGAAAGGCAAGAGAATTTCTTCATTGCCC
TTGGTGAACCGAAATGGATGGAACGTGGAATATCAGATGTGACAGACAGGAAGCTGACTATGGAAGAA
GAGGAGGCCAAGAGGATAGCAGAGATGGGAAAGCCAGTATTGGGTGAACACCCCAAACTAGAAGTCAT
CATTGAAGAGTCCTATGAGTTCAAGACTACGGTGGACAAACTGATCAAGAAGACAAACCTGGCCTTGG
TTGTGGGGACCCATTCCTGGAGGGACCAGTTCATGGAGGCCATCACCGTCAGTGCAGCAGGGGATGAG
GATGAGGATGAATCCGGGGAGGAGAGGCTGCCCTCCTGCTTTGACTACGTCATGCACTTCCTGACTGT
CTTCTGGAAGGTGCTGTTTGCCTGTGTGCCCCCCACAGAGTACTGCCACGGCTGGGCCTGCTTCGCCG
TCTCCATCCTCATCATTGGCATGCTCACCGCCATCATTGGGGACCTGGCCTCGCACTTCGGCTGCACC
ATTGGTCTCAAAGATTCAGTCACAGCTGTTGTTTTCGTGGCATTTGGCACCTCTGTCCCAGATACGTT
TGCCAGCAAAGCTGCTGCCCTCCAGGATGTATATGCAGACGCCTCCATTGGCAACGTGACGGGCAGCA
ACGCCGTCAATGTCTTCCTGGGCATCGGCCTGGCCTGGTCCGTGGCCGCCATCTACTGGGCTCTGCAG
GGACAGGAGTTCCACGTGTCGGCCGGCACACTGGCCTTCTCCGTCACCCTCTTCACCATCTTTGCATT
TGTCTGCATCAGCGTGCTCTTGTACCGAAGGCGGCCGCACCTGGGAGGGGAGCTTGGTGGCCCCCGTG
GCTGCAAGCTCGCCACAACATGGCTCTTTGTGAGCCTGTGGCTCCTCTACATACTCTTTGCCACACTA
GAGGCCTATTGCTACATCAAGGGGTTCCTCGAG

NOV55f, CG56258-03 Protein Sequence	SEQ ID NO: 866	891 aa	MW at 98998.6kD

EAGGSGDVPSTGQNNESCSGSSDCKEGVILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRF
MASIEVITSQEREVTIKKPNGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLG
PSTIVGSAAFNMFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEG

LLTLFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFLDGNLV
PLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQATRMMTGAGNILK
KHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLTVVRKGGDMSKTMYVDYKTEDG
SANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSL
PLPRAVLASPCVATVTILDDDHAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAK
GGGEDFEDTYGELEFKNDETVKTIRVKIVDEEEYERQENFFIALGEPKWMERGISDVTDRKLTMEEEE
AKRIAEMGKPVLGEHPKLEVIIEESYEFKTTVDKLIKKTNLALVVGTHSWRDQFMEAITVSAAGDEDE
DESGEERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHGWACFAVSILIIGMLTAIIGDLASHFGCTIG
LKDSVTAVVFVAFGTSVPDTFASKAAALQDVYADASIGNVTGSNAVNVFLGIGLAWSVAAIYWALQGQ
EFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHLGGELGGPRGCKLATTWLFVSLWLLYILFATLEA
YCYIKGF

NOV55g, CG56258-05 DNA Sequence	SEQ ID NO: 867	2778 bp
	ORF Start: ATG at 10	ORF Stop: at 2773

GGATCCACCATGGCGTGGTTAAGGTTGCAGCCTCTCACCTCTGCCTTCCTCCATTTTGGGCTGGTTAC
CTTTGTGCTCTTCCTGAATGGTCTTCGAGCAGAGGCTGGTGGCTCAGGGGACGTGCCAAGCACAGGGC
AGAACAATGAGTCCTGTTCAGGGTCATCGGACTGCAAGGAGGGTGTCATCCTGCCAATCTGGTACCCG
GAGAACCCTTCCCTTGGGGACAAGATTGCCAGGGTCATTGTCTATTTTGTGGCCCTGATATACATGTT
CCTTGGGGTGTCCATCATTGCTGACCGCTTCATGGCATCTATTGAAGTCATCACCTCTCAAGAGAGGG
AGGTGACAATTAAGAAACCCAATGGAGAAACCAGCACAACCACTATTCGGGTCTGGAATGAAACTGTC
TCCAACCTGACCCTTATGGCCCTGGGTTCCTCTGCTCCTGAGATACTCCTCTCTTTAATTGAGGTGTG
TGGTCATGGGTTCATTGCTGGTGATCTGGGACCTTCTACCATTGTAGGGAGTGCAGCCTTCAACATGT
TCATCATCATTGGCATCTGTGTCTACGTGATCCCAGACGGAGAGACTCGCAAGATCAAGCATCTACGA
GTCTTCTTCATCACCGCTGCTTGGAGTATCTTTGCCTACATCTGGCTCTATATGATTCTGGCAGTCTT
CTCCCCTGGTGTGGTCCAGGTTTGGGAAGGCCTCCTCACTCTCTTCTTCTTTCCAGTGTGTGTCCTTC
TGGCCTGGGTGGCAGATAAACGACTGCTCTTCTACAAATACATGCACAAAAAGTACCGCACAGACAAA
CACCGAGGAATTATCATAGAGACAGAGGGTGACCACCCTAAGGGCATTGAGATGGATGGGAAAATGAT
GAATTCCCATTTTCTAGATGGGAACCTGGTGCCCCTGGAAGGGAAGGAAGTGGATGAGTCCCGCAGAG
AGATGATCCGGATTCTCAAGGATCTGAAGCAAAAACACCCAGAGAAGGACTTAGATCAGCTGGTGGAG
ATGGCCAATTACTATGCTCTTTCCCACCAACAGAAGAGCCGCGCCTTCTACCGTATCCAAGCCACTCG
TATGATGACTGGTGCAGGCAATATCCTGAAGAAACATGCAGCAGAACAAGCCAAGAAGGCCTCCAGCA
TGAGCGAGGTGCACACCGATGAGCCTGAGGACTTTATTTCCAAGGTCTTCTTTGACCCATGTTCTTAC
CAGTGCCTGGAGAACTGTGGGGCTGTACTCCTGACAGTGGTGAGGAAAGGGGGAGACATGTCAAAGAC
CATGTATGTGGACTACAAAACAGAGGATGGTTCTGCCAATGCAGGGGCTGACTATGAGTTCACAGAGG
GCACGGTGGTTCTGAAGCCAGGAGAGACCCAGAAGGAGTTCTCCGTGGGCATAATTGATGACGACATT
TTTGAGGAGGATGAACACTTCTTTGTAAGGTTGAGCAATGTCCGCATAGAGGAGGAGCAGCCAGAGGA
GGGGATGCCTCCAGCAATATTCAACAGTCTTCCCTTGCCTCGGGCTGTCCTAGCCTCCCCTTGTGTGG
CCACAGTTACCATCTTGGATGATGACCATGCAGGCATCTTCACTTTTGAATGTGATACTATTCATGTC
AGTGAGAGTATTGGTGTTATGGAGGTCAAGGTTCTGCGGACATCAGGTGCCCGGGGTACAGTCATCGT
CCCCTTTAGGACAGTAGAAGGGACAGCCAAGGGTGGCGGTGAGGACTTTGAAGACACATATGGGGAGT
TGGAATTCAAGAATGATGAAACTGTGAAAACCATAAGGGTTAAAATAGTAGATGAGGAGGAATACGAA
AGGCAAGAGAATTTCTTCATTGCCCTTGGTGAACCGAAATGGATGGAACGTGGAATATCAGATGTGAC
AGACAGGAAGCTGACTATGGAAGAAGAGGAGGCCAAGAGGATAGCAGAGATGGGAAAGCCAGTATTGG
GTGAACACCCCAAACTAGAAGTCATCATTGAAGAGTCCTATGAGTTCAAGACTACGGTGGACAAACTG
ATCAAGAAGACAAACCTGGCCTTGGTTGTGGGGACCCATTCCTGGAGGGACCAGTTCATGGAGGCCAT
CACCGTCAGTGCAGCAGGGGATGAGGATGAGGATGAATCCGGGGAGGAGAGGCTGCCCTCCTGCTTTG
ACTACGTCATGCACTTCCTGACTGTCTTCTGGAAGGTGCTGTTTGCCTGTGTGCCCCCCACAGAGTAC
TGCCACGGCTGGGCCTGCTTCGCCGTCTCCATCCTCATCATTGGCATGCTCACCGCCATCATTGGGGA
CCTGGCCTCGCACTTCGGCTGCACCATTGGTCTCAAAGATTCAGTCACAGCTGTTGTTTTCGTGGCAT
TTGGCACCTCTGTCCCAGATACGTTTGCCAGCAAAGCTGCTGCCCTCCAGGATGTATATGCAGACGCC
TCCATTGGCAACGTGACGGGCAGCAACGCCGTCAATGTCTTCCTGGGCATCGGCCTGGCCTGGTCCGT
GGCCGCCATCTACTGGGCTCTGCAGGGACAGGAGTTCCACGTGTCGGCCGGCACACTGGCCTTCTCCG
TCACCCTCTTCACCATCTTTGCATTTGTCTGCATCAGCGTGCTCTTGTACCGAAGGCGGCCGCACCTG
GGAGGGGAGCTTGGTGGCCCCCGTGGCTGCAAGCTCGCCACAACATGGCTCTTTGTGAGCCTGTGGCT
CCTCTACATACTCTTTGCCACACTAGAGGCCTATTGCTACATCAAGGGGTTCCTCGAG

NOV55g, CG56258-05 Protein Sequence	SEQ ID NO: 868	921 aa	MW at 102413.7kD

MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGVILPIWYPENP

SLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGETSTTTIRVWNETVSNL
TLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFIIIGICVYVIPDGETRKIKHLRVFF
ITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRG
IIIETEGDHPKGIEMDGKMMNSHFLDGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMAN
YYALSHQQKSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCL
ENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEE
DEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIHVSES
IGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKTIRVKIVDEEEYERQE
NFFIALGEPKWMERGISDVTDRKLTMEEEEAKRIAEMGKPVLGEHPKLEVIIEESYEFKTTVDKLIKK
TNLALVVGTHSWRDQFMEAITVSAAGDEDEDESGEERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHG
WACFAVSILIIGMLTAIIGDLASHFGCTIGLKDSVTAVVFVAFGTSVPDTFASKAAALQDVYADASIG
NVTGSNAVNVFLGIGLAWSVAAIYWALQGQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHLGGE
LGGPRGCKLATTWLFVSLWLLYILFATLEAYCYIKGF

NOV55h, CG56258-06 DNA Sequence	SEQ ID NO: 869	2813 bp
	ORF Start: ATG at 9	ORF Stop: TAG at 2793

TCTCGTGT**ATG**GCGTGGTTAAGGTTGCAGCCTCTCACCTCTGCCTTCCTCCATTTTGGGCTGGTTACC
TTTGTGCTCTTCCTGAATGGTCTTCGAGCAGAGGCTGGTGGCTCAGGGGACGTGCCAAGCACAGGGCA
GAACAATGAGTCCTGTTCAGGGTCATCGGACTGCAAGGAGGGTGTCATCCTGCCAATCTGGTACCCGG
AGAACCCTTCCCTTGGGGACAAGATTGCCAGGGTCATTGTCTATTTTGTGGCCCTGATATACATGTTC
CTTGGGGTGTCCATCATTGCTGACCGCTTCATGGCATCTATTGAAGTCATCACCTCTCAAGAGAGGGA
GGTGACAATTAAGAAACCCAATGGAGAAACCAGCACAACCACTATTCGGGTCTGGAATGAAACTGTCT
CCAACCTGACCCTTATGGCCCTGGGTTCCTCTGCTCCTGAGATACTCCTCTCTTTAATTGAGGTGTGT
GGTCATGGGTTCATTGCTGGTGATCTGGGACCTTCTACCATTGTAGGGAGTGCAGCCTTCAACATGTT
CATCATCATTGGCATCTGTGTCTACGTGATCCCAGACGGAGAGACTCGCAAGATCAAGCATCTACGAG
TCTTCTTCATCACCGCTGCTTGGAGTATCTTTGCCTACATCTGGCTCTATATGATTCTGGCAGTCTTC
TCCCCTGGTGTGGTCCAGGTTTGGGAAGGCCTCCTCACTCTCTTCTTCTTTCCAGTGTGTGTCCTTCT
GGCCTGGGTGGCAGATAAACGACTGCTCTTCTACAAATACATGCACAAAAAGTACCGCACAGACAAAC
ACCGAGGAATTATCATAGAGACAGAGGGTGACCACCCTAAGGGCATTGAGATGGATGGGAAAATGATG
AATTCCCATTTTCTAGATGGGAACCTGGTGCCCCTGGAAGGGAAGGAAGTGGATGAGTCCCGCAGAGA
GATGATCCGGATTCTCAAGGATCTGAAGCAAAAACACCCAGAGAAGGACTTAGATCAGCTGGTGGAGA
TGGCCAATTACTATGCTCTTTCCCACCAACAGAAGAGCCGTGCCTTCTACCGTATCCAAGCCACTCGT
ATGATGACTGGTGCAGGCAATATCCTGAAGAAACATGCAGCAGAACAAGCCAAGAAGGCCTCCAGCAT
GAGCGAGGTGCACACCGATGAGCCTGAGGACTTTATTTCCAAGGTCTTCTTTGACCCATGTTCTTACC
AGTGCCTGGAGAACTGTGGGGCTGTACTCCTGACAGTGGTGAGGAAAGGGGGAGACATGTCAAAGACC
ATGTATGTGGACTACAAAACAGAGGATGGTTCTGCCAATGCAGGGGCTGACTATGAGTTCACAGAGGG
CACGGTGGTTCTGAAGCCAGGAGAGACCCAGAAGGAGTTCTCCGTGGGCATAATTGATGACGACATTT
TTGAGGAGGATGAACACTTCTTTGTAAGGTTGAGCAATGTCCGCATAGAGGAGGAGCAGCCAGAGGAG
GGGATGCCTCCAGCAATATTCAACAGTCTTCCCTTGCCTCGGGCTGTCCTAGCCTCCCCTTGTGTGGC
CACAGTTACCATCTTGGATGATGACCATGCAGGCATCTTCACTTTTGAATGTGATACTATTCATGTCA
GTGAGAGTATTGGTGTTATGGAGGTCAAGGTTCTGCGGACATCAGGTGCCCGGGGTACAGTCATCGTC
CCCTTTAGGACAGTAGAAGGGACAGCCAAGGGTGGCGGTGAGGACTTTGAAGACACATATGGGGAGTT
GGAATTCAAGAATGATGAAACTGTGAAAACTCTTCAGGTGAAGATAGTTGATGACGAGGAATATGAGA
AAAAGGATAATTTCTTCATTGAGCTGGGCCAGCCCCAGTGGCTTAAGCGAGGGATTTCAGCTCTGCTA
CTCAATCAAGGGGATGGGGACAGGAAGCTAACAGCCGAGGAGGAGGAGGCTCGGAGGATAGCAGAGAT
GGGCAAGCCAGTTCTTGGGGAGAACTGCCGGCTGGAGGTCATCATCGAGGAGTCATATGATTTTAAGA
ACACGGTGGATAAACTCATCAAGAAAACGAACTTGGCCTTGGTAATTGGGACCCATTCATGGAGGGAG
CAGTTTTTAGAGGCAATTACGGTGAGCGCAGGGGACGAGGAGGAGGAGGAGGACGGGTCCCGGGAGGA
GCGGCTGCCGTCGTGCTTTGACTACGTGATGCACTTCCTGACGGTGTTCTGGAAGGTGCTCTTCGCCT
GTGTGCCCCCCACCGAGTACTGCCACGGCTGGGCCTGCTTTGGTGTCTCCATCCTGGTCATCGGCCTG
CTCACCGCCCTCATTGGGGACCTCGCCTCCCACTTCGGCTGCACCGTTGGCCTCAAGGACTCTGTCAA
TGCTGTTGTCTTCGTTGCCCTGGGCACCTCCATCCCTGACACGTTCGCCAGCAAGGTGGCGGCGCTGC
AGGACCAGTGCGCCGACGCGTCCATCGGCAACGTGACCGGCTCCAACGCGGTGAACGTGTTCCTTGGC
CTGGGCGTCGCCTGGTCTGTGGCCGCCGTGTACTGGGCGGTGCAGGGCCGCCCCTTCGAGGTGCGCAC
TGGCACGCTGGCCTTCTCCGTCACGCTCTTCACCGTCTTCGCCTTCGTGGGCATTGCCGTGCTGCTGT
ACCGGCGCCGGCCGCACATCGGCGGCGAGCTGGGCGGCCCGCGCGGACCCAAGCTCGCCACCACCGCG
CTCTTCCTGGGCCTCTGGCTCCTGTACATCCTCTTCGCCAGCCTGGAGGCGTACTGCCACATCCGGGG
CTTC**TAG**GGCCTCGCGCAGAGACTC

NOV55h, CG56258-06 Protein Sequence	SEQ ID NO: 870	928 aa	MW at 102900.1kD

MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGVILPIWYPENP
SLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGETSTTTIRVWNETVSNL
TLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFIIIGICVYVIPDGETRKIKHLRVFF
ITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRG
IIIETEGDHPKGIEMDGKMMNSHFLDGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMAN
YYALSHQQKSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCL
ENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEE
DEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIHVSES
IGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKTLQVKIVDDEEYEKKD
NFFIELGQPQWLKRGISALLLNQGDGDRKLTAEEEEARRIAEMGKPVLGENCRLEVIIEESYDFKNTV
DKLIKKTNLALVIGTHSWREQFLEAITVSAGDEEEEEDGSREERLPSCFDYVMHFLTVFWKVLFACVP
PTEYCHGWACFGVSILVIGLLTALIGDLASHFGCTVGLKDSVNAVVFVALGTSIPDTFASKVAALQDQ
CADASIGNVTGSNAVNVFLGLGVAWSVAAVYWAVQGRPFEVRTGTLAFSVTLFTVFAFVGIAVLLYRR
RPHIGGELGGPRGPKLATTALFLGLWLLYILFASLEAYCHIRGF

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 55B.

Table 55B. Comparison of the NOV55 protein sequences.

```
NOV55a  ---MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEG
NOV55b  ---MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEG
NOV55c  GSTMAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEG
NOV55d  ------------------------------GSEAGGSGDVPSTGQNNESCSGSSDCKEG
NOV55e  ---MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEG
NOV55f  ---------------------------------EAGGSGDVPSTGQNNESCSGSSDCKEG
NOV55g  ---MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEG
NOV55h  ---MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEG

NOV55a  VILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKP
NOV55b  VILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKP
NOV55c  VILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKP
NOV55d  VILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKP
NOV55e  VILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKP
NOV55f  VILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKP
NOV55g  VILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKP
NOV55h  VILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKP

NOV55a  NGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFN
NOV55b  NGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFN
NOV55c  NGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFN
NOV55d  NGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFN
NOV55e  NGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFN
NOV55f  NGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFN
NOV55g  NGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFN
NOV55h  NGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFN

NOV55a  MFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLT
NOV55b  MFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLT
NOV55c  MFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLT
NOV55d  MFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLT
NOV55e  MFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLT
```

```
NOV55f  MFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLT
NOV55g  MFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLT
NOV55h  MFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLT

NOV55a  LFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFL
NOV55b  LFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFL
NOV55c  LFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFL
NOV55d  LFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFL
NOV55e  LFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFL
NOV55f  LFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFL
NOV55g  LFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFL
NOV55h  LFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFL

NOV55a  DGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQ
NOV55b  DGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQ
NOV55c  DGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQ
NOV55d  DGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQ
NOV55e  DGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQ
NOV55f  DGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQ
NOV55g  DGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQ
NOV55h  DGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQ

NOV55a  ATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLT
NOV55b  ATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLT
NOV55c  ATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLT
NOV55d  ATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLT
NOV55e  ATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLT
NOV55f  ATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLT
NOV55g  ATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLT
NOV55h  ATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLT

NOV55a  VVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEED
NOV55b  VVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEED
NOV55c  VVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEED
NOV55d  VVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEED
NOV55e  VVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEED
NOV55f  VVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEED
NOV55g  VVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEED
NOV55h  VVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEED

NOV55a  EHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECD
NOV55b  EHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECD
NOV55c  EHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECD
NOV55d  EHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECD
NOV55e  EHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECD
NOV55f  EHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECD
NOV55g  EHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECD
NOV55h  EHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECD

NOV55a  TIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKT
NOV55b  TIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKT
NOV55c  TIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKT
NOV55d  TIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKT
NOV55e  TIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKT
```

```
NOV55f  TIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKT
NOV55g  TIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKT
NOV55h  TIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKT

NOV55a  IRVKIVDEEEYERQENFFIALGEP----KWMERGISDVT---DRKLTMEEEEAKRIAEMG
NOV55b  IHIKVIDDEAYEKNKNYFIEMMGPRMVDMSFQKALLLSP---DRKLTMEEEEAKRIAEMG
NOV55c  IRVKIVDEEEYERQENFFIALGEP----KWMERGISDVT---DRKLTMEEEEAKRIAEMG
NOV55d  IRVKIVDEEEYERQENFFIALGEP----KWMERGISDVT---DRKLTMEEEEAKRIAEMG
NOV55e  LQVKIVDDEEYEKKDNFFIELGQPQ-WLKRGISALLLNQGDGDRKLTAEEEEARRIAEMG
NOV55f  IRVKIVDEEEYERQENFFIALGEP----KWMERGISDVT---DRKLTMEEEEAKRIAEMG
NOV55g  IRVKIVDEEEYERQENFFIALGEP----KWMERGISDVT---DRKLTMEEEEAKRIAEMG
NOV55h  LQVKIVDDEEYEKKDNFFIELGQPQ-WLKRGISALLLNQGDGDRKLTAEEEEARRIAEMG

NOV55a  KPVLGEHPKLEVIIEESYEFKTTVDKLIKKTNLALVVGTHSWRDQFMEAITVSAAGDEDE
NOV55b  KPVLGEHPKLEVIIEESYEFKTTVDKLIKKTNLALVVGTHSWRDQFMEAITVSAAGDEDE
NOV55c  KPVLGEHPKLEVIIEESYEFKTTVDKLIKKTNLALVVGTHSWRDQFMEAITVSAAGDEDE
NOV55d  KPVLGEHPKLEVIIEESYEFKTTVDKLIKKTNLALVVGTHSWRDQFMEAITVSAAGDEDE
NOV55e  KPVLGENCRLEVIIEESYDFKNTVDKLIKKTNLALVIGTHSWREQFLEAITVSAG-DEEE
NOV55f  KPVLGEHPKLEVIIEESYEFKTTVDKLIKKTNLALVVGTHSWRDQFMEAITVSAAGDEDE
NOV55g  KPVLGEHPKLEVIIEESYEFKTTVDKLIKKTNLALVVGTHSWRDQFMEAITVSAAGDEDE
NOV55h  KPVLGENCRLEVIIEESYDFKNTVDKLIKKTNLALVIGTHSWREQFLEAITVSAG-DEEE

NOV55a  DESG--EERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHGWACFAVSILIIGMLTAIIGD
NOV55b  DESG--EERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHGWACFAVSILIIGMLTAIIGD
NOV55c  DESG--EERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHGWACFAVSILIIGMLTAIIGD
NOV55d  DESG--EERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHGWACFAVSILIIGMLTAIIGD
NOV55e  EEDGSREERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHGWACFGVSILVIGLLTALIGD
NOV55f  DESG--EERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHGWACFAVSILIIGMLTAIIGD
NOV55g  DESG--EERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHGWACFAVSILIIGMLTAIIGD
NOV55h  EEDGSREERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHGWACFGVSILVIGLLTALIGD

NOV55a  LASHFGCTIGLKDSVTAVVFVAFGTSVPDTFASKAAALQDVYADASIGNVTGSNAVNVFL
NOV55b  LASHFGCTIGLKDSVTAVVFVAFGTSVPDTFASKAAALQDVYADASIGNVTGSNAVNVFL
NOV55c  LASHFGCTIGLKDSVTAVVFVAFGTSVPDTFASKAAALQDVYADASIGNVTGSNAVNVFL
NOV55d  LASHFGCTIGLKDSVTAVVFVAFGTSVPDTFASKAAALQDVYADASIGNVTGSNAVNVFL
NOV55e  LASHFGCTVGLKDSVNAVVFVALGTSIPDTFASKVAALQDQCADASIGNVTGSNAVNVFL
NOV55f  LASHFGCTIGLKDSVTAVVFVAFGTSVPDTFASKAAALQDVYADASIGNVTGSNAVNVFL
NOV55g  LASHFGCTIGLKDSVTAVVFVAFGTSVPDTFASKAAALQDVYADASIGNVTGSNAVNVFL
NOV55h  LASHFGCTVGLKDSVNAVVFVALGTSIPDTFASKVAALQDQCADASIGNVTGSNAVNVFL

NOV55a  GIGLAWSVAAIYWALQGQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHLGGELGGP
NOV55b  GIGLAWSVAAIYWALQGQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHLGGELGGP
NOV55c  GIGLAWSVAAIYWALQGQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHLGGELGGP
NOV55d  GIGLAWSVAAIYWALQGQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHLGGELGGP
NOV55e  GLGVAWSVAAVYWAVQGRPFEVRTGTLAFSVTLFTVFAFVGIAVLLYRRRPHIGGELGGP
NOV55f  GIGLAWSVAAIYWALQGQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHLGGELGGP
NOV55g  GIGLAWSVAAIYWALQGQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHLGGELGGP
NOV55h  GLGVAWSVAAVYWAVQGRPFEVRTGTLAFSVTLFTVFAFVGIAVLLYRRRPHIGGELGGP

NOV55a  RGCKLATTWLFVSLWLLYILFATLEAYCYIKGF--
NOV55b  RGCKLATTWLFVSLWLLYILFATLEAYCYIKGF--
NOV55c  RGCKLATTWLFVSLWLLYILFATLEAYCYIKGFLE
NOV55d  RGCKLATTWLFVSLWLLYILFATLEAYCYIKGFLE
NOV55e  RGPKLATTALFLGLWLLYILFASLEAYCHIRGF--
```

```
NOV55f  RGCKLATTWLFVSLWLLYILFATLEAYCYIKGF--
NOV55g  RGCKLATTWLFVSLWLLYILFATLEAYCYIKGF--
NOV55h  RGPKLATTALFLGLWLLYILFASLEAYCHIRGF--

NOV55a  (SEQ ID NO:  856)
NOV55b  (SEQ ID NO:  858)
NOV55c  (SEQ ID NO:  860)
NOV55d  (SEQ ID NO:  862)
NOV55e  (SEQ ID NO:  864)
NOV55f  (SEQ ID NO:  866)
NOV55g  (SEQ ID NO:  868)
NOV55h  (SEQ ID NO:  870)
```

Further analysis of the NOV55a protein yielded the following properties shown in Table 55C.

Table 55C. Protein Sequence Properties NOV55a

SignalP analysis:	Cleavage site between residues 31 and 32

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 5;  pos.chg 1;  neg.chg 0
      H-region:  length 23;  peak value  11.61
      PSG score:   7.21

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   0.67
      possible cleavage site: between 30 and 31

>>> Seems to have a cleavable signal peptide (1 to 30)

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 31
      Tentative number of TMS(s) for the threshold  0.5:  11
      INTEGRAL    Likelihood = -8.92   Transmembrane   78 -  94
      INTEGRAL    Likelihood = -6.64   Transmembrane  170 - 186
      INTEGRAL    Likelihood = -3.24   Transmembrane  208 - 224
      INTEGRAL    Likelihood = -8.49   Transmembrane  235 - 251
      INTEGRAL    Likelihood =  0.21   Transmembrane  508 - 524
      INTEGRAL    Likelihood = -1.12   Transmembrane  724 - 740
      INTEGRAL    Likelihood = -9.61   Transmembrane  750 - 766
      INTEGRAL    Likelihood =  0.05   Transmembrane  775 - 791
      INTEGRAL    Likelihood = -3.35   Transmembrane  823 - 839
      INTEGRAL    Likelihood = -8.39   Transmembrane  858 - 874
      INTEGRAL    Likelihood = -3.56   Transmembrane  893 - 909
      PERIPHERAL  Likelihood =  1.06 (at 137)
      ALOM score:  -9.61  (number of TMSs: 11)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 15
      Charge difference: -3.5   C(-1.0) - N( 2.5)
      N >= C: N-terminal side will be inside
```

```
>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:        2       Hyd Moment(75):  4.35
      Hyd Moment(95):  3.78     G content:       2
      D/E content:     1        S/T content:     3
      Score: -3.73

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 39  LRA|EA

NUCDISC: discrimination of nuclear localization signals
      pat4: RRRP (4) at  876
      pat7: none
      bipartite: none
      content of basic residues:   9.0%
      NLS Score: -0.22

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: AWLR


      KKXX-like motif in the C-terminus: YIKG

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: found
      ILPI at 59

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none
```

```
checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

---------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       11.1 %: nuclear
       11.1 %: vesicles of secretory system
       11.1 %: mitochondrial

>> prediction for CG56258-04 is end (k=9)
```

A search of the NOV55a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 55D.

Table 55D. Geneseq Results for NOV55a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV55a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABB83246	Human transporter protein - Homo sapiens, 921 aa. [WO200233086-A2, 25-APR-2002]	1..921 1..921	921/921 (100%) 921/921 (100%)	0.0
ABB81913	Human ion exchanger protein #1 - Homo sapiens, 921 aa. [WO200259316-A2, 01-AUG-2002]	1..921 1..921	921/921 (100%) 921/921 (100%)	0.0
ABP74104	Human TRICH SEQ ID NO 9 - Homo sapiens, 921 aa. [WO200246415-A2, 13-JUN-2002]	1..921 1..921	921/921 (100%) 921/921 (100%)	0.0
ABB81915	Human ion exchanger protein #1 Asp/Gly mutant - Homo sapiens, 921 aa. [WO200259316-A2, 01-AUG-2002]	1..921 1..921	920/921 (99%) 920/921 (99%)	0.0
ABB81916	Human ion exchanger protein #1 Ala mutant - Homo sapiens, 922 aa. [WO200259316-A2, 01-AUG-2002]	1..921 1..922	921/922 (99%) 921/922 (99%)	0.0

In a BLAST search of public sequence databases, the NOV55a protein was found to have homology to the proteins shown in the BLASTP data in Table 55E.

Table 55E. Public BLASTP Results for NOV55a				
Protein Accession Number	**Protein/Organism/Length**	**NOV55a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8IUF0	Na+/Ca2+ exchanger isoform 3 splice variant 2 - Homo sapiens (Human), 921 aa.	1..921 1..921	921/921 (100%) 921/921 (100%)	0.0
CAD12716	Sequence 1 from Patent WO0183744 - Homo sapiens (Human), 927 aa (fragment).	1..921 1..927	921/927 (99%) 921/927 (99%)	0.0
Q8NFI7	Na+/Ca2+ exchanger isoform 3 - Homo sapiens (Human), 927 aa.	1..921 1..927	921/927 (99%) 921/927 (99%)	0.0
Q96QG1	Sodium/calcium exchanger SCL8A3 - Homo sapiens (Human), 924 aa.	1..921 1..924	918/924 (99%) 919/924 (99%)	0.0
P70549	Sodium/calcium exchanger 3 precursor (Na(+)/Ca(2+)-exchange protein 3) - Rattus norvegicus (Rat), 927 aa.	1..921 1..927	897/927 (96%) 911/927 (97%)	0.0

PFam analysis indicates that the NOV55a protein contains the domains shown in the Table 55F.

Table 55F. Domain Analysis of NOV55a			
Pfam Domain	**NOV55a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Na_Ca_Ex	89..248	41/181 (23%) 134/181 (74%)	1.6e-33
Calx-beta	386..485	50/106 (47%) 88/106 (83%)	1.4e-45
Calx-beta	519..619	51/106 (48%) 95/106 (90%)	3.5e-46
Na_Ca_Ex	757..910	44/178 (25%) 128/178 (72%)	8.7e-31

Example 56.

The NOV56 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 56A.

Table 56A. NOV56 Sequence Analysis		
NOV56a, CG56262-01 DNA Sequence	SEQ ID NO: 871	1551 bp
	ORF Start: ATG at 108	ORF Stop: TGA at 1512

GCGGCCGCGGGAGCTGACCCTGCGGGGTCCCGGGGGGGGAGGGGGAGCCGCGAAGCCCCCACTGAGGC
CGCCGCTGCCGGGCCTCCCCTCCCCCCCGGGCGGGCGCC**ATG**CGGGGGAGCCCGGGCGACGCGGAGCG
GCGGCAGCGCTGGGGTCGCCTGTTCGAGGAGCTGGACAGTAACAAGGATGGCCGCGTGGACGTGCACG
AGTTGCGCCAGGGGCTGGCCAGGCTGGGCGGGGGCAACCCAGACCCCGGCGCCCAACAGGGTATCTCC
TCTGAGGGTGATGCTGACCCAGATGGCGGGCTCGACCTGGAGGAATTTTCCCGCTATCTGCAGGAGCG
GGAACAGCGTCTGCTGCTCATGTTTCACAGTCTTGACCGGAACCAGGATGGTCACATTGATGTCTCTG
AGATCCAACAGAGTTTCCGAGCTCTGGGCATTTCCATCTCGCTGGAGCAGGCTGAGAAAATTTTGCAC
AGCATGGACCGAGACGGCACAATGACCATTGACTGGCAAGAATGGCGCGACCACTTCCTGTTGCATTC
GCTGGAAAATGTGGAGGACGTGCTGTATTTCTGGAAGCATTCCACGGTCCTGGACATTGGCGAGTGCC
TGACAGTGCCGGACGAGTTCTCAAAGCAAGAGAAGCTGACGGGCATGTGGTGGAAACAGCTGGTGGCC
GGCGCAGTGGCAGGTGCCGTGTCACGGACAGGCACGGCCCCTCTGGACCGCCTCAAGGTCTTCATTCA
GGTCCATGCCTCAAAGACCAACCGGCTGAACATCCTTGGGGGGCTTCGAAGCATGGTCCTTGAGGGAG
GCATCCGCTGCCTGTGGCGCGGCAATGGTATTAATGTACTCAAGATTGCCCCCGAGTCAGCTATCAAG
TTCATGGCCTATGAACAGGTGAGGAGGGCCATCCTGGGGCAGCAGGAGACACTGCATGTGCAGGAGCG
CTTCGTGGCTGGCTCCCTGGCTGGTGCCACAGCCCAAACCATCATTTACCCTATGGAGGTGCTGAAGA
CGCGGCTGACCTTGCGCCGGACGGGCCAGTATAAGGGGCTGCTGGACTGCGCCAGGCGTATCCTGGAG
AGGGAGGGGCCCCGTGCCTTCTACCGCGGCTACCTCCCCAACGTGCTGGGCATCATCCCCTATGCGGG
CATCGACCTGGCCGTCTACGAGGTCCTGAAGAACTGGTGGCTTCAGCAGTACAGCCACGACTCGGCAG
ACCCAGGCATCCTCGTGCTCCTGGCCTGCGGTACCATATCCAGCACCTGCGGCCAGATAGCCAGTTAC
CCGCTGGCCCTGGTCCGGACCCGCATGCAGGCACAAGCCTCCATCGAGGGTGGCCCCCAGCTGTCCAT
GCTGGGTCTGCTACGTCACATCCTGTCCCAGGAGGGCATGCGGGGCCTCTACCGGGGGATCGCCCCCA
ACTTCATGAAGGTTATTCCAGCTGTGAGCATCTCCTATGTGGTCTACGAGAACATGAAGCAGGCCTTG
GGGGTCACGTCCAGG**TGA**GGGACCCGGAGCCCGTCCCCCCAATCCCTCACCCCCC

NOV56a, CG56262-01 Protein Sequence	SEQ ID NO: 872	468 aa	MW at 52387.5kD

MRGSPGDAERRQRWGRLFEELDSNKDGRVDVHELRQGLARLGGGNPDPGAQQGISSEGDADPDGGLDL
EEFSRYLQEREQRLLLMFHSLDRNQDGHIDVSEIQQSFRALGISISLEQAEKILHSMDRDGTMTIDWQ
EWRDHFLLHSLENVEDVLYFWKHSTVLDIGECLTVPDEFSKQEKLTGMWWKQLVAGAVAGAVSRTGTA
PLDRLKVFIQVHASKTNRLNILGGLRSMVLEGGIRCLWRGNGINVLKIAPESAIKFMAYEQVRRAILG
QQETLHVQERFVAGSLAGATAQTIIYPMEVLKTRLTLRRTGQYKGLLDCARRILEREGPRAFYRGYLP
NVLGIIPYAGIDLAVYEVLKNWWLQQYSHDSADPGILVLLACGTISSTCGQIASYPLALVRTRMQAQA
SIEGGPQLSMLGLLRHILSQEGMRGLYRGIAPNFMKVIPAVSISYVVYENMKQALGVTSR

NOV56b, 266120550 DNA Sequence	SEQ ID NO: 873	1426 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCACCATGCGGGGGAGCCCGGGCGACGCGGAGCGGCGGCAGCGCTGGGGTCGCCTGTTCG
AGGAGCTGGACAGTAACAAGGATGGCCGCGTGGACGTGCACGAGTTGCGCCAGGGGCTGGCCAGGCTG
GGCGGGGGCAACCCAGACCCCGGCGCCCAACAGGGTATCTCCTCTGAGGGTGATGCTGACCCAGATGG
CGGGCTCGACCTGGAGGAATTTTCCCGCTATCTGCAGGAGCGGGAACAGCGTCTGCTGCTCATGTTTC
ACAGTCTTGACCGGAACCAGGATGGTCACATTGATGTCTCTGAGATCCAACAGAGTTTCCGAGCTCTG
GGCATTTCCATCTCGCTGGAGCAGGCTGAGAAAATTTTGCACAGCATGGACCGAGACGGCACAATGAC
CATTGACTGGCAAGAATGGCGCGACCACTTCCTGTTGCATTCGCTGGAAAATGTGGAGGACGTGCTGT
ATTTCTGGAAGCATTCCACGGTCCTGGACATTGGCGAGTGCCTGACAGTGCCGGACGAGTTCTCAAAG
CAAGAGAAGCTGACGGGCATGTGGTGGAAACAGCTGGTGGCCGGCGCAGTGGCAGGTGCCGTGTCACG
GACAGGCACGGCCCCTCTGGACCGCCTCAAGGTCTTCATGCAGGTCCATGCCTCAAAGACCAACCGGC
TGAACATCCTTGGGGGGCTTCGAAGCATGGTCCTTGAGGGAGGCATCCGCTCCCTGTGGCGCGGCAAT
GGTATTAATGTACTCAAGATTGCCCCCGAGTCAGCTATCAAGTTCATGGCCTATGAACAGATCAAGAG
GGCCATCCTGGGGCAGCAGGAGACACTGCATGTGCAGGAGCGCTTCGTGGCTGGCTCCCTGGCTGGTG
CCACAGCCCAAACCATCATTTACCCTATGGAGGTGCTGAAGACGCGGCTGACCTTGCGCCGGACGGGC
CAGTATAAGGGGCTGCTGGACTGCGCCAGGCGTATCCTGGAGAGGGAGGGGCCCCGTGCCTTCTACCG
CGGCTACCTCCCCAACGTGCTGGGCATCATCCCCTATGCGGGCATCGACCTGGCCGTCTACGAGACTC
TGAAGAACTGGTGGCTTCAGCAGTACAGCCACGACTCGGCAGACCCAGGCATCCTCGTGCTCCTGGCC
TGCGGTACCATATCCAGCACCTGCGGCCAGATAGCCAGTTACCCGCTGGCCCTGGTCCGGACCCGCAT
GCAGGCACAAGCCTCCATCGAGGGTGGCCCCCAGCTGTCCATGCTGGGTCTGCTACGTCACATCCTGT
CCCAGGAGGGCATGCGGGGCCTCTACCGGGGGATCGCCCCCAACTTCATGAAGGTTATTCCAGCTGTG
AGCATCTCCTATGTGGTCTACGAGAACATGAAGCAGGCCTTGGGGGTCACGTCCAGGCTCGAGGGC

NOV56b, 266120550 Protein Sequence	SEQ ID NO: 874	475 aa	MW at 53023.1kD

TGSTMRGSPGDAERRQRWGRLFEELDSNKDGRVDVHELRQGLARLGGGNPDPGAQQGISSEGDADPDG
GLDLEEFSRYLQEREQRLLLMFHSLDRNQDGHIDVSEIQQSFRALGISISLEQAEKILHSMDRDGTMT
IDWQEWRDHFLLHSLENVEDVLYFWKHSTVLDIGECLTVPDEFSKQEKLTGMWWKQLVAGAVAGAVSR
TGTAPLDRLKVFMQVHASKTNRLNILGGLRSMVLEGGIRSLWRGNGINVLKIAPESAIKFMAYEQIKR
AILGQQETLHVQERFVAGSLAGATAQTIIYPMEVLKTRLTLRRTGQYKGLLDCARRILEREGPRAFYR
GYLPNVLGIIPYAGIDLAVYETLKNWWLQQYSHDSADPGILVLLACGTISSTCGQIASYPLALVRTRM
QAQASIEGGPQLSMLGLLRHILSQEGMRGLYRGIAPNFMKVIPAVSISYVVYENMKQALGVTSRLEG

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 56B.

Table 56B. Comparison of the NOV56 protein sequences.

```
NOV56a  ----MRGSPGDAERRQRWGRLFEELDSNKDGRVDVHELRQGLARLGGGNPDPGAQQGISS
NOV56b  TGSTMRGSPGDAERRQRWGRLFEELDSNKDGRVDVHELRQGLARLGGGNPDPGAQQGISS

NOV56a  EGDADPDGGLDLEEFSRYLQEREQRLLLMFHSLDRNQDGHIDVSEIQQSFRALGISISLE
NOV56b  EGDADPDGGLDLEEFSRYLQEREQRLLLMFHSLDRNQDGHIDVSEIQQSFRALGISISLE
```

```
NOV56a  QAEKILHSMDRDGTMTIDWQEWRDHFLLHSLENVEDVLYFWKHSTVLDIGECLTVPDEFS
NOV56b  QAEKILHSMDRDGTMTIDWQEWRDHFLLHSLENVEDVLYFWKHSTVLDIGECLTVPDEFS

NOV56a  KQEKLTGMWWKQLVAGAVAGAVSRTGTAPLDRLKVFIQVHASKTNRLNILGGLRSMVLEG
NOV56b  KQEKLTGMWWKQLVAGAVAGAVSRTGTAPLDRLKVFMQVHASKTNRLNILGGLRSMVLEG

NOV56a  GIRCLWRGNGINVLKIAPESAIKFMAYEQVRRAILGQQETLHVQERFVAGSLAGATAQTI
NOV56b  GIRSLWRGNGINVLKIAPESAIKFMAYEQIKRAILGQQETLHVQERFVAGSLAGATAQTI

NOV56a  IYPMEVLKTRLTLRRTGQYKGLLDCARRILEREGPRAFYRGYLPNVLGIIPYAGIDLAVY
NOV56b  IYPMEVLKTRLTLRRTGQYKGLLDCARRILEREGPRAFYRGYLPNVLGIIPYAGIDLAVY

NOV56a  EVLKNWWLQQYSHDSADPGILVLLACGTISSTCGQIASYPLALVRTRMQAQASIEGGPQL
NOV56b  ETLKNWWLQQYSHDSADPGILVLLACGTISSTCGQIASYPLALVRTRMQAQASIEGGPQL

NOV56a  SMLGLLRHILSQEGMRGLYRGIAPNFMKVIPAVSISYVVYENMKQALGVTSR---
NOV56b  SMLGLLRHILSQEGMRGLYRGIAPNFMKVIPAVSISYVVYENMKQALGVTSRLEG

NOV56a  (SEQ ID NO:  872)
NOV56b  (SEQ ID NO:  874)
```

Further analysis of the NOV56a protein yielded the following properties shown in Table 56C.

Table 56C. Protein Sequence Properties NOV56a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 3;  neg.chg 2
      H-region:  length 1;  peak value -14.40
      PSG score: -18.80

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -7.60
      possible cleavage site: between 46 and 47

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   0
      PERIPHERAL  Likelihood =  0.85 (at 342)
      ALOM score:  -0.90  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75): 10.59
      Hyd Moment(95):  7.77    G content:      2
      D/E content:     2       S/T content:    1
      Score: -5.97
```

```
Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 12  MRG|SP

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  11.3%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: RGSP

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:
      Leucine zipper pattern (PS00029):  *** found ***
         LEEFSRYLQEREQRLLLMFHSL at 68
 none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues
```

```
--------------------------

Final Results (k = 9/23):

       52.2 %: cytoplasmic
       30.4 %: nuclear
        8.7 %: mitochondrial
        4.3 %: vacuolar
        4.3 %: vesicles of secretory system

>> prediction for CG56262-01 is cyt (k=23)
```

A search of the NOV56a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 56D.

Table 56D. Geneseq Results for NOV56a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV56a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE22927	Human transporter and ion channel (TRICH) 26 - Homo sapiens, 468 aa. [WO200222684-A2, 21-MAR-2002]	1..468 1..468	463/468 (98%) 466/468 (98%)	0.0
AAU27869	Human contig polypeptide sequence #22 - Homo sapiens, 509 aa. [WO200164834-A2, 07-SEP-2001]	1..468 39..506	463/468 (98%) 466/468 (98%)	0.0
AAU27697	Human full-length polypeptide sequence #22 - Homo sapiens, 471 aa. [WO200164834-A2, 07-SEP-2001]	1..468 1..468	463/468 (98%) 466/468 (98%)	0.0
ABG22637	Novel human diagnostic protein #22628 - Homo sapiens, 508 aa. [WO200175067-A2, 11-OCT-2001]	1..468 39..508	442/470 (94%) 450/470 (95%)	0.0
ABG30434	Human protein sequence #2 used for determining sequence of unknown gene - Homo sapiens, 456 aa. [JP2002176980-A, 25-JUN-2002]	1..409 1..409	403/409 (98%) 406/409 (98%)	0.0

In a BLAST search of public sequence databases, the NOV56a protein was found to have homology to the proteins shown in the BLASTP data in Table 56E.

Table 56E. Public BLASTP Results for NOV56a

Protein Accession Number	Protein/Organism/Length	NOV56a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD55563	Putative calcium binding transporter - Homo sapiens (Human), 438 aa.	1..407 1..407	402/407 (98%) 405/407 (98%)	0.0
Q96NQ4	Hypothetical protein FLJ30339 - Homo sapiens (Human), 384 aa.	85..468 1..384	379/384 (98%) 382/384 (98%)	0.0
CAD20531	Sequence 1 from Patent WO0174854 - Homo sapiens (Human), 460 aa (fragment).	1..407 1..406	374/407 (91%) 388/407 (94%)	0.0
AAH43834	Similar to hypothetical protein MGC36388 - Xenopus laevis (African clawed frog), 514 aa.	6..468 53..514	302/463 (65%) 368/463 (79%)	e-177
Q8BHG0	Weakly similar to peroxisomal CA-dependent solute carrier - Mus musculus (Mouse), 502 aa.	6..468 41..502	305/463 (65%) 366/463 (78%)	e-176

PFam analysis indicates that the NOV56a protein contains the domains shown in the Table 56F.

Table 56F. Domain Analysis of NOV56a

Pfam Domain	NOV56a Match Region	Identities/ Similarities for the Matched Region	Expect Value
efhand	13..41	8/29 (28%) 26/29 (90%)	0.00022
efhand	81..109	8/29 (28%) 21/29 (72%)	0.012
mito_carr	184..276	38/124 (31%) 78/124 (63%)	2.1e-24
mito_carr	278..369	38/124 (31%) 81/124 (65%)	3.4e-32
mito_carr	375..468	31/124 (25%) 76/124 (61%)	2.4e-23

Example 57.

The NOV57 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 57A.

5

Table 57A. NOV57 Sequence Analysis

NOV57a, CG56315-01 DNA Sequence	SEQ ID NO: 875	728 bp
	ORF Start: ATG at 28	ORF Stop: TGA at 697

AGTCTTGCCTTCTTTTGAGCCTAAGTC**ATG**AGTTGGATGTTCCTCAGAGATCTCCTGAGTGGAGTAAA
TAAATACTCCACTGGGACTGGATGGATTTGGCTGGCTGTCGTGTTTGTCTTCCGTTTGCTGGTCTACA
TGGTGGCAGCAGAGCACGTGTGGAAAGATGAGCAGAAAGAGTTTGAGTGCAACAGTAGACAGCCCGGT
TGCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGACTTTGGGCCTTACAACTGAT
AATGGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAAAGGC
ACAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTATGGTACGCTTATCTTATCAGC
CTCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTATAAGCTATATGATGGCTTTAG
TGTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCCAAAC
CCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGTGTATTGTGTTGAATTTC
ATTGAACTGAGTTTTTTGGTTCTCAAGTGCTTTATTAAGTGCTGTCTCCAAAAATATTTAAAAAAACC
TCAAGTCCTCAGTGTG**TGA**GTGCCACAGCCTCAGATATGTTGAATGTG

NOV57a, CG56315-01 Protein Sequence	SEQ ID NO: 876	223 aa	MW at 25859.7kD

MSWMFLRDLLSGVNKYSTGTGWIWLAVVFVFRLLVYMVAAEHVWKDEQKEFECNSRQPGCKNVCFDDF
FPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLYVSPGTMDGGLWYAYLISLIVKTGFEI
GFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFISKPTEKTIFILFLVITSCLCIVLNFIELSFLVLK
CFIKCCLQKYLKKPQVLSV

NOV57b, 247679382 DNA Sequence	SEQ ID NO: 877	700 bp	
	ORF Start: at 2	ORF Stop: end of sequence	

AGGCTCCGCGGCCGCCCCCTTCACCGGATCCATGAGTTGGATGTTCCTCAGAGATCTCCTGAGTGGAG
TAAATAAATACTCCACTGGGATTGGATGGATTTGGCTGGCTGTCGTGTTTGTCTTCCGTTTGCTGGTC
TACATGGTGGCAGCAGAGCACGTGTGGAAAGATGAGCAGAAAGAGTTTGAGTGCAACAGTAGACAGCC
CGGTTGCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGACTTTGGGCCGTACAAC
TGATAATGGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAA
AGGCACAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTATGGTACGCTTATCTTAT
CAGCCTCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTATAAGCTATATGATGGCT
TTAGTGTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCC
AAACCCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGCGTATTGTGTTGAA
TTTCATTGAACTGAGTTTTTTGCTCGAGGGCAAGGGTGGGCGCGCCGACCCAGCTTTCTTGTACAAAA
GTTGGCATTTTATAAGAAAA

NOV57b, 247679382 Protein Sequence	SEQ ID NO: 878	233 aa	MW at 26752.3kD

GSAAAPFTGSMSWMFLRDLLSGVNKYSTGIGWIWLAVVFVFRLLVYMVAAEHVWKDEQKEFECNSRQP
GCKNVCFDDFFPISQVRLWAVQLIMVSTPSLLVVLHVAYHEGREKRHRKKLYVSPGTMDGGLWYAYLI
SLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFISKPTEKTIFILFLVITSCLRIVLN
FIELSFLLEGKGGRADPAFLYKSWHFIRK

NOV57c, 247678321 DNA Sequence	SEQ ID NO: 879	724 bp
	ORF Start: at 2	ORF Stop: end of sequence

AGGCTCCGCGGCCGCCCCCTTCACCGGATCCATGAGTTGGATGTTCCTCAGAGATCTCCTGAGTGGAG
TAAATAAATACTCCACTGGGATTGGATGGATTTGGCTGGCTGTCGTGTTTGTCTTCCGTTTGCTGGTC
TACATGGTGGCAGCAGAGCACGTGTGGAAAGATGAGCAGAAAGAGTTTGAGTGCAACAGTAGACAGCC
CGGTTGCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGACTTTGGGCCTTACAAC

TGATAATGGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAA
AGGCACAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTATGGTACGCTTATCTTAT
CAGCCTCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTATAAGCTATATGATGGCT
TTAGTGTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCC
AAACCCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGCGTATTGTGTTGAA
TTCCATTGAACTGAGTTTTTTGGTTCTCAAGTGCCTTATTAAGTGCTGTCTCCAAAAATATTTAAAAA
AACCTCAAGTCCTCAGTGTGCTCGAGGGCAAGGGTGGGCGCGCC

NOV57c, 247678321 Protein Sequence	SEQ ID NO: 880	241 aa	MW at 27446.4kD

GSAAAPFTGSMSWMFLRDLLSGVNKYSTGIGWIWLAVVFVFRLLVYMVAAEHVWKDEQKEFECNSRQP
GCKNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLYVSPGTMDGGLWYAYLI
SLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFISKPTEKTIFILFLVITSCLRIVLN
SIELSFLVLKCLIKCCLQKYLKKPQVLSVLEGKGGRA

NOV57d, 247679418 DNA Sequence	SEQ ID NO: 881	538 bp
	ORF Start: at 2	ORF Stop: end of sequence

AGGCTCCGCGGCCGCCCCCTTCACCGGATCCGAGCACGTGTGGAAAGATGAGCAGAAAGAGTTTGAGT
GCAACAGTAGACAGCCCGGTTGCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGA
CTTTGGGCCTTACAACTGATAATGGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCA
TGAGGGTAGAGAGAAAAGGCACAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTAT
GGTACGCTTATCTTATCAGCCTCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTAT
AAGCTATATGATGGCTTTAGTGTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGT
GGACTGCTTCATCTCCAAACCCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCT
TGCGTATTGTGTTGAATTTCATTGAACTGAGTTTTTTGCTCGAGGGCAAGGGTGGGCGCGCC

NOV57d, 247679418 Protein Sequence	SEQ ID NO: 882	179 aa	MW at 20339.7kD

GSAAAPFTGSEHVWKDEQKEFECNSRQPGCKNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYH
EGREKRHRKKLYVSPGTMDGGLWYAYLISLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTV
DCFISKPTEKTIFILFLVITSCLRIVLNFIELSFLLEGKGGRA

NOV57e, 247679395 DNA Sequence	SEQ ID NO: 883	658 bp
	ORF Start: at 2	ORF Stop: end of sequence

AGGCTCCGCGGCCGCCCCCTTCACCGGATCCATGAGTTGGATGTTCCTCAGAGATCTCCTGAGTGGAG
TAAATAAATACTCCACTGGGATTGGATGGATTTGGCTGGCTGTCGTGTTTGTCTTCCGTTTGCTGGTC
TACATGGTGGCAGCAGAGCACGTGTGGAAAGATGAGCAGAAAGAGTTTGAGTGCAACAGTAGACAGCC
CGGTTGCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGACTTTGGGCCTTACAAC
TGATAATGGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAA
AGGCACAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTATGGTACGCTTATCTTAT
CAGCCTCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTATAAGCTATATGATGGCT
TTAGTGTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCC
AAACCCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGCGTATTGTGTTGAA
TTTCATTGAACTGAGTTTTTTGCTCGAGGGCAAGGGTGGGCGCGCC

NOV57e, 247679395 Protein Sequence	SEQ ID NO: 884	219 aa	MW at 24976.2kD

GSAAAPFTGSMSWMFLRDLLSGVNKYSTGIGWIWLAVVFVFRLLVYMVAAEHVWKDEQKEFECNSRQP
GCKNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLYVSPGTMDGGLWYAYLI
SLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFISKPTEKTIFILFLVITSCLRIVLN
FIELSFLLEGKGGRA

NOV57f, 247679328 DNA Sequence	SEQ ID NO: 885	604 bp
	ORF Start: at 2	ORF Stop: end of sequence

AGGCTCCGCGGCCGCCCCCTTCACCGGATCCGAGCACGTGTGGAAAGATGAGCAGAAAGAGTTTGAGT
GCAACAGTAGACAGCCCGGTTGCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGA
CTTTGGGCCTTACAACTGATAATGGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCA
TGAGGGTAGAGAGAAAAGGCACAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTAT

GGTACGCTTATCTTATCAGCCTCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTAT
AAGCTATATGATGGCTTTAGTGTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGT
GGACTGCTTCATCTCCAAACCCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCT
TGCGTATTGTGTTGAATTCCATTGAACTGAGTTTTTTGGTTCTCAAGTGCCTTATTAAGTGCTGTCTC
CAAAAATATTTAAAAAAACCTCAAGTCCTCAGTGTGCTCGAGGGCAAGGGTGGGCGCGCC

NOV57f, 247679328 Protein Sequence	SEQ ID NO: 886	201 aa	MW at 22809.8kD

GSAAAPFTGSEHVWKDEQKEFECNSRQPGCKNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYH
EGREKRHRKKLYVSPGTMDGGLWYAYLISLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTV
DCFISKPTEKTIFILFLVITSCLRIVLNSIELSFLVLKCLIKCCLQKYLKKPQVLSVLEGKGGRA

NOV57g, CG56315-02 DNA Sequence	SEQ ID NO: 887	727 bp
	ORF Start: ATG at 27	ORF Stop: TGA at 696

AGTCTTGCTTCTTTTGAGCCTAAGTC**ATG**AGTTGGATGTTCCTCAGAGATCTCCTGAGTGGAGTAAAT
AAATACTCCACTGGGATTGGATGGATTTGGCTGGCTGTCGTGTTTGTCTTCCGTTTGCTGGTCTACAT
GGTGGCAGCAGAGCACGTGTGGAAAGATGAGCAGAAAGAGTTTGAGTGCAACAGTAGACAGCCCGGTT
GCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGACTTTGGGCCTTACAACTGATA
ATGGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAAAGGCA
CAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTATGGTACGCTTATCTTATCAGCC
TCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTATAAGCTATATGATGGCTTTAGT
GTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCCAAACC
CACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGTGTATTGTGTTGAATTTCA
TTGAACTGAGTTTTTTGGTTCTCAAGTGCTTTATTAAGTGCTGTCTCCAAAAATATTTAAAAAAACCT
CAAGTCCTCAGTGTG**TGA**GTGCCACAGCCTCAGATATGTTGAATGTG

NOV57g, CG56315-02 Protein Sequence	SEQ ID NO: 888	223 aa	MW at 25871.7kD

MSWMFLRDLLSGVNKYSTGIGWIWLAVVFVFRLLVYMVAAEHVWKDEQKEFECNSRQPGCKNVCFDDF
FPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLYVSPGTMDGGLWYAYLISLIVKTGFEI
GFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFISKPTEKTIFILFLVITSCLCIVLNFIELSFLVLK
CFIKCCLQKYLKKPQVLSV

NOV57h, CG56315-03 DNA Sequence	SEQ ID NO: 889	24 bp
	ORF Start: at 1	ORF Stop: end of sequence

TTTGAGCAAAACAGAAGACAGCCC

NOV57h, CG56315-03 Protein Sequence	SEQ ID NO: 890	8 aa	MW at 1074.2kD

FEQNRRQP

NOV57i, CG56315-04 DNA Sequence	SEQ ID NO: 891	24 bp
	ORF Start: at 1	ORF Stop: end of sequence

TTTGAGTGCAACAGGAGACAGCCC

NOV57i, CG56315-04 Protein Sequence	SEQ ID NO: 892	8 aa	MW at 1049.2kD

FECNRRQP

NOV57j, CG56315-05 DNA Sequence	SEQ ID NO: 893	24 bp
	ORF Start: at 1	ORF Stop: end of sequence

TTTGAGCAAAACAGTAGACAGCCC

NOV57j, CG56315-05 Protein Sequence	SEQ ID NO: 894	8 aa	MW at 1005.1kD

FEQNSRQP

10

NOV57k, CG56315-06 DNA Sequence	SEQ ID NO: 895	24 bp
	ORF Start: at 1	ORF Stop: end of sequence

TTTGAGTGCAACAGTAGACAGCCC

NOV57k, CG56315-06 Protein Sequence	SEQ ID NO: 896	8 aa	MW at 980.1kD

FECNSRQP

NOV57l, CG56315-07 DNA Sequence	SEQ ID NO: 897	24 bp
	ORF Start: at 1	ORF Stop: end of sequence

TTTGAGCAAAACAGTAGACAGGCC

NOV57l, CG56315-07 Protein Sequence	SEQ ID NO: 898	8 aa	MW at 979.0kD

FEQNSRQA

NOV57m, CG56315-08 DNA Sequence	SEQ ID NO: 899	24 bp
	ORF Start: at 1	ORF Stop: end of sequence

TTTGAGTGCAACAGTAGACAGGCC

NOV57m, CG56315-08 Protein Sequence	SEQ ID NO: 900	8 aa	MW at 954.0kD

FECNSRQA

NOV57n, SNP13381650 of CG56315-01, DNA Sequence	SEQ ID NO: 901	728 bp
	ORF Start: ATG at 28	ORF Stop: TGA at 697
	SNP Pos: 86	SNP Change: C to T

AGTCTTGCCTTCTTTTGAGCCTAAGTC**ATG**AGTTGGATGTTCCTCAGAGATCTCCTGAGTGGAGTAAA
TAAATACTCCACTGGGA**T**TGGATGGATTTGGCTGGCTGTCGTGTTTGTCTTCCGTTTGCTGGTCTACA
TGGTGGCAGCAGAGCACGTGTGGAAAGATGAGCAGAAAGAGTTTGAGTGCAACAGTAGACAGCCCGGT
TGCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGACTTTGGGCCTTACAACTGAT
AATGGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAAAGGC
ACAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTATGGTACGCTTATCTTATCAGC
CTCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTATAAGCTATATGATGGCTTTAG
TGTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCCAAAC
CCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGTGTATTGTGTTGAATTTC
ATTGAACTGAGTTTTTTGGTTCTCAAGTGCTTTATTAAGTGCTGTCTCCAAAAATATTTAAAAAAACC
TCAAGTCCTCAGTGTG**TGA**GTGCCACAGCCTCAGATATGTTGAATGTG

NOV57n, SNP13381650 of CG56315-01, Protein Sequence	SEQ ID NO: 902	223 aa	MW at 25871.7kD
	SNP Pos: 20		SNP Change: Thr to Ile

MSWMFLRDLLSGVNKYSTG**I**GWIWLAVVFVFRLLVYMVAAEHVWKDEQKEFECNSRQPGCKNVCFDDF
FPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLYVSPGTMDGGLWYAYLISLIVKTGFEI
GFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFISKPTEKTIFILFLVITSCLCIVLNFIELSFLVLK
CFIKCCLQKYLKKPQVLSV

NOV57o, SNP13381651 of CG56315-01, DNA Sequence	SEQ ID NO: 903	728 bp
	ORF Start: ATG at 28	ORF Stop: TGA at 697
	SNP Pos: 154	SNP Change: G to A

AGTCTTGCCTTCTTTTGAGCCTAAGTC**ATG**AGTTGGATGTTCCTCAGAGATCTCCTGAGTGGAGTAAA
TAAATACTCCACTGGGACTGGATGGATTTGGCTGGCTGTCGTGTTTGTCTTCCGTTTGCTGGTCTACA
TGGTGGCAGCAGAGCAC**A**TGTGGAAAGATGAGCAGAAAGAGTTTGAGTGCAACAGTAGACAGCCCGGT
TGCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGACTTTGGGCCTTACAACTGAT
AATGGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAAAGGC

ACAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTATGGTACGCTTATCTTATCAGC
CTCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTATAAGCTATATGATGGCTTTAG
TGTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCCAAAC
CCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGTGTATTGTGTTGAATTTC
ATTGAACTGAGTTTTTTGGTTCTCAAGTGCTTTATTAAGTGCTGTCTCCAAAAATATTTAAAAAAACC
TCAAGTCCTCAGTGTGTGAGTGCCACAGCCTCAGATATGTTGAATGTG

NOV57o, SNP13381651 of CG56315-01, Protein Sequence	SEQ ID NO: 904	223 aa	MW at 25891.7kD
	SNP Pos: 43		SNP Change: Val to Met

MSWMFLRDLLSGVNKYSTGTGWIWLAVVFVFRLLVYMVAAEHMWKDEQKEFECNSRQPGCKNVCFDDF
FPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLYVSPGTMDGGLWYAYLISLIVKTGFEI
GFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFISKPTEKTIFILFLVITSCLCIVLNFIELSFLVLK
CFIKCCLQKYLKKPQVLSV

NOV57p, SNP13381652 of CG56315-01, DNA Sequence	SEQ ID NO: 905	728 bp
	ORF Start: ATG at 28	ORF Stop: TGA at 697
	SNP Pos: 276	SNP Change: G to A

AGTCTTGCCTTCTTTTGAGCCTAAGTCATGAGTTGGATGTTCCTCAGAGATCTCCTGAGTGGAGTAAA
TAAATACTCCACTGGGACTGGATGGATTTGGCTGGCTGTCGTGTTTGTCTTCCGTTTGCTGGTCTACA
TGGTGGCAGCAGAGCACGTGTGGAAAGATGAGCAGAAAGAGTTTGAGTGCAACAGTAGACAGCCCGGT
TGCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGACTTTGGGCCTTACAACTGAT
AATAGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAAAGGC
ACAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTATGGTACGCTTATCTTATCAGC
CTCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTATAAGCTATATGATGGCTTTAG
TGTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCCAAAC
CCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGTGTATTGTGTTGAATTTC
ATTGAACTGAGTTTTTTGGTTCTCAAGTGCTTTATTAAGTGCTGTCTCCAAAAATATTTAAAAAAACC
TCAAGTCCTCAGTGTGTGAGTGCCACAGCCTCAGATATGTTGAATGTG

NOV57p, SNP13381652 of CG56315-01, Protein Sequence	SEQ ID NO: 906	223 aa	MW at 25841.6kD
	SNP Pos: 83		SNP Change: Met to Ile

MSWMFLRDLLSGVNKYSTGTGWIWLAVVFVFRLLVYMVAAEHVWKDEQKEFECNSRQPGCKNVCFDDF
FPISQVRLWALQLIIVSTPSLLVVLHVAYHEGREKRHRKKLYVSPGTMDGGLWYAYLISLIVKTGFEI
GFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFISKPTEKTIFILFLVITSCLCIVLNFIELSFLVLK
CFIKCCLQKYLKKPQVLSV

NOV57q, SNP13381653 of CG56315-01, DNA Sequence	SEQ ID NO: 907	728 bp
	ORF Start: ATG at 28	ORF Stop: TGA at 697
	SNP Pos: 557	SNP Change: C to T

AGTCTTGCCTTCTTTTGAGCCTAAGTCATGAGTTGGATGTTCCTCAGAGATCTCCTGAGTGGAGTAAA
TAAATACTCCACTGGGACTGGATGGATTTGGCTGGCTGTCGTGTTTGTCTTCCGTTTGCTGGTCTACA
TGGTGGCAGCAGAGCACGTGTGGAAAGATGAGCAGAAAGAGTTTGAGTGCAACAGTAGACAGCCCGGT
TGCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGACTTTGGGCCTTACAACTGAT
AATGGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAAAGGC
ACAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTATGGTACGCTTATCTTATCAGC
CTCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTATAAGCTATATGATGGCTTTAG
TGTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCCAAAC
CCACTGAGAAGATGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGTGTATTGTGTTGAATTTC
ATTGAACTGAGTTTTTTGGTTCTCAAGTGCTTTATTAAGTGCTGTCTCCAAAAATATTTAAAAAAACC
TCAAGTCCTCAGTGTGTGAGTGCCACAGCCTCAGATATGTTGAATGTG

NOV57q, SNP13381653 of CG56315-01, Protein Sequence	SEQ ID NO: 908	223 aa	MW at 25889.8kD
	SNP Pos: 177		SNP Change: Thr to Met

MSWMFLRDLLSGVNKYSTGTGWIWLAVVFVFRLLVYMVAAEHVWKDEQKEFECNSRQPGCKNVCFDDF
FPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLYVSPGTMDGGLWYAYLISLIVKTGFEI
GFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFISKPTEKMIFILFLVITSCLCIVLNFIELSFLVLK
CFIKCCLQKYLKKPQVLSV

NOV57r, SNP13381654 of CG56315-01, DNA Sequence	SEQ ID NO: 909	728 bp
	ORF Start: ATG at 28	ORF Stop: TGA at 697
	SNP Pos: 643	SNP Change: T to C

AGTCTTGCCTTCTTTTGAGCCTAAGTCATGAGTTGGATGTTCCTCAGAGATCTCCTGAGTGGAGTAAA
TAAATACTCCACTGGGACTGGATGGATTTGGCTGGCTGTCGTGTTTGTCTTCCGTTTGCTGGTCTACA
TGGTGGCAGCAGAGCACGTGTGGAAAGATGAGCAGAAAGAGTTTGAGTGCAACAGTAGACAGCCCGGT
TGCAAAAATGTGTGTTTTGATGACTTCTTCCCCATTTCCCAAGTCAGACTTTGGGCCTTACAACTGAT
AATGGTCTCCACACCTTCACTTCTGGTGGTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAAAGGC
ACAGAAAGAAACTCTATGTCAGCCCAGGTACAATGGATGGGGGCCTATGGTACGCTTATCTTATCAGC
CTCATTGTTAAAACTGGTTTTGAAATTGGCTTCCTTGTTTTATTTTATAAGCTATATGATGGCTTTAG
TGTTCCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCCAAAC
CCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGTGTATTGTGTTGAATTTC
ATTGAACTGAGTTTTTTGGTTCTCAAGTGCCTTATTAAGTGCTGTCTCCAAAAATATTTAAAAAAACC
TCAAGTCCTCAGTGTGTGAGTGCCACAGCCTCAGATATGTTGAATGTG

NOV57r, SNP13381654 of CG56315-01, Protein Sequence	SEQ ID NO: 910	223 aa	MW at 25825.6kD
	SNP Pos: 206		SNP Change: Phe to Leu

MSWMFLRDLLSGVNKYSTGTGWIWLAVVFVFRLLVYMVAAEHVWKDEQKEFECNSRQPGCKNVCFDDF
FPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLYVSPGTMDGGLWYAYLISLIVKTGFEI
GFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFISKPTEKTIFILFLVITSCLCIVLNFIELSFLVLK
CLIKCCLQKYLKKPQVLSV

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 57B.

Table 57B. Comparison of the NOV57 protein sequences.

```
NOV57a  ----------MSWMFLRDLLSGVNKYSTGTGWIWLAVVFVFRLLVYMVAAEHVWKDEQKE
NOV57b  GSAAAPFTGSMSWMFLRDLLSGVNKYSTGIGWIWLAVVFVFRLLVYMVAAEHVWKDEQKE
NOV57c  GSAAAPFTGSMSWMFLRDLLSGVNKYSTGIGWIWLAVVFVFRLLVYMVAAEHVWKDEQKE
NOV57d  GSAAAPFTG-----------------------------------------SEHVWKDEQKE
NOV57e  GSAAAPFTGSMSWMFLRDLLSGVNKYSTGIGWIWLAVVFVFRLLVYMVAAEHVWKDEQKE
NOV57f  GSAAAPFTG-----------------------------------------SEHVWKDEQKE
NOV57g  ----------MSWMFLRDLLSGVNKYSTGIGWIWLAVVFVFRLLVYMVAAEHVWKDEQKE
NOV57h  ------------------------------------------------------------
NOV57i  ------------------------------------------------------------
NOV57j  ------------------------------------------------------------
NOV57k  ------------------------------------------------------------
NOV57l  ------------------------------------------------------------
NOV57m  ------------------------------------------------------------

NOV57a  FECNSRQPGCKNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLY
NOV57b  FECNSRQPGCKNVCFDDFFPISQVRLWAVQLIMVSTPSLLVVLHVAYHEGREKRHRKKLY
NOV57c  FECNSRQPGCKNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLY
NOV57d  FECNSRQPGCKNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLY
NOV57e  FECNSRQPGCKNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLY
NOV57f  FECNSRQPGCKNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLY
NOV57g  FECNSRQPGCKNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRHRKKLY
NOV57h  FEQNRRQP------------------------------------------------------
NOV57i  FECNRRQP------------------------------------------------------
NOV57j  FEQNSRQP------------------------------------------------------
NOV57k  FECNSRQP------------------------------------------------------
NOV57l  FEQNSRQA------------------------------------------------------
```

```
NOV57m  FECNSRQA----------------------------------------------------

NOV57a  VSPGTMDGGLWYAYLISLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFI
NOV57b  VSPGTMDGGLWYAYLISLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFI
NOV57c  VSPGTMDGGLWYAYLISLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFI
NOV57d  VSPGTMDGGLWYAYLISLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFI
NOV57e  VSPGTMDGGLWYAYLISLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFI
NOV57f  VSPGTMDGGLWYAYLISLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFI
NOV57g  VSPGTMDGGLWYAYLISLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCDLKPCPNTVDCFI
NOV57h  ------------------------------------------------------------
NOV57i  ------------------------------------------------------------
NOV57j  ------------------------------------------------------------
NOV57k  ------------------------------------------------------------
NOV57l  ------------------------------------------------------------
NOV57m  ------------------------------------------------------------

NOV57a  SKPTEKTIFILFLVITSCLCIVLNFIELSFLVLKCFIKCCLQKYLKKPQVLSV-------
NOV57b  SKPTEKTIFILFLVITSCLRIVLNFIELSFLLEGKGGRADPAFLYKSWHFIRK-------
NOV57c  SKPTEKTIFILFLVITSCLRIVLNSIELSFLVLKCLIKCCLQKYLKKPQVLSVLEGKGGR
NOV57d  SKPTEKTIFILFLVITSCLRIVLNFIELSFLLEGKGGRA----------------------
NOV57e  SKPTEKTIFILFLVITSCLRIVLNFIELSFLLEGKGGRA----------------------
NOV57f  SKPTEKTIFILFLVITSCLRIVLNSIELSFLVLKCLIKCCLQKYLKKPQVLSVLEGKGGR
NOV57g  SKPTEKTIFILFLVITSCLCIVLNFIELSFLVLKCFIKCCLQKYLKKPQVLSV-------
NOV57h  ------------------------------------------------------------
NOV57i  ------------------------------------------------------------
NOV57j  ------------------------------------------------------------
NOV57k  ------------------------------------------------------------
NOV57l  ------------------------------------------------------------
NOV57m  ------------------------------------------------------------

NOV57a  -
NOV57b  -
NOV57c  A
NOV57d  -
NOV57e  -
NOV57f  A
NOV57g  -
NOV57h  -
NOV57i  -
NOV57j  -
NOV57k  -
NOV57l  -
NOV57m  -

NOV57a  (SEQ ID NO:  876)
NOV57b  (SEQ ID NO:  878)
NOV57c  (SEQ ID NO:  880)
NOV57d  (SEQ ID NO:  882)
NOV57e  (SEQ ID NO:  884)
NOV57f  (SEQ ID NO:  886)
NOV57g  (SEQ ID NO:  888)
NOV57h  (SEQ ID NO:  890)
NOV57i  (SEQ ID NO:  892)
NOV57j  (SEQ ID NO:  894)
```

NOV57k	(SEQ ID NO: 896)
NOV57l	(SEQ ID NO: 898)
NOV57m	(SEQ ID NO: 900)

Further analysis of the NOV57a protein yielded the following properties shown in Table 57C.

Table 57C. Protein Sequence Properties NOV57a

SignalP analysis:	Cleavage site between residues 41 and 42

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 8;  pos.chg 1;  neg.chg 1
      H-region:  length 6;  peak value  -2.35
      PSG score:  -6.75

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -4.71
      possible cleavage site: between 40 and 41

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   5
      INTEGRAL    Likelihood = -7.64   Transmembrane   23 -  39
      INTEGRAL    Likelihood = -2.60   Transmembrane   81 -  97
      INTEGRAL    Likelihood = -3.24   Transmembrane  125 - 141
      INTEGRAL    Likelihood =-12.26   Transmembrane  177 - 193
      INTEGRAL    Likelihood = -2.81   Transmembrane  195 - 211
      PERIPHERAL  Likelihood =  6.36 (at 63)
      ALOM score: -12.26  (number of TMSs: 5)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 30
      Charge difference: -1.5   C( 0.5) - N( 2.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       2       Hyd Moment(75): 11.62
      Hyd Moment(95):  9.22    G content:      3
      D/E content:     2       S/T content:    5
      Score: -4.01

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 42  FRL|LV

NUCDISC: discrimination of nuclear localization signals
      pat4: KRHR (3) at  103
      pat4: RHRK (3) at  104
```

```
      pat4: HRKK (3) at  105
      pat7: none
      bipartite: none
      content of basic residues:  11.2%
      NLS Score:  0.09

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

---------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       33.3 %: mitochondrial
       11.1 %: nuclear

>> prediction for CG56315-01 is end (k=9)
```

A search of the NOV57a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 57D.

Table 57D. Geneseq Results for NOV57a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV57a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE26418	Human transmembrane protein (TMP)-4 protein - Homo sapiens, 223 aa. [WO200234783-A2, 02-MAY-2002]	1..223 1..223	221/223 (99%) 221/223 (99%)	e-131
ABU56673	Lung cancer-associated polypeptide #266 - Unidentified, 273 aa. [WO200286443-A2, 31-OCT-2002]	1..209 1..216	123/216 (56%) 161/216 (73%)	6e-73
ABU56449	Lung cancer-associated polypeptide #42 - Unidentified, 273 aa. [WO200286443-A2, 31-OCT-2002]	1..209 1..216	123/216 (56%) 161/216 (73%)	6e-73
AAY36192	Human secreted protein #64 - Homo sapiens, 273 aa. [WO9925825-A2, 27-MAY-1999]	1..209 1..216	123/216 (56%) 161/216 (73%)	6e-73
AAY36145	Human secreted protein #17 - Homo sapiens, 273 aa. [WO9925825-A2, 27-MAY-1999]	1..209 1..216	123/216 (56%) 161/216 (73%)	6e-73

In a BLAST search of public sequence databases, the NOV57a protein was found to have homology to the proteins shown in the BLASTP data in Table 57E.

Table 57E. Public BLASTP Results for NOV57a				
Protein Accession Number	**Protein/Organism/Length**	**NOV57a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96KP0	Connexin25 - Homo sapiens (Human), 223 aa.	1..223 1..223	222/223 (99%) 223/223 (99%)	e-131
Q91XA4	Similar to gap junction membrane channel protein beta 5 (Gap junction protein) (Connexin) - Mus musculus (Mouse), 271 aa.	1..209 1..214	123/214 (57%) 166/214 (77%)	6e-73
Q02739	Gap junction beta-5 protein (Connexin 31.1) (Cx31.1) - Mus musculus (Mouse), 271 aa.	1..209 1..214	123/214 (57%) 166/214 (77%)	6e-73
O95377	Gap junction beta-5 protein (Connexin 31.1) (Cx31.1) - Homo sapiens (Human), 273 aa.	1..209 1..216	123/216 (56%) 161/216 (73%)	2e-72
P28232	Gap junction beta-5 protein (Connexin 31.1) (Cx31.1) - Rattus norvegicus (Rat), 271 aa.	1..213 1..218	122/218 (55%) 167/218 (75%)	4e-72

PFam analysis indicates that the NOV57a protein contains the domains shown in the Table 57F.

Table 57F. Domain Analysis of NOV57a			
Pfam Domain	**NOV57a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
connexin	2..201	109/247 (44%) 166/247 (67%)	2.2e-111

Example 58.

The NOV58 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 58A.

Table 58A. NOV58 Sequence Analysis		
NOV58a, CG56398-01 DNA Sequence	SEQ ID NO: 911	2105 bp
	ORF Start: ATG at 31	ORF Stop: TAG at 2056
CCTCAGGATCCAGAGGTCTCGTTCAGGACCATGGAGAGCGGCACCAGCAGCCCTCAGCCTCCACAGTT		

AGATCCCCTGGATGCGTTTCCCCAGAAGGGCTTGGAGCCTGGGGACATCGCGGTGCTAGTTCTGTACT
TCCTCTTTGTCCTGGCTGTTGGACTATGGTCCACAGTGAAGACCAAAAGAGACACAGTGAAAGGCTAC
TTCCTGGCTGGAGGGGACATGGTGTGGTGGCCAGTGGGTGCATCCTTGTTTGCCAGCAATGTTGGAAG
TGGACATTTCATTGGCCTGGCAGGGTCAGGTGCTGCTACGGGCATTTCTGTATCAGCTTATGAACTTA
ATGGCTTGTTTTCTGTGCTGATGTTGGCCTGGATCTTCCTACCCATCTACATTGCTGGTCAGGTGACC
ACGATGCCAGAATACCTACGGAAGCGCTTCGGTGGCATCAGAATCCCCATCATCCTGGCTGTACTCTA
CCTATTTATCTACATCTTCACCAAGATCTCGGTAGACATGTATGCAGGTGCCATCTTCATCCAGCAGT
CTTTGCACCTGGATCTGTACCTGGCCATAGTTGGGCTACTGGCCATCACTGCTGTATACACGGTTGCT
GGTGGCCTGGCTGCTGTGATCTACACGGATGCCCTGCAGACGCTGATCATGCTTATAGGAGCGCTCAC
CTTGATGGGCTACAGTTTCGCCGCGGTTGGTGGGATGGAAGGACTGAAGGAGAAGTACTTCTTGGCCC
TGGCTAGCAACCGGAGTGAGAACAGCAGCTGCGGGCTGCCCCGGGAAGATGCCTTCCATATTTTCCGA
GATCCGCTGACATCTGATCTCCCGTGGCCGGGGGTCCTATTTGGAATGTCCATCCCATCCCTCTGGTA
CTGGTGCACGGATCAGGTAATTGTCCAGCGGACTCTGGCTGCCAAGAACCTGTCCCATGCCAAAGGAG
GTGCTCTGATGGCTGCATACCTGAAGGTGCTGCCCCTCTTCATAATGGTGTTCCCTGGGATGGTCAGC
CGCATCCTCTTCCCAGATCAAGTGGCCTGTGCAGATCCAGAGATCTGCCAGAAGATCTGCAGCAACCC
CTCAGGCTGTTCGGACATCGCGTATCCCAAACTCGTGCTGGAACTCCTGCCCACAGGTCTCCGTGGGC
TGATGATGGCTGTGATGGTGGCGGCTCTCATGTCCTCCCTCACCTCCATCTTTAACAGTGCCAGCACC
ATCTTCACCATGGACCTCTGGAATCACCTCCGGCCTCGGGCATCTGAGAAGGAGCTCATGATTGTGGG
CAGGGTGTTTGTGCTGCTGCTGGTCCTGGTCTCCATCCTCTGGATCCCTGTGGTCCAGGCCAGCCAGG
GCGGCCAGCTCTTCATCTATATCCAGTCCATCAGCTCCTACCTGCAGCCGCCTGTGGCGGTGGTCTTC
ATCATGGGATGTTTCTGGAAGAGGACCAATGAAAAGGGTGCCTTCTGGGGCCTGATCTCGGGCCTGCT
CCTGGGCTTGGTTAGGCTGGTCCTGGACTTTATTTACGTGCAGCCTCGATGCGACCAGCCAGATGAGC
GCCCGGTCCTGGTGAAGAGCATTCACTACCTCTACTTCTCCATGATCCTGTCCACGGTCACCCTCATC
ACTGTCTCCACCGTGAGCTGGTTCACAGAGCCACCCTCCAAGGAGATGGTCAGCCACCTGACCTGGTT
TACTCGTCACGACCCCGTGGTCCAGAAGGAACAAGCACCACCAGCAGCTCCCTTGTCTCTTACCCTCT
CTCAGAACGGGATGCCAGAGGCCAGCAGCAGCAGCAGCGTCCAGTTCGAGATGGTTCAAGAAAACACG
TCTAAAACCCACAGCGGTGACATGACCCCAAAGCAGTCCAAAGTGGTGAAGGCCATCCTGTGGCTCTG
TGGAATACAGGAGAAGGGCAAGGAAGAGCTCCCGGCCAGAGCAGAAGCCATCATAGTTTCCCTGGAAG
AAAACCCCTTGGTGAAGACCCTCCTGGACGTCAACCTCATTTTCTGCGTGAGCTGCGCCATCTTTATC
TGGGGCTATTTTGCT**TAG**TGTGGGGTGAACCCAGGGGTCCAAACTCTGTTTCTCTTCAGTGCTCC

NOV58a, CG56398-01 Protein Sequence	SEQ ID NO: 912	675 aa	MW at 73989.3kD

MESGTSSPQPPQLDPLDAFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTKRDTVKGYFLAGGDMVWW
PVGASLFASNVGSGHFIGLAGSGAATGISVSAYELNGLFSVLMLAWIFLPIYIAGQVTTMPEYLRKRF
GGIRIPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDLYLAIVGLLAITAVYTVAGGLAAVIYTD
ALQTLIMLIGALTLMGYSFAAVGGMEGLKEKYFLALASNRSENSSCGLPREDAFHIFRDPLTSDLPWP
GVLFGMSIPSLWYWCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVAC
ADPEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTMDLWNHL
RPRASEKELMIVGRVFVLLLVLVSILWIPVVQASQGGQLFIYIQSISSYLQPPVAVVFIMGCFWKRTN
EKGAFWGLISGLLLGLVRLVLDFIYVQPRCDQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWFTE
PPSKEMVSHLTWFTRHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSGDMTP
KQSKVVKAILWLCGIQEKGKEELPARAEAIIVSLEENPLVKTLLDVNLIFCVSCAIFIWGYFA

NOV58b, 265726152 DNA Sequence	SEQ ID NO: 913	1309 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCTACTTCCTGGCTGGAGGGGACATGGTGTGGTGGCCAGTGGGTGCATCCTTGTTTGCCA
GCAATGTTGGAAGTGGACATTTCATTGGCCTGGCAGGGTCAGGTGCTGCTACGGGCATTTCTGTATCA
GCTTATGAACTTAATGGCTTGTTTTCTGTGCTGATGTTGGCCTGGATCTTCCTACCCATCTACATTGC
TGGTCAGGTCACCACGATGCCAGAATACCTACGGAAGCGCTTCGGTGGCATCAGAATCCCCATCATCC
TGGCTGTACTCTACCTATTTATCTACATCTTCACCAAGATCTCGGTAGACATGTATGCAGGTGCCATC
TTCATCCAGCAGTCTTTGCACCTGGATCTGTACCTGGCCATAGTTGGGCTACTGGCCATCACTGCTGT
ATACACGGTTGCTGGTGGCCTGGCTGCTGTGATCTACACGGATGCCCTGCAGACGCTGATCATGCTTA
TAGGAGCGCTCACCTTGATGGGCTACAGTTTCGCCGCGGTTGGTGGGATGGAAGGACTGAAGGAGAAG
TACTTCTTGGCCCTGGCTAGCAACCGGAGTGAGAACAGCAGCTGCGGGCTGCCCCGGGAAGATGCCTT
CCATATTTTCCGAGATCCGCTGACATCTGATCTCCCGTGGCCGGGGGTCCTATTTGGAATGTCCATCC
CATCCCTCTGGTACTGGTGCACGGATCAGGTGATTGTCCAGCGGACTCTGGCTGCCAAGAACCTGTCC
CATGCCAAAGGAGGTGCTCTGATGGCTGCATACCTGAAGGTGCTGCCCCTCTTCATAATGGTGTTCCC
TGGGATGGTCAGCCGCATCCTCTTCCCAGATCAAGTGGCCTGTGCAGATCCAGAGATCTGCCAGAAGA

TCTGCAGCAACCCCTCAGGCTGTTCGGACATCGCGTATCCCAAACTCGTGCTGGAACTCCTGCCCACA
GGGCTCCGTGGGCTGATGATGGCTGTGATGGTGGCGGCTCTCATGTCCTCCCTCACCTCCATCTTTAA
CAGTGCCAGCACCATCTTCACCATGGACCTCTGGAATCACCTCCGGCCTCGGGCATCTGAGAAGGAGC
TCATGATTGTGGGCAGGGTGTTTGTGCTGCTGCTGGTCCTGGTCTCCATCCTCTGGATCCCTGTGGTC
CAGGCCAGCCAGGGCGGCCAGCTCTTCATCTATATCCAGTCCATCAGCTCCTACCTGCAGCCGCCTGT
GGCGGTGGTCTTCATCATGGGATGTTTCTGGAAGAGGACCAATGAAAAGGGTGCCTTCTGGGGCCTGA
TCTCGGGCCTCGAGGGC

NOV58b, 265726152 Protein Sequence	SEQ ID NO: 914	436 aa	MW at 47358.5kD

TGSYFLAGGDMVWWPVGASLFASNVGSGHFIGLAGSGAATGISVSAYELNGLFSVLMLAWIFLPIYIA
GQVTTMPEYLRKRFGGIRIPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDLYLAIVGLLAITAV
YTVAGGLAAVIYTDALQTLIMLIGALTLMGYSFAAVGGMEGLKEKYFLALASNRSENSSCGLPREDAF
HIFRDPLTSDLPWPGVLFGMSIPSLWYWCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFP
GMVSRILFPDQVACADPEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFN
SASTIFTMDLWNHLRPRASEKELMIVGRVFVLLLVLVSILWIPVVQASQGGQLFIYIQSISSYLQPPV
AVVFIMGCFWKRTNEKGAFWGLISGLEG

NOV58c, 267254040 DNA Sequence	SEQ ID NO: 915	1912 bp
	ORF Start: at 2	ORF Stop: end of sequence

AGAATTCGCCTTCACCGGATCCAAAAGAGACACAGTGAAAGGCTACTTCCTGGCTGGAGGGGACATGG
TGTGGTGGCCAGTGGGTGCATCCTTGTTTGCCAGCAATGTTGGAAGTGGACATTTCATTGGCCTGGCA
GGGTCAGGTGCTGCTACGGGCATTTCTGTATCAGCTTATGAACTTAATGGCTTGTTTTCTGTGCTGAT
GTTGGCCTGGATCTTCCTACCCATCTACATTGCTGGTCAGGTCACCACGATGCCAGAATACCTACGGA
AGCGCTTCGGTGGCATCAGAATCCCCATCATCCTGGCTGTACTCTACCTATTTATCTACATCTTCACC
AAGATCTCGGTAGACATGTATGCAGGTGCCATCTTCATCCAGCAGTCTTTGCACCTGGATCTGTACCT
GGCCATAGTTGGGCTACTGGCCATCACTGCTGTATACACGGTTGCTGGTGGCCTGGCTGCTGTGATCT
ACACGGATGCCCTGCAGACGCTGATCATGCTTATAGGAGCGCTCACCTTGATGGGCTACAGTTTCGCC
GCGGTTGGTGGGATGGAAGGACTGAAGGAGAAGTACTTCTTGGCCCTGGCTAGCAACCGGAGTGAGAA
CAGCAGCTGCGGGCTGCCCCGGGAAGATGCCTTCCATATTTTCCGAGATCCGCTGACATCTGATCTCC
CGTGGCCGGGGGTCCTATTTGGAATGTCCATCCCATCCCTCTGGTACTGGTGCACGGATCAGGTGATT
GTCCAGCGGACTCTGGCTGCCAAGAACCTGTCCCATGCCAAAGGAGGTGCTCTGATGGCTGCATACCT
GAAGGTGCTGCCCCTCTTCATAATGGTGTTCCCTGGGATGGTCAGCCGCATCCTCTTCCCAGATCAAG
TGGCCTGTGCAGATCCAGAGATCTGCCAGAAGATCTGCAGCAACCCCTCAGGCTGTTCGGACATCGCG
TATCCCAAACTCGTGCTGGAACTCCTGCCCACAGGGCTCCGTGGGCTGATGATGGCTGTGATGGTGGC
GGCTCTCATGTCCTCCCTCACCTCCATCTTTAACAGTGCCAGCACCATCTTCACCATGGACCTCTGGA
ATCACCTCCGGCCTCGGGCATCTGAGAAGGAGCTCATGATTGTGGGCAGGGTGTTTGTGCTGCTGCTG
GTCCTGGTCTCCATCCTCTGGATCCCTGTGGTCCAGGCCAGCCAGGGCGGCCAGCTCTTCATCTATAT
CCAGTCCATCAGCTCCTACCTGCAGCCGCCTGTGGCGGTGGTCTTCATCATGGGATGTTTCTGGAAGA
GGACCAATGAAAAGGGTGCCTTCTGGGGCCTGATCTCGGGCCTGCTCCTGGGCTTGGTTAGGCTGGTC
CTGGACTTTATTTACGTGCAGCCTCGATGCGACCAGCCAGATGAGCGCCCGGTCCTGGTGAAGAGCAT
TCACTACCTCTACTTCTCCATGATCCTGTCCACGGTCACCCTCATCACTGTCTCCACCGTGAGCTGGT
TCACAGAGCCACCCTCCAAGGAGATGGTCAGCCACCTGACCTGGTTTACTCGTCACGACCCCGTGGTC
CAGAAGGAACAAGCACCACCAGCAGCTCCCTTGTCTCTTACCCTCTCTCAGAACGGGATGCCAGAGGC
CAGCAGCAGCAGCAGCGTCCAGTTCGAGATGGTTCAAGAAAACACGTCTAAAACCCACAGCTGTGACA
TGACCCCAAAGCAGTCCAAAGTGGTGAAGGCCATCCTGTGGCTCTGTGGAATACAGGAGAAGGGCAAG
GAAGAGCTCCCGGCCAGAGCAGAAGCCATCATAGTTTCCCTGGAAGAAAACCCCTTGGTGAAGACCCT
CCTGGACGTCAACCTCATTTTCTGCGTGAGCTGCGCCATCTTTATCTGGGGCTATTTTGCTCTCGAGG
GCAGGGCG

NOV58c, 267254040 Protein Sequence	SEQ ID NO: 916	637 aa	MW at 69945.6kD

EFAFTGSKRDTVKGYFLAGGDMVWWPVGASLFASNVGSGHFIGLAGSGAATGISVSAYELNGLFSVLM
LAWIFLPIYIAGQVTTMPEYLRKRFGGIRIPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDLYL
AIVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGYSFAAVGGMEGLKEKYFLALASNRSEN
SSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWYWCTDQVIVQRTLAAKNLSHAKGGALMAAYL
KVLPLFIMVFPGMVSRILFPDQVACADPEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVA
ALMSSLTSIFNSASTIFTMDLWNHLRPRASEKELMIVGRVFVLLLVLVSILWIPVVQASQGGQLFIYI
QSISSYLQPPVAVVFIMGCFWKRTNEKGAFWGLISGLLLGLVRLVLDFIYVQPRCDQPDERPVLVKSI

HYLYFSMILSTVTLITVSTVSWFTEPPSKEMVSHLTWFTRHDPVVQKEQAPPAAPLSLTLSQNGMPEA
SSSSSVQFEMVQENTSKTHSCDMTPKQSKVVKAILWLCGIQEKGKEELPARAEAIIVSLEENPLVKTL
LDVNLIFCVSCAIFIWGYFALEGRA

NOV58d, SNP13379242 of CG56398-01, DNA Sequence	SEQ ID NO: 917	2105 bp
	ORF Start: ATG at 31	ORF Stop: TAG at 2056
	SNP Pos: 804	SNP Change: C to G

CCTCAGGATCCAGAGGTCTCGTTCAGGACCATGGAGAGCGGCACCAGCAGCCCTCAGCCTCCACAGTT
AGATCCCCTGGATGCGTTTCCCCAGAAGGGCTTGGAGCCTGGGGACATCGCGGTGCTAGTTCTGTACT
TCCTCTTTGTCCTGGCTGTTGGACTATGGTCCACAGTGAAGACCAAAAGAGACACAGTGAAAGGCTAC
TTCCTGGCTGGAGGGGACATGGTGTGGTGGCCAGTGGGTGCATCCTTGTTTGCCAGCAATGTTGGAAG
TGGACATTTCATTGGCCTGGCAGGGTCAGGTGCTGCTACGGGCATTTCTGTATCAGCTTATGAACTTA
ATGGCTTGTTTTCTGTGCTGATGTTGGCCTGGATCTTCCTACCCATCTACATTGCTGGTCAGGTGACC
ACGATGCCAGAATACCTACGGAAGCGCTTCGGTGGCATCAGAATCCCCATCATCCTGGCTGTACTCTA
CCTATTTATCTACATCTTCACCAAGATCTCGGTAGACATGTATGCAGGTGCCATCTTCATCCAGCAGT
CTTTGCACCTGGATCTGTACCTGGCCATAGTTGGGCTACTGGCCATCACTGCTGTATACACGGTTGCT
GGTGGCCTGGCTGCTGTGATCTACACGGATGCCCTGCAGACGCTGATCATGCTTATAGGAGCGCTCAC
CTTGATGGGCTACAGTTTCGCCGCGGTTGGTGGGATGGAAGGACTGAAGGAGAAGTACTTCTTGGCCC
TGGCTAGCAACCGGAGTGAGAACAGCAGCTGCGGGCTGCCCCGGGAAGATGCCTTGCATATTTTCCGA
GATCCGCTGACATCTGATCTCCCGTGGCCGGGGGTCCTATTTGGAATGTCCATCCCATCCCTCTGGTA
CTGGTGCACGGATCAGGTAATTGTCCAGCGGACTCTGGCTGCCAAGAACCTGTCCCATGCCAAAGGAG
GTGCTCTGATGGCTGCATACCTGAAGGTGCTGCCCCTCTTCATAATGGTGTTCCCTGGGATGGTCAGC
CGCATCCTCTTCCCAGATCAAGTGGCCTGTGCAGATCCAGAGATCTGCCAGAAGATCTGCAGCAACCC
CTCAGGCTGTTCGGACATCGCGTATCCCAAACTCGTGCTGGAACTCCTGCCCACAGGTCTCCGTGGGC
TGATGATGGCTGTGATGGTGGCGGCTCTCATGTCCTCCCTCACCTCCATCTTTAACAGTGCCAGCACC
ATCTTCACCATGGACCTCTGGAATCACCTCCGGCCTCGGGCATCTGAGAAGGAGCTCATGATTGTGGG
CAGGGTGTTTGTGCTGCTGCTGGTCCTGGTCTCCATCCTCTGGATCCCTGTGGTCCAGGCCAGCCAGG
GCGGCCAGCTCTTCATCTATATCCAGTCCATCAGCTCCTACCTGCAGCCGCCTGTGGCGGTGGTCTTC
ATCATGGGATGTTTCTGGAAGAGGACCAATGAAAAGGGTGCCTTCTGGGGCCTGATCTCGGGCCTGCT
CCTGGGCTTGGTTAGGCTGGTCCTGGACTTTATTTACGTGCAGCCTCGATGCGACCAGCCAGATGAGC
GCCCGGTCCTGGTGAAGAGCATTCACTACCTCTACTTCTCCATGATCCTGTCCACGGTCACCCTCATC
ACTGTCTCCACCGTGAGCTGGTTCACAGAGCCACCCTCCAAGGAGATGGTCAGCCACCTGACCTGGTT
TACTCGTCACGACCCCGTGGTCCAGAAGGAACAAGCACCACCAGCAGCTCCCTTGTCTCTTACCCTCT
CTCAGAACGGGATGCCAGAGGCCAGCAGCAGCAGCAGCGTCCAGTTCGAGATGGTTCAAGAAAACACG
TCTAAAACCCACAGCGGTGACATGACCCCAAAGCAGTCCAAAGTGGTGAAGGCCATCCTGTGGCTCTG
TGGAATACAGGAGAAGGGCAAGGAAGAGCTCCCGGCCAGAGCAGAAGCCATCATAGTTTCCCTGGAAG
AAAACCCCTTGGTGAAGACCCTCCTGGACGTCAACCTCATTTTCTGCGTGAGCTGCGCCATCTTTATC
TGGGGCTATTTTGCTTAGTGTGGGGTGAACCCAGGGGTCCAAACTCTGTTTCTCTTCAGTGCTCC

NOV58d, SNP13379242 of CG56398-01, Protein Sequence	SEQ ID NO: 918	675 aa	MW at 73955.2kD
	SNP Pos: 258		SNP Change: Phe to Leu

MESGTSSPQPPQLDPLDAFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTKRDTVKGYFLAGGDMVWW
PVGASLFASNVGSGHFIGLAGSGAATGISVSAYELNGLFSVLMLAWIFLPIYIAGQVTTMPEYLRKRF
GGIRIPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDLYLAIVGLLAITAVYTVAGGLAAVIYTD
ALQTLIMLIGALTLMGYSFAAVGGMEGLKEKYFLALASNRSENSSCGLPREDALHIFRDPLTSDLPWP
GVLFGMSIPSLWYWCTDQVIVQRTLAAKNLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVAC
ADPEICQKICSNPSGCSDIAYPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTMDLWNHL
RPRASEKELMIVGRVFVLLLVLVSILWIPVVQASQGGQLFIYIQSISSYLQPPVAVVFIMGCFWKRTN
EKGAFWGLISGLLLGLVRLVLDFIYVQPRCDQPDERPVLVKSIHYLYFSMILSTVTLITVSTVSWFTE
PPSKEMVSHLTWFTRHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQENTSKTHSGDMTP
KQSKVVKAILWLCGIQEKGKEELPARAEAIIVSLEENPLVKTLLDVNLIFCVSCAIFIWGYFA

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 58B.

Table 58B. Comparison of the NOV58 protein sequences.

```
NOV58a  MESGTSSPQPPQLDPLDAFPQKGLEPGDIAVLVLYFLFVLAVGLWSTVKTKRDTVKGYFL
NOV58b  ----------------------------------------------------TGSYFL
NOV58c  ------------------------------------------EFAFTGSKRDTVKGYFL

NOV58a  AGGDMVWWPVGASLFASNVGSGHFIGLAGSGAATGISVSAYELNGLFSVLMLAWIFLPIY
NOV58b  AGGDMVWWPVGASLFASNVGSGHFIGLAGSGAATGISVSAYELNGLFSVLMLAWIFLPIY
NOV58c  AGGDMVWWPVGASLFASNVGSGHFIGLAGSGAATGISVSAYELNGLFSVLMLAWIFLPIY

NOV58a  IAGQVTTMPEYLRKRFGGIRIPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDLYLA
NOV58b  IAGQVTTMPEYLRKRFGGIRIPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDLYLA
NOV58c  IAGQVTTMPEYLRKRFGGIRIPIILAVLYLFIYIFTKISVDMYAGAIFIQQSLHLDLYLA

NOV58a  IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGYSFAAVGGMEGLKEKYFLAL
NOV58b  IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGYSFAAVGGMEGLKEKYFLAL
NOV58c  IVGLLAITAVYTVAGGLAAVIYTDALQTLIMLIGALTLMGYSFAAVGGMEGLKEKYFLAL

NOV58a  ASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWYWCTDQVIVQRTLAAK
NOV58b  ASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWYWCTDQVIVQRTLAAK
NOV58c  ASNRSENSSCGLPREDAFHIFRDPLTSDLPWPGVLFGMSIPSLWYWCTDQVIVQRTLAAK

NOV58a  NLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVACADPEICQKICSNPSGCSDIA
NOV58b  NLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVACADPEICQKICSNPSGCSDIA
NOV58c  NLSHAKGGALMAAYLKVLPLFIMVFPGMVSRILFPDQVACADPEICQKICSNPSGCSDIA
                                                                    .
NOV58a  YPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTMDLWNHLRPRASEKELMIV
NOV58b  YPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTMDLWNHLRPRASEKELMIV
NOV58c  YPKLVLELLPTGLRGLMMAVMVAALMSSLTSIFNSASTIFTMDLWNHLRPRASEKELMIV

NOV58a  GRVFVLLLVLVSILWIPVVQASQGGQLFIYIQSISSYLQPPVAVVFIMGCFWKRTNEKGA
NOV58b  GRVFVLLLVLVSILWIPVVQASQGGQLFIYIQSISSYLQPPVAVVFIMGCFWKRTNEKGA
NOV58c  GRVFVLLLVLVSILWIPVVQASQGGQLFIYIQSISSYLQPPVAVVFIMGCFWKRTNEKGA

NOV58a  FWGLISGLLLGLVRLVLDFIYVQPRCDQPDERPVLVKSIHYLYFSMILSTVTLITVSTVS
NOV58b  FWGLISGLEG----------------------------------------------------
NOV58c  FWGLISGLLLGLVRLVLDFIYVQPRCDQPDERPVLVKSIHYLYFSMILSTVTLITVSTVS

NOV58a  WFTEPPSKEMVSHLTWFTRHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQE
NOV58b  ----------------------------------------------------------
NOV58c  WFTEPPSKEMVSHLTWFTRHDPVVQKEQAPPAAPLSLTLSQNGMPEASSSSSVQFEMVQE

NOV58a  NTSKTHSGDMTPKQSKVVKAILWLCGIQEKGKEELPARAEAIIVSLEENPLVKTLLDVNL
NOV58b  ----------------------------------------------------------
NOV58c  NTSKTHSCDMTPKQSKVVKAILWLCGIQEKGKEELPARAEAIIVSLEENPLVKTLLDVNL

NOV58a  IFCVSCAIFIWGYFA-----
NOV58b  --------------------
NOV58c  IFCVSCAIFIWGYFALEGRA

NOV58a  (SEQ ID NO:  912)
NOV58b  (SEQ ID NO:  914)
NOV58c  (SEQ ID NO:  916)
```

Further analysis of the NOV58a protein yielded the following properties shown in Table 58C.

Table 58C. Protein Sequence Properties NOV58a

SignalP analysis:	Cleavage site between residues 51 and 52

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 2;  pos.chg 0;  neg.chg 1
      H-region:  length 11;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -0.94
      possible cleavage site: between 41 and 42

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:  11
      INTEGRAL    Likelihood =-10.61   Transmembrane   29 -  45
      INTEGRAL    Likelihood = -6.64   Transmembrane  106 - 122
      INTEGRAL    Likelihood = -7.64   Transmembrane  139 - 155
      INTEGRAL    Likelihood = -5.79   Transmembrane  177 - 193
      INTEGRAL    Likelihood = -3.24   Transmembrane  210 - 226
      INTEGRAL    Likelihood = -4.88   Transmembrane  309 - 325
      INTEGRAL    Likelihood = -3.88   Transmembrane  376 - 392
      INTEGRAL    Likelihood =-13.69   Transmembrane  423 - 439
      INTEGRAL    Likelihood = -5.47   Transmembrane  484 - 500
      INTEGRAL    Likelihood = -0.53   Transmembrane  524 - 540
      INTEGRAL    Likelihood = -5.47   Transmembrane  654 - 670
      PERIPHERAL  Likelihood =  1.32 (at 454)
      ALOM score: -13.69  (number of TMSs: 11)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 36
      Charge difference:  6.0   C( 3.0) - N(-3.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  4.59
      Hyd Moment(95):  4.89     G content:       1
      D/E content:     2        S/T content:     4
      Score: -6.67

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found
```

```
NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: PEYLRKR (3) at  129
      bipartite: none
      content of basic residues:   6.7%
      NLS Score: -0.22

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       11.1 %: vacuolar
       11.1 %: mitochondrial
       11.1 %: Golgi
```

```
>> prediction for CG56398-01 is end (k=9)
```

A search of the NOV58a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 58D.

Table 58D. Geneseq Results for NOV58a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV58a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB80599	Human sbg1007026SGLT protein - Homo sapiens, 675 aa. [WO200222802-A1, 21-MAR-2002]	1..675 1..675	674/675 (99%) 674/675 (99%)	0.0
AAB60093	Human transport protein TPPT-13 - Homo sapiens, 675 aa. [WO200078953-A2, 28-DEC-2000]	1..675 1..675	674/675 (99%) 674/675 (99%)	0.0
AAB85102	Novel human transporter protein (NHP) - Homo sapiens, 675 aa. [WO200142469-A1, 14-JUN-2001]	1..675 1..675	673/675 (99%) 673/675 (99%)	0.0
ABP69833	Human polypeptide SEQ ID NO 1880 - Homo sapiens, 675 aa. [WO200270539-A2, 12-SEP-2002]	1..675 1..675	672/675 (99%) 672/675 (99%)	0.0
AAU77134	Human sodium-sugar symporter 32620 polypeptide - Homo sapiens, 675 aa. [WO200216582-A2, 28-FEB-2002]	1..675 1..675	671/675 (99%) 672/675 (99%)	0.0

In a BLAST search of public sequence databases, the NOV58a protein was found to have homology to the proteins shown in the BLASTP data in Table 58E.

Table 58E. Public BLASTP Results for NOV58a				
Protein Accession Number	**Protein/Organism/Length**	**NOV58a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8WWX8	Sodium/glucose cotransporter KST1 - Homo sapiens (Human), 675 aa.	1..675 1..675	674/675 (99%) 674/675 (99%)	0.0
CAC83728	Sodium-dependent glucose transporter - Homo sapiens (Human), 675 aa.	1..675 1..675	673/675 (99%) 673/675 (99%)	0.0
Q96PP5	Putative sodium-coupled cotransporter RKST1 - Homo sapiens (Human), 675 aa.	1..675 1..675	672/675 (99%) 672/675 (99%)	0.0
CAD43585	Sequence 1 from Patent WO0216582 - Homo sapiens (Human), 675 aa.	1..675 1..675	671/675 (99%) 672/675 (99%)	0.0
Q8K0E3	Similar to putative sodium-coupled cotransporter RKST1 - Mus musculus (Mouse), 673 aa.	1..675 1..673	583/675 (86%) 627/675 (92%)	0.0

PFam analysis indicates that the NOV58a protein contains the domains shown in the Table 58F.

Table 58F. Domain Analysis of NOV58a			
Pfam Domain	**NOV58a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
SSF	58..487	199/447 (45%) 370/447 (83%)	1.9e-199

Example 59.

The NOV59 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 59A.

Table 59A. NOV59 Sequence Analysis		
NOV59a, CG56605-03 DNA Sequence	SEQ ID NO: 919	557 bp
	ORF Start: ATG at 26	ORF Stop: TAG at 524
ATGGTGGCAGAGTCTACATAGAACTATGCTTCGTGGTGTTCTGGGGAAAACCTTTCGACTTGTTGGCT		

ATACTATTCAATATGGCTGTATAGCTCATTGTGCTTTTGAATACGTTGGTGGTGTTGTCATGTGTTCT
GGACCATCAATGGAGCCTACAATTCAAAATTCAGATATTGTCTTTGCAGAAAATCTTAGTCGACATTT
TTATGGTATCCAAAGAGGTGACATTGTGATTGCAAAAAGCCCAAGTGATCCAAAATCAAATATTTGTA
AAAGAGTAATTGGTTTGGAAGGAGACAAAATCCTCACCACTAGTCCATCAGATTTCTTTAAAAGCCAT
AGTTATGTGCCAATGGGTCATGTTTGGTTAGAAGGTGACAATCTACAGAATTCTACAGATTCCAGGTG
CTATGGACCTATTCCATATGGACTAATAAGAGGACGAATCTTCTTTAAGATTTGGCCTCTGAGTGATT
TTGGATTTTTACGTGCCAGCCCTAATGGCCACAGATTTTCTGATGATTAGTAAGCATTTATTCTTTTG
ACTTGATTATTGT

NOV59a, CG56605-03 Protein Sequence	SEQ ID NO: 920	166 aa	MW at 18504.0kD

MLRGVLGKTFRLVGYTIQYGCIAHCAFEYVGGVVMCSGPSMEPTIQNSDIVFAENLSRHFYGIQRGDI
VIAKSPSDPKSNICKRVIGLEGDKILTTSPSDFFKSHSYVPMGHVWLEGDNLQNSTDSRCYGPIPYGL
IRGRIFFKIWPLSDFGFLRASPNGHRFSDD

NOV59b, CG56605-01 DNA Sequence	SEQ ID NO: 921	513 bp
	ORF Start: ATG at 16	ORF Stop: TAA at 445

AGTCTACATAGAACTATGCTTCGTGGTGCTCTGGGGAAACCTTCTGACCTTGTTGGCTATGCTATTCA
GTATGGCTGTATAGCTCACTGTGCTTCTGAATACGTTGGTGGTGTTGTCATGTGTTCTGGACCATCAA
TGGAGCCTACAATTCAAAATTCAGATACTGTCTTTGCACAAAATCTTAGTCGACATTTTGATAGTATC
CAAAGAGGTGACATTGTGATTGCAAAAAGCCCAAGTGATCCAACATCAAATATTTGTAAAAGAGTAAC
TGGTTTGGAAGGAGACAAAATCCTCACCACTAGTCCATCAGATTTCTTTAAAAGCTACAGTTATGTCC
CAGTGGGTCATGTTTGGTTAGAAGGTGATAATCTACAGAATTCTACAGATTCCAGTTACTATGGACCT
ATTCCATATGAACTAATAAGAGGACGAATCTTCTTTTAAGATTCGGCCTCTGAGTGATTTTGGATTTT
TATGTGCCAGCCTTAATGGCCACAGATTTTCTGATGA

NOV59b, CG56605-01 Protein Sequence	SEQ ID NO: 922	143 aa	MW at 15622.4kD

MLRGALGKPSDLVGYAIQYGCIAHCASEYVGGVVMCSGPSMEPTIQNSDTVFAQNLSRHFDSIQRGDI
VIAKSPSDPTSNICKRVTGLEGDKILTTSPSDFFKSYSYVPVGHVWLEGDNLQNSTDSSYYGPIPYEL
IRGRIFF

NOV59c, CG56605-02 DNA Sequence	SEQ ID NO: 923	507 bp
	ORF Start: ATG at 11	ORF Stop: at 506

ACATAGAACTATGCTTCGTGGTGCTCTGGGGAAAACCTTTCGACTTGTTGGCTATACTATTCAATATG
GCTGTATAGCTCATTGTGCTTTTGAATACGTTGGTGGTGTTGTCATGTGTTCTGGACCATCAATGGAG
CCTACAATTCAAAATTCAGATATTGTCTTTGCAGAAAATCTTAGTCGACATTTTTATGGTATCCAAAG
AGGTGACATTGTGATTGCAAAAAGCCCAAGTGATCCAAAATCAAATATTTGTAAAAGAGTAATTGGTT
TGGAAGGAGACAAAATCCTCACCACTAGTCCATCAGATTTCTTTAAAAGCCATAGTTATGTGCCAATG
GGTCATGTTTGGTTAGAAGGTGACAATCTACAGAATTCTACAGATTCCAGGTGCTATGGACCTATTCC
ATATGGACTAATAAGAGGACGAATCTTCTTTAAGATTTGGCCTCTGAGTGATTTTGGATTTTTACGTG
CCAGCCTTAATGGCCACAGATTTTCTGATGA

NOV59c, CG56605-02 Protein Sequence	SEQ ID NO: 924	165 aa	MW at 18376.9kD

MLRGALGKTFRLVGYTIQYGCIAHCAFEYVGGVVMCSGPSMEPTIQNSDIVFAENLSRHFYGIQRGDI
VIAKSPSDPKSNICKRVIGLEGDKILTTSPSDFFKSHSYVPMGHVWLEGDNLQNSTDSRCYGPIPYGL
IRGRIFFKIWPLSDFGFLRASLNGHRFSD

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 59B.

Table 59B. Comparison of the NOV59 protein sequences.

NOV59a	MLRGVLGKTFRLVGYTIQYGCIAHCAFEYVGGVVMCSGPSMEPTIQNSDIVFAENLSRHF
NOV59b	MLRGALGKPSDLVGYAIQYGCIAHCASEYVGGVVMCSGPSMEPTIQNSDTVFAQNLSRHF

```
NOV59c  MLRGALGKTFRLVGYTIQYGCIAHCAFEYVGGVVMCSGPSMEPTIQNSDIVFAENLSRHF

NOV59a  YGIQRGDIVIAKSPSDPKSNICKRVIGLEGDKILTTSPSDFFKSHSYVPMGHVWLEGDNL
NOV59b  DSIQRGDIVIAKSPSDPTSNICKRVTGLEGDKILTTSPSDFFKSYSYVPVGHVWLEGDNL
NOV59c  YGIQRGDIVIAKSPSDPKSNICKRVIGLEGDKILTTSPSDFFKSHSYVPMGHVWLEGDNL

NOV59a  QNSTDSRCYGPIPYGLIRGRIFFKIWPLSDFGFLRASPNGHRFSDD
NOV59b  QNSTDSSYYGPIPYELIRGRIFF----------------------
NOV59c  QNSTDSRCYGPIPYGLIRGRIFFKIWPLSDFGFLRASLNGHRFSD-

NOV59a  (SEQ ID NO:  920)
NOV59b  (SEQ ID NO:  922)
NOV59c  (SEQ ID NO:  924)
```

Further analysis of the NOV59a protein yielded the following properties shown in Table 59C.

Table 59C. Protein Sequence Properties NOV59a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 3;  neg.chg 0
      H-region:  length 16;  peak value   6.29
      PSG score:   1.89

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -7.69
      possible cleavage site: between 32 and 33

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  1.70 (at 20)
      ALOM score:   1.70  (number of TMSs: 0)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 6
      Charge difference:  0.0   C( 2.0) - N( 2.0)
      N >= C: N-terminal side will be inside

MITDISC: discrimination of mitochondrial targeting seq
      R content:       2       Hyd Moment(75): 14.36
      Hyd Moment(95):  9.31    G content:      4
      D/E content:     1       S/T content:    2
      Score: -2.96

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 21  FRL|VG
```

```
NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  10.2%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      XXRR-like motif in the N-terminus: LRGV

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 89

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):
```

```
       43.5 %: cytoplasmic
       34.8 %: mitochondrial
        8.7 %: nuclear
        8.7 %: peroxisomal
        4.3 %: plasma membrane

>> prediction for CG56605-03 is cyt (k=23)
```

A search of the NOV59a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 59D.

5

Table 59D. Geneseq Results for NOV59a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV59a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG70899	Human S26 signal peptidase family member 33449 - Homo sapiens, 166 aa. [US2002115630-A1, 22-AUG-2002]	1..166 1..166	166/166 (100%) 166/166 (100%)	2e-97
AAB47563	Protease PRTS-5 - Homo sapiens, 166 aa. [WO200171004-A2, 27-SEP-2001]	1..166 1..166	166/166 (100%) 166/166 (100%)	2e-97
ABG08149	Novel human diagnostic protein #8140 - Homo sapiens, 166 aa. [WO200175067-A2, 11-OCT-2001]	1..166 1..166	166/166 (100%) 166/166 (100%)	2e-97
ABB64326	Drosophila melanogaster polypeptide SEQ ID NO 19770 - Drosophila melanogaster, 166 aa. [WO200171042-A2, 27-SEP-2001]	5..150 3..162	76/160 (47%) 100/160 (62%)	2e-38
AAB74688	Human protease and protease inhibitor PPIM-21 - Homo sapiens, 94 aa. [WO200110903-A2, 15-FEB-2001]	108..166 36..94	58/59 (98%) 58/59 (98%)	6e-30

In a BLAST search of public sequence databases, the NOV59a protein was found to have homology to the proteins shown in the BLASTP data in Table 59E.

Table 59E. Public BLASTP Results for NOV59a				
Protein Accession Number	**Protein/Organism/Length**	**NOV59a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96LU5	Hypothetical protein FLJ25059 - Homo sapiens (Human), 166 aa.	1..166 1..166	166/166 (100%) 166/166 (100%)	5e-97
Q9CQU8	1500034J20Rik protein (2610528O17Rik protein) (RIKEN cDNA 1500034J20 gene) (Unknown EST) - Mus musculus (Mouse), 166 aa.	1..166 1..166	156/166 (93%) 157/166 (93%)	1e-89
Q96SH9	DJ1137O17.1 (Similar to putative mitochondrial inner membrane protease subnunit 2) - Homo sapiens (Human), 144 aa (fragment).	1..144 1..144	144/144 (100%) 144/144 (100%)	2e-82
Q9VXR8	CG9240 protein - Drosophila melanogaster (Fruit fly), 166 aa.	5..150 3..162	76/160 (47%) 100/160 (62%)	5e-38
Q8SZ24	RE22928p - Drosophila melanogaster (Fruit fly), 166 aa.	5..150 3..162	76/160 (47%) 99/160 (61%)	1e-37

PFam analysis indicates that the NOV59a protein contains the domains shown in the Table 59F.

Table 59F. Domain Analysis of NOV59a			
Pfam Domain	**NOV59a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Peptidase_S26	1..152	37/221 (17%) 104/221 (47%)	9.1e-07

Example 60.

The NOV60 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 60A.

Table 60A. NOV60 Sequence Analysis

NOV60a, CG56645-03 DNA Sequence	SEQ ID NO: 925	1862 bp
	ORF Start: ATG at 35	ORF Stop: TGA at 1829

CGGGCTCATACCTAGTGCCTGCGGCAGGACAGCC**ATG**GCCGCCAACTCCACCAGCGACCTCCACACTC
CCGGGACGCAGCTGAGCGTGGCTGACATCATCGTCATCACTGTGTATTTTGCTCTGAACGTGGCCGTG
GGCATATGGTCCTCTTGTCGGGCCAGTAGGAACACGGTGAATGGCTACTTCCTGGCAGGCCGGGACAT
GACGTGGTGGCCGATTGGAGCCTCCCTCTTCGCCAGCAGCGAGGGCTCTGGCCTCTTCATTGGACTGG
CGGGCTCAGGCGCGGCAGGAGGTCTGGCCGTGGCAGGCTTCGAGTGGAATGCCACGTACGTGCTGCTG
GCACTGGCATGGGTGTTCGTGCCCATCTACATCTCCTCAGAGATCGTCACCTTACCTGAGTACATTCA
GAAGCGCTACGGGGGCCAGCGGATCCGCATGTACCTGTCTGTCCTGTCCCTGCTACTGTCTGTCTTCA
CCAAGATATCGCTGGACCTGTACGCGGGGGCTCTGTTTGTGCACATCTGCCTGGGCTGGAACTTCTAC
CTCTCCACCATCCTCACGCTCGGCATCACAGCCCTGTACACCATCGCAGGGGGCCTGGCTGCTGTAAT
CTACACGGACGCCCTGCAGACGCTCATCATGGTGGTGGGGGCTGTCATCCTGACAATCAAAGCTTTTG
ACCAGATCGGTGGTTACGGGCAGCTGGAGGCAGCCTACGCCCAGGCCATTCCCTCCAGGACCATTGCC
AACACCACCTGCCACCTGCCACGTACAGACGCCATGCACATGTTTCGAGACCCCCACACAGGGGACCT
GCCGTGGACCGGGATGACCTTTGGCCTGACCATCATGGCCACCTGGTACTGGTGCACCGACCAGGTCA
TCGTGCAGCGATCACTGTCAGCCCGGGACCTGAACCATGCCAAGGCGGGCTCCATCCTGGCCAGCTAC
CTCAAGATGCTCCCCATGGGCCTGATCATCATGCCGGGCATGATCAGCCGCGCATTGTTCCCAGGTGC
TCATGTCTATGAGGAGAGACACCAAGTGTCCGTCTCTCGAACAGATGATGTGGGCTGCGTGGTGCCGT
CCGAGTGCCTGCGGGCCTGCGGGGCCGAGGTCGGCTGCTCCAACATCGCCTACCCCAAGCTGGTCATG
GAACTGATGCCCATCGGTCTGCGGGGGCTGATGATCGCAGTGATGCTGGCGGCGCTCATGTCGTCGCT
GACCTCCATCTTCAACAGCAGCAGCACCCTCTTCACTATGGACATCTGGAGGCGGCTGCGTCCCCGCT
CCGGCGAGCGGGAGCTCCTGCTGGTGGGACGGCTGGTCATAGTGGCACTCATCGGCGTGAGTGTGGCC
TGGATCCCCGTCCTGCAGGACTCCAACAGCGGGCAACTCTTCATCTACATGCAGTCAGTGACCAGCTC
CCTGGCCCCACCAGTGACTGCAGTCTTTGTCCTGGGCCTGATAGCAGGGCTGGTGGTGGGGGCCACGA
GGCTGGTCCTGGAATTCCTGAACCCAGCCCCACCGTGCGGAGAGCCAGACACGCGGCCAGCCGTCCTG
GGGAGCATCCACTACCTGCACTTCGCTGTCGCCCTCTTTGCACTCAGTGGTGCTGTTGTGGTGGCTGG
AAGCCTGCTGACCCCACCCCCACAGAGTGTCCAGATTGAGAACCTTACCTGGTGGAACCTGGCTCAGG
ATGTGCCCTTGGGAACTAAAGCAGGTGATGGCCAAACACCCCAGAAACACGCCTTCTGGGCCCGTGTC
TGTGGCTTCAATGCCATCCTCCTCATGTGTGTCAACATATTCTTTTATGCCTACTTCGCC**TGA**CACTG
CCATCCTGGACAGAAAGGCAGGAGCT

NOV60a, CG56645-03 Protein Sequence	SEQ ID NO: 926	598 aa	MW at 64485.7kD

MAANSTSDLHTPGTQLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGYFLAGRDMTWWPIGASLFA
SSEGSGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEIVTLPEYIQKRYGGQRIRMY
LSVLSLLLSVFTKISLDLYAGALFVHICLGWNFYLSTILTLGITALYTIAGGLAAVIYTDALQTLIMV
VGAVILTIKAFDQIGGYGQLEAAYAQAIPSRTIANTTCHLPRTDAMHMFRDPHTGDLPWTGMTFGLTI
MATWYWCTDQVIVQRSLSARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFPGAHVYEERHQVSV
SRTDDVGCVVPSECLRACGAEVGCSNIAYPKLVMELMPIGLRGLMIAVMLAALMSSLTSIFNSSSTLF
TMDIWRRLRPRSGERELLLVGRLVIVALIGVSVAWIPVLQDSNSGQLFIYMQSVTSSLAPPVTAVFVL
GLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYLHFAVALFALSGAVVVAGSLLTPPPQSVQ
IENLTWWNLAQDVPLGTKAGDGQTPQKHAFWARVCGFNAILLMCVNIFFYAYFA

NOV60b, CG56645-04 DNA Sequence	SEQ ID NO: 927	1801 bp
	ORF Start: ATG at 35	ORF Stop: TGA at 1790

CGGGCTCATACCTAGTGCCTGCGGCAGGACAGCC**ATG**GCCGCCAACTCCACCAGCGACCTCCACACTC
CCGGGACGCAGCTGAGCGTGGCTGACATCATCGTCATCACTGTGTATTTTGCTCTGAACGTGGCCGTG
GGCATATGGTCCTCTTGTCGGGCCAGTAGGAACACGGTGAATGGCTACTTCCTGGCAGGCCGGGACAT
GACGTGGTGGCCGATTGGAGCCTCCCTCTTCGCCAGCAGCGAGGGCTCTGGCCTCTTCATTGGACTGG
CGGGCTCAGGCGCGGCAGGAGGTCTGGCCGTGGCAGGCTTCGAGTGGAATGCCACGTACGTGCTGCTG
GCACTGGCATGGGTGTTCGTGCCCATCTACATCTCCTCAGAGATCGTCACCTTACCTGAGTACATTCA
GAAGCGCTACGGGGGCCAGCGGATCCGCATGTACCTGTCTGTCCTGTCCCTGCTACTGTCTGTCTTCA
CCAAGATATCGCTGGACCTGTACGCGGGGGCTCTGTTTGTGCACATCTGCCTGGGCTGGAACTTCTAC
CTCTCCACCATCCTCACGCTCGGCATCACAGCCCTGTACACCATCGCAGCTTTTGACCAGATCGGTGG

TTACGGGCAGCTGGAGGCAGCCTACGCCCAGGCCATTCCCTCCAGGACCATTGCCAACACCACCTGCC
AGCTGCCACGTACAGACGCCATGCACATGTTTCGAGACCCCCACACAGGGGACCTGCCGTGGACCGGG
ATGACCTTTGGCCTGACCATCATGGCCACCTGGTACTGGTGCACCGACCAGGTCATCGTGGAGGGATC
ACTGTCAGCCCGGGACCTGAACCATGCCAAGGGGGGCTCCATCCTGGGCAGCTACCTCAAGATGCTCC
CCATGGGCCTGATCATCATGCCGGGCATGATCAGCCGCGCATTGTTCCCAGGTGCTCATGTCTATGAG
GAGAGACACCAAGTGTCCGTCTCTCGAACAGATGATGTGGGCTGCGTGGTGCCGTCCGAGTGCCTGCG
GGCCTGCGGGGCCGAGGTCGGCTGCTCCAACATCGCCTACCCCAAGCTGGTCATGGAACTGATGCCCA
TCGGTCTGCGGGGGCTGATGATCGCAGTGATGCTGGCGGCGCTCATGTCGTCGCTGACCTCCATCTTC
AACAGCAGCAGCACCCTCTTCACTATGGACATCTGGAGGCGGCTGCGTCCCCGCTCCGGCGAGCGGGA
GCTCCTGCTGGTGGGACGGCTGGTCATAGTGGCACTCATCGGCGTGAGTGTGGCCTGGATCCCCGTCC
TGCAGGACTCCAACAGCGGGCAACTCTTCATCTACATGCAGTCAGTGACCAGCTCCCTGGCCCCACCA
GTGACTGCAGTCTTTGTCCTGGGCGTCTTCTGGCGACGTGCCAACGAGCAGGGGGCCTTCTGGGGCCT
GATAGCAGGGCTGGTGGTGGGGGCCACGAGGCTGGTCCTGGAATTCCTGAACCCGGCCCCACCGTGCG
GAGAGCCAGACACGCGGCCAGCCGTCCTGGGGAGCATCCACTACCTGCACTTCGCTGTCGCCCTCTTT
GCACTCAGTGGTGCTGTTGTGGTGGCTGGAAGCCTGCTGACCCCACCCCCACAGAGTGTCCAGATTGA
GAACCTTACCTGGTGGACCCTGGCTCAGGATGTGCCCTTGGGAACTAAAGCAGGTGATGGCCAAACAC
CCCAGAAACACGCCTTCTGGGCCCGTGTCTGTGGCTTCAATGCCATCCTCCTCATGTGTGTCAACATA
TTCTTTTATGCCTACTTCGCCTGACACTGCCAT

NOV60b, CG56645-04 Protein Sequence	SEQ ID NO: 928	585 aa	MW at 63317.1kD

MAANSTSDLHTPGTQLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGYFLAGRDMTWWPIGASLFA
SSEGSGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEIVTLPEYIQKRYGGQRIRMY
LSVLSLLLSVFTKISLDLYAGALFVHICLGWNFYLSTILTLGITALYTIAAFDQIGGYGQLEAAYAQA
IPSRTIANTTCQLPRTDAMHMFRDPHTGDLPWTGMTFGLTIMATWYWCTDQVIVEGSLSARDLNHAKG
GSILGSYLKMLPMGLIIMPGMISRALFPGAHVYEERHQVSVSRTDDVGCVVPSECLRACGAEVGCSNI
AYPKLVMELMPIGLRGLMIAVMLAALMSSLTSIFNSSSTLFTMDIWRRLRPRSGERELLLVGRLVIVA
LIGVSVAWIPVLQDSNSGQLFIYMQSVTSSLAPPVTAVFVLGVFWRRANEQGAFWGLIAGLVVGATRL
VLEFLNPAPPCGEPDTRPAVLGSIHYLHFAVALFALSGAVVVAGSLLTPPPQSVQIENLTWWTLAQDV
PLGTKAGDGQTPQKHAFWARVCGFNAILLMCVNIFFYAYFA

NOV60c, CG56645-01 DNA Sequence	SEQ ID NO: 929	1914 bp
	ORF Start: ATG at 51	ORF Stop: TGA at 1839

TTGCCCCTCAGTCCCTCGGGCTCATACCTAGTGCCTGCGGCAGGACAGCCATGGCCGCCAACTCCACC
AGCGACCTCCACACTCCCGGGACGCAGCTGAGCGTGGCTGACATCATCGTCATCACTGTGTATTTTGC
TCTGAACGTGGCCGTGGGCATATGGTCCTCTTGTCGGGCCAGTAGGAACACGGTGAATGGCTACTTCC
TGGCAGGCCGGGACATGACGTGGTGGCCGATTGGAGCCTCCCTCTTCGCCAGCAGCGAGGGCTCTGGC
CTCTTCATTGGACTGGCGGGCTCAGGCGCGGCAGGAGGTCTGGCCGTGGCAGGCTTCGAGTGGAATGC
CACGTACGTGCTGCTGGCACTGGCATGGGTGTTCGTGCCCATCTACATCTCCTCAGAGATCGTCACCT
TACCTGAGTACATTCAGAAGCGCTACGGGGGCCAGCGGATCCGCATGTACCTGTCTGTCCTGTCCCTG
CTACTGTCTGTCTTCACCAAGATATCGCTGGACCTGTACGCGGGGGCTCTGTTTGTGCACATCTGCCT
GGGCTGGAACTTCTACCTCTCCACCATCCTCACGCTCGGCATCACAGCCCTGTACACCATCGCAGGGG
GCCTGGCTGCTGTAATCTACACGGACGCCCTGCAGACGCTCATCATGGTGGTGGGGGCTGTCATCCTG
ACAATCAAAGCTTTTGACCAGATCGGTGGTTACGGGCAGCTGGAGGCAGCCTACGCCCAGGCCATTCC
CTCCAGGACCATTGCCAACACCACCTGCCACCTGCCACGTACAGACGCCATGCACATGTTTCGAGACC
CCCACACAGGGGACCTGCCGTGGACCGGGATGACCTTTGGCCTGACCATCATGGCCACCTGGTACTGG
TGCACCGACCAGGTGATCGTGCAGCGATCACTGTCAGCCCGGGACCTGAACCATGCCAAGGCGGGCTC
CATCCTGGCCAGCTACCTCAAGATGCTCCCCATGGGCCTGATCATAATGCCGGGCATGATCAGCCGCG
CATTGTTCCCAGATGATGTGGGCTGCGTGGTGCCGTCCGAGTGCCTGCGGGCCTGCGGGGCCGAGGTC
GGCTGCTCCAACATCGCCTACCCCAAGCTGGTCATGGAACTGATGCCCATCGGTCTGCGGGGGCTGAT
GATCGCAGTGATGCTGGCGGCGCTCATGTCGTCGCTGACCTCCATCTTCAACAGCAGCAGCACCCTCT
TCACTATGGACATCTGGAGGCGGCTGCGTCCCCGCTCCGGCGAGCGGGAGCTCCTGCTGGTGGGACGG
CTGGTCATAGTGGCACTCATCGGCGTGAGTGTGGCCTGGATCCCCGTCCTGCAGGACTCCAACAGCGG
GCAACTCTTCATCTACATGCAGTCAGTGACCAGCTCCCTGGCCCCACCAGTGACTGCAGTCTTTGTCC
TGGGCGTCTTCTGGCGACGTGCCAACGAGCAGGGGGCCTTCTGGGGCCTGATAGCAGGGCTGGTGGTG
GGGGCCACGAGGCTGGTCCTGGAATTCCTGAACCCAGCCCCACCGTGCGGAGAGCCAGACACGCGGCC
AGCCGTCCTGGGGAGCATCCACTACCTGCACTTCGCTGTCGCCCTCTTTGCACTCAGTGGTGCTGTTG
TGGTGGCTGGAAGCCTGCTGACCCCACCCCCACAGAGTGTCCAGATTGAGAACCTTACCTGGTGGACC
CTGGCTCAGGATGTGCCCTTGGGAACTAAAGCAGGTGATGGCCAAACACCCCAGAAACACGCCTTCTG

GGCCCGTGTCTGTGGCTTCAATGCCATCCTCCTCATGTGTGTCAACATATTCTTTTATGCCTACTTCG
CCTGACACTGCCATCCTGGACAGAAAGGCAGGAGCTCTGAGTCCTCAGGTCCACCCATTTCCCTCATG
GGGATCCCGA

NOV60c, CG56645-01 Protein Sequence	SEQ ID NO: 930	596 aa	MW at 64341.6kD

MAANSTSDLHTPGTQLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGYFLAGRDMTWWPIGASLFA
SSEGSGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEIVTLPEYIQKRYGGQRIRMY
LSVLSLLLSVFTKISLDLYAGALFVHICLGWNFYLSTILTLGITALYTIAGGLAAVIYTDALQTLIMV
VGAVILTIKAFDQIGGYGQLEAAYAQAIPSRTIANTTCHLPRTDAMHMFRDPHTGDLPWTGMTFGLTI
MATWYWCTDQVIVQRSLSARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFPDDVGCVVPSECLR
ACGAEVGCSNIAYPKLVMELMPIGLRGLMIAVMLAALMSSLTSIFNSSSTLFTMDIWRRLRPRSGERE
LLLVGRLVIVALIGVSVAWIPVLQDSNSGQLFIYMQSVTSSLAPPVTAVFVLGVFWRRANEQGAFWGL
IAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYLHFAVALFALSGAVVVAGSLLTPPPQSVQIE
NLTWWTLAQDVPLGTKAGDGQTPQKHAFWARVCGFNAILLMCVNIFFYAYFA

NOV60d, CG56645-02 DNA Sequence	SEQ ID NO: 931	1912 bp
	ORF Start: ATG at 35	ORF Stop: TGA at 1871

CGGGCTCATACCTAGTGCCTGCGGCAGGACAGCCATGGCCGCCAACTCCACCAGCGACCTCCACACTC
CCGGGACGCAGCTGAGCGTGGCTGACATCATCGTCATCACTGTGTATTTTGCTCTGAACGTGGCCGTG
GGCATATGGTCCTCTTGTCGGGCCAGTAGGAACACGGTGAATGGCTACTTCCTGGCAGGCCGGGACAT
GACGTGGTGGCCGATTGGAGCCTCCCTCTTCGCCAGCAGCGAGGGCTCTGGCCTCTTCATTGGACTGG
CGGGCTCAGGCGCGGCAGGAGGTCTGGCCGTGGCAGGCTTCGAGTGGAATGCCACGTACGTGCTGCTG
GCACTGGCATGGGTGTTCGTGCCCATCTACATCTCCTCAGAGATCGTCACCTTACCTGAGTACATTCA
GAAGCGCTACGGGGGCCAGCGGATCCGCATGTACCTGTCTGTCCTGTCCCTGCTACTGTCTGTCTTCA
CCAAGATATCGCTGGACCTGTACGCGGGGGCTCTGTTTGTGCACATCTGCCTGGGCTGGAACTTCTAC
CTCTCCACCATCCTCACGCTCGGCATCACAGCCCTGTACACCATCGCAGGGGGCCTGGCTGCTGTAAT
CTACACGGACGCCCTGCAGACGCTCATCATGGTGGTGGGGGCTGTCATCCTGACAATCAAAGCTTTTG
ACCAGATCGGTGGTTACGGGCAGCTGGAGGCAGCCTACGCCCAGGCCATTCCCTCCAGGACCATTGCC
AACACCACCTGCCACCTGCCACGTACAGACGCCATGCACATGTTTCGAGACCCCCACACAGGGGACCT
GCCGTGGACCGGGATGACCTTTGGCCTGACCATCATGGCCACCTGGTACTGGTGCACCGACCAGGTCA
TCGTGCAGCGATCACTGTCAGCCCGGGACCTGAACCATGCCAAGGCGGGCTCCATCCTGGCCAGCTAC
CTCAAGATGCTCCCCATGGGCCTGATCATCATGCCGGGCATGATCAGCCGCGCATTGTTCCCAGGTGC
TCATGTCTATGAGGAGAGACACCAAGTGTCCGTCTCTCGAACAGATGATGTGGGCTGCGTGGTGCCGT
CCGAGTGCCTGCGGGCCTGCGGGGCCGAGGTCGGCTGCTCCAACATCGCCTACCCCAAGCTGGTCATG
GAACTGATGCCCATCGGTCTGCGGGGGCTGATGATCGCAGTGATGCTGGCGGCGCTCATGTCGTCGCT
GACCTCCATCTTCAACAGCAGCAGCACCCTCTTCACTATGGACATCTGGAGGCGGCTGCGTCCCCGCT
CCGGCGAGCGGGAGCTCCTGCTGGTGGGACGGCTGGTCATAGTGGCACTCATCGGCGTGAGTGTGGCC
TGGATCCCCGTCCTGCAGGACTCCAACAGCGGGCAACTCTTCATCTACATGCAGTCAGTGACCAGCTC
CCTGGCCCCACCAGTGACTGCAGTCTTTGTCCTGGGCGTCTTCTGGCGACGTGCCAACGAGCAGGGGG
CCTTCTGGGGCCTGATAGCAGGGCTGGTGGTGGGGGCCACGAGGCTGGTCCTGGAATTCCTGAACCCA
GCCCCACCGTGCGGAGAGCCAGACACGCGGCCAGCCGTCCTGGGGAGCATCCACTACCTGCACTTCGC
TGTCGCCCTCTTTGCACTCAGTGGTGCTGTTGTGGTGGCTGGAAGCCTGCTGACCCCACCCCCACAGA
GTGTCCAGATTGAGAACCTTACCTGGTGGACCCTGGCTCAGGATGTGCCCTTGGGAACTAAAGCAGGT
GATGGCCAAACACTCCAGAAACACGCCTTCTGGGCCCGTGTCTGTGGCTTCAATGCCATCCTCCTCAT
GTGTGTCAACATATTCTTTTATGCCTACTTCGCCTGACACTGCCATCCTGGACAGAAAGGCAGGAGCT
CTGAGTCC

NOV60d, CG56645-02 Protein Sequence	SEQ ID NO: 932	612 aa	MW at 66194.7kD

MAANSTSDLHTPGTQLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGYFLAGRDMTWWPIGASLFA
SSEGSGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEIVTLPEYIQKRYGGQRIRMY
LSVLSLLLSVFTKISLDLYAGALFVHICLGWNFYLSTILTLGITALYTIAGGLAAVIYTDALQTLIMV
VGAVILTIKAFDQIGGYGQLEAAYAQAIPSRTIANTTCHLPRTDAMHMFRDPHTGDLPWTGMTFGLTI
MATWYWCTDQVIVQRSLSARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFPGAHVYEERHQVSV
SRTDDVGCVVPSECLRACGAEVGCSNIAYPKLVMELMPIGLRGLMIAVMLAALMSSLTSIFNSSSTLF
TMDIWRRLRPRSGERELLLVGRLVIVALIGVSVAWIPVLQDSNSGQLFIYMQSVTSSLAPPVTAVFVL
GVFWRRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYLHFAVALFALSGAVV
VAGSLLTPPPQSVQIENLTWWTLAQDVPLGTKAGDGQTLQKHAFWARVCGFNAILLMCVNIFFYAYFA

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 60B.

Table 60B. Comparison of the NOV60 protein sequences.

```
NOV60a  MAANSTSDLHTPGTQLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGYFLAGRDMTWW
NOV60b  MAANSTSDLHTPGTQLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGYFLAGRDMTWW
NOV60c  MAANSTSDLHTPGTQLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGYFLAGRDMTWW
NOV60d  MAANSTSDLHTPGTQLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGYFLAGRDMTWW

NOV60a  PIGASLFASSEGSGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEIVTL
NOV60b  PIGASLFASSEGSGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEIVTL
NOV60c  PIGASLFASSEGSGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEIVTL
NOV60d  PIGASLFASSEGSGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEIVTL

NOV60a  PEYIQKRYGGQRIRMYLSVLSLLLSVFTKISLDLYAGALFVHICLGWNFYLSTILTLGIT
NOV60b  PEYIQKRYGGQRIRMYLSVLSLLLSVFTKISLDLYAGALFVHICLGWNFYLSTILTLGIT
NOV60c  PEYIQKRYGGQRIRMYLSVLSLLLSVFTKISLDLYAGALFVHICLGWNFYLSTILTLGIT
NOV60d  PEYIQKRYGGQRIRMYLSVLSLLLSVFTKISLDLYAGALFVHICLGWNFYLSTILTLGIT

NOV60a  ALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGQLEAAYAQAIPSRTIANT
NOV60b  ALYTIA--------------------------AFDQIGGYGQLEAAYAQAIPSRTIANT
NOV60c  ALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGQLEAAYAQAIPSRTIANT
NOV60d  ALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGQLEAAYAQAIPSRTIANT

NOV60a  TCHLPRTDAMHMFRDPHTGDLPWTGMTFGLTIMATWYWCTDQVIVQRSLSARDLNHAKAG
NOV60b  TCQLPRTDAMHMFRDPHTGDLPWTGMTFGLTIMATWYWCTDQVIVEGSLSARDLNHAKGG
NOV60c  TCHLPRTDAMHMFRDPHTGDLPWTGMTFGLTIMATWYWCTDQVIVQRSLSARDLNHAKAG
NOV60d  TCHLPRTDAMHMFRDPHTGDLPWTGMTFGLTIMATWYWCTDQVIVQRSLSARDLNHAKAG

NOV60a  SILASYLKMLPMGLIIMPGMISRALFPGAHVYEERHQVSVSRTDDVGCVVPSECLRACGA
NOV60b  SILGSYLKMLPMGLIIMPGMISRALFPGAHVYEERHQVSVSRTDDVGCVVPSECLRACGA
NOV60c  SILASYLKMLPMGLIIMPGMISRALFP---------------DDVGCVVPSECLRACGA
NOV60d  SILASYLKMLPMGLIIMPGMISRALFPGAHVYEERHQVSVSRTDDVGCVVPSECLRACGA

NOV60a  EVGCSNIAYPKLVMELMPIGLRGLMIAVMLAALMSSLTSIFNSSSTLFTMDIWRRLRPRS
NOV60b  EVGCSNIAYPKLVMELMPIGLRGLMIAVMLAALMSSLTSIFNSSSTLFTMDIWRRLRPRS
NOV60c  EVGCSNIAYPKLVMELMPIGLRGLMIAVMLAALMSSLTSIFNSSSTLFTMDIWRRLRPRS
NOV60d  EVGCSNIAYPKLVMELMPIGLRGLMIAVMLAALMSSLTSIFNSSSTLFTMDIWRRLRPRS

NOV60a  GERELLLVGRLVIVALIGVSVAWIPVLQDSNSGQLFIYMQSVTSSLAPPVTAVFVL----
NOV60b  GERELLLVGRLVIVALIGVSVAWIPVLQDSNSGQLFIYMQSVTSSLAPPVTAVFVLGVFW
NOV60c  GERELLLVGRLVIVALIGVSVAWIPVLQDSNSGQLFIYMQSVTSSLAPPVTAVFVLGVFW
NOV60d  GERELLLVGRLVIVALIGVSVAWIPVLQDSNSGQLFIYMQSVTSSLAPPVTAVFVLGVFW

NOV60a  ----------GLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYLHFAVALFALS
NOV60b  RRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYLHFAVALFALS
NOV60c  RRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYLHFAVALFALS
NOV60d  RRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYLHFAVALFALS

NOV60a  GAVVVAGSLLTPPPQSVQIENLTWWNLAQDVPLGTKAGDGQTPQKHAFWARVCGFNAILL
NOV60b  GAVVVAGSLLTPPPQSVQIENLTWWTLAQDVPLGTKAGDGQTPQKHAFWARVCGFNAILL
NOV60c  GAVVVAGSLLTPPPQSVQIENLTWWTLAQDVPLGTKAGDGQTPQKHAFWARVCGFNAILL
```

```
NOV60d  GAVVVAGSLLTPPPQSVQIENLTWWTLAQDVPLGTKAGDGQTLQKHAFWARVCGFNAILL

NOV60a  MCVNIFFYAYFA
NOV60b  MCVNIFFYAYFA
NOV60c  MCVNIFFYAYFA
NOV60d  MCVNIFFYAYFA

NOV60a  (SEQ ID NO:  926)
NOV60b  (SEQ ID NO:  928)
NOV60c  (SEQ ID NO:  930)
NOV60d  (SEQ ID NO:  932)
```

Further analysis of the NOV60a protein yielded the following properties shown in Table 60C.

Table 60C. Protein Sequence Properties NOV60a

SignalP analysis:	Cleavage site between residues 43 and 44

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 8;  pos.chg 0;  neg.chg 1
      H-region:  length 11;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -2.19
      possible cleavage site: between 42 and 43

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:  12
      INTEGRAL    Likelihood = -6.00   Transmembrane   21 -  37
      INTEGRAL    Likelihood = -1.17   Transmembrane   75 -  91
      INTEGRAL    Likelihood = -4.51   Transmembrane   97 - 113
      INTEGRAL    Likelihood = -1.33   Transmembrane  136 - 152
      INTEGRAL    Likelihood = -2.92   Transmembrane  177 - 193
      INTEGRAL    Likelihood = -3.88   Transmembrane  196 - 212
      INTEGRAL    Likelihood = -0.27   Transmembrane  301 - 317
      INTEGRAL    Likelihood = -5.04   Transmembrane  384 - 400
      INTEGRAL    Likelihood =-10.03   Transmembrane  431 - 447
      INTEGRAL    Likelihood = -8.65   Transmembrane  470 - 486
      INTEGRAL    Likelihood = -6.74   Transmembrane  520 - 536
      INTEGRAL    Likelihood = -3.45   Transmembrane  578 - 594
      PERIPHERAL  Likelihood =  0.63 (at 159)
      ALOM score: -10.03  (number of TMSs: 12)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 28
      Charge difference:  4.5   C( 3.0) - N(-1.5)
      C > N: C-terminal side will be inside
```

```
>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  8.68
      Hyd Moment(95):  8.07    G content:       1
      D/E content:     2       S/T content:     6
      Score: -5.13

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   5.5%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1
```

```
COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       77.8 %: endoplasmic reticulum
       11.1 %: mitochondrial
       11.1 %: vacuolar

>> prediction for CG56645-03 is end (k=9)
```

A search of the NOV60a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 60D.

Table 60D. Geneseq Results for NOV60a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV60a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB99496	Amino acid sequence of sodium/glucose cotransporter Fbh68723pat - Homo sapiens, 664 aa. [WO200283857-A2, 24-OCT-2002]	1..598 48..659	597/612 (97%) 597/612 (97%)	0.0
ABB99497	Amino acid sequence of sodium/glucose cotransporter h68723 - Homo sapiens, 643 aa. [WO200283857-A2, 24-OCT-2002]	1..598 48..643	581/612 (94%) 581/612 (94%)	0.0
AAO14199	Human transporter and ion channel TRICH-16 - Homo sapiens, 596 aa. [WO200204520-A2, 17-JAN-2002]	1..598 1..596	581/612 (94%) 581/612 (94%)	0.0
ABB80588	Human sbg1020829SGLT protein - Homo sapiens, 596 aa. [WO200222802-A1, 21-MAR-2002]	1..598 1..596	581/612 (94%) 581/612 (94%)	0.0
ABG31594	Human transporter protein - Homo sapiens, 596 aa. [US2002081678-A1, 27-JUN-2002]	1..598 1..596	581/612 (94%) 581/612 (94%)	0.0

In a BLAST search of public sequence databases, the NOV60a protein was found to have homology to the proteins shown in the BLASTP data in Table 60E.

Table 60E. Public BLASTP Results for NOV60a

Protein Accession Number	Protein/Organism/Length	NOV60a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC51188	Sequence 144 from Patent WO0149728 - Homo sapiens (Human), 596 aa.	1..598 1..596	581/612 (94%) 581/612 (94%)	0.0
Q28610	Na+/glucose cotransporter-related protein - Oryctolagus cuniculus (Rabbit), 597 aa.	1..598 2..597	518/612 (84%) 544/612 (88%)	0.0
Q96LQ1	Hypothetical protein FLJ25217 - Homo sapiens (Human), 517 aa.	157..598 62..517	441/456 (96%) 441/456 (96%)	0.0
Q8BZW1	Weakly similar to NA+-glucose cotransporter type 1 - Mus musculus (Mouse), 685 aa.	6..551 19..562	339/560 (60%) 425/560 (75%)	0.0
Q8BGU9	Weakly similar to NA+-glucose cotransporter type 1 - Mus musculus (Mouse), 685 aa.	6..551 19..562	338/560 (60%) 424/560 (75%)	0.0

PFam analysis indicates that the NOV60a protein contains the domains shown in the Table 60F.

Table 60F. Domain Analysis of NOV60a

Pfam Domain	NOV60a Match Region	Identities/ Similarities for the Matched Region	Expect Value
SSF	50..485	187/463 (40%) 362/463 (78%)	1.3e-167

Example 61.

The NOV61 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 61A.

Table 61A. NOV61 Sequence Analysis

NOV61a, CG56667-01 DNA Sequence	SEQ ID NO: 933	1027 bp
	ORF Start: ATG at 1	ORF Stop: TAG at 940
ATGATCTGCTCAGCTATCAACCTACACTTACTACTGGCAGTTAAGATGATTCACCCTGTCTGGATTCT		

TGCTCCTCGGGAGCAAGGGCTGTTTCTGCTGATTTATCTGGCAGTGCTGGTGGGGAACCTGCTCATCA
TTGCAGTCATCACTCTCGATCAGCATCTTCACACACCCATGTACTTCTTCCTGAAGAACCTCTCCGTT
TTGGATCTGTGCTACATCTCAGTCACTGTGCCTAAATCCATCCGTAACTCCCTGACTCGCAGAAGCTC
CATCTCTTATCTTGGCTGTGTGGCTCAAGTCTATTTTTTCTCTGCCTTTGCATCTGCTGAGCTGGCCT
TCCTTACTGTCATGTCTTATGACCGCTATGTTGCCATTTGCCACCCCCTCCAATACAGAGCCGTGATG
ACATCAGGAGGGTGCTATCAGATGGCAGTCACCACCTGGCTAAGCTGCTTTTCCTACGCAGCCGTCCA
CACTGGCAACATGTTTCGGGAGCACGTTTGCAGATCCAGTGTGATCCACCAGTTCTTCCGTGACATCC
CTCATGTGTTGGCCCTGGTTTCCTGTGAGGTTTTCTTTGTAGAGTTTTTGACCCTGGCCCTGAGCTCA
TGCTTGGTTCTGGGATGCTTTATTCTCATGATGATCTCCTATTTCCAAATCTTCTCAACGGTGCTCAG
AATCCCTTCAGGACAGAGTCGAGCAAAAGCCTTCTCCACCTGCTCCCCCCAGCTCATTGTCATCATGC
TCTTTCTTACCACAGGGCTCTTTGCTGCCTTAGGACCAATTGCAAAAGCTCTGTCCATTCAGGATTTA
GTGATTGCTCTGACATACACAGTTTTGCCTCCCTTCCTCAATCCCATCATATATAGTCTTAGGAATAA
GGAGATTAAAACAGCCATGTGGAGACTCTTTGTGAAGATATATTTTCTGCAAAAGTAGAACATCCTGG
TCTTTACTATAGAAGATCTGCAACAAAACCCCAAAAAAGCATAAATACTTTATGACAAAAAAAGATGA
AAAAATT

NOV61a, CG56667-01 Protein Sequence	SEQ ID NO: 934	313 aa	MW at 35319.9kD

MICSAINLHLLLAVKMIHPVWILAPREQGLFLLIYLAVLVGNLLIIAVITLDQHLHTPMYFFLKNLSV
LDLCYISVTVPKSIRNSLTRRSSISYLGCVAQVYFFSAFASAELAFLTVMSYDRYVAICHPLQYRAVM
TSGGCYQMAVTTWLSCFSYAAVHTGNMFREHVCRSSVIHQFFRDIPHVLALVSCEVFFVEFLTLALSS
CLVLGCFILMMISYFQIFSTVLRIPSGQSRAKAFSTCSPQLIVIMLFLTTGLFAALGPIAKALSIQDL
VIALTYTVLPPFLNPIIYSLRNKEIKTAMWRLFVKIYFLQK

NOV61b, 268952113 DNA Sequence	SEQ ID NO: 935	817 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCGTCATCACTCTCGATCAGCATCTTCACACACCCATGTACTTCTTCCTGAAGAACCTCT
CTGTTTTGGATCTGTGCTACATCTCAGTCACTGTGCCTAAATCCATCCGTAACTCCCTGACTCGCAGA
AGCTCCATCTCTTATCTTGGCTGTGTGGCTCAAGTCTATTTTTTCTCTGCCTTTGCATCTGCTGAGCT
GGCCTTCCTTACTGTCATGTCTTATGACCGCTATGTTGCCATTTGCCACCCCCTCCAATACAGAGCCG
TGATGACATCAGGAGGGTGCTATCAGATGGCAGTCACCACCTGGCTAAGCTGCTTTTCCTACGCAGCC
GTCCACACTGGCAACATGTTTCGGGAGCACGTTTGCAGATCCAATGTGATCCACCAGTTCTTCCGTGA
CATCCCTCATGTGTTGGCCCTGGTTTCCTGTGAGGTTTTCTTTGTAGAGTTTTTGACCCTGGCCCTGA
GCTCATGCTTGGTTCTGGGATGCTTTATTCTCATGATGATCTCCTATTTCCAAATCTTCTCAACGGTG
CTCAGAATCCCTTCAGGACAGAGTCGAGCAAAAGCCTTCTCCACCTGCTCCCCCCAGCTCATTGTCAT
CATGCTCTTTCTTACCACAGGGCTCTTTGCTGCCTTAGGACCAATTGCAAAAGCTCTGTCCATTCAGG
ATTTAGTGATTGCTCTGACATACACAGTTTTGCCTCCCTTCCTCAATCCCATCATATATAGTCTTAGG
AATAAGGAGATTAAAACAGCCATGTGGAGACTCTTTGTGAAGATATATTTTCTGCAAAAGCTCGAGGG
C

NOV61b, 268952113 Protein Sequence	SEQ ID NO: 936	272 aa	MW at 30688.0kD

TGSVITLDQHLHTPMYFFLKNLSVLDLCYISVTVPKSIRNSLTRRSSISYLGCVAQVYFFSAFASAEL
AFLTVMSYDRYVAICHPLQYRAVMTSGGCYQMAVTTWLSCFSYAAVHTGNMFREHVCRSNVIHQFFRD
IPHVLALVSCEVFFVEFLTLALSSCLVLGCFILMMISYFQIFSTVLRIPSGQSRAKAFSTCSPQLIVI
MLFLTTGLFAALGPIAKALSIQDLVIALTYTVLPPFLNPIIYSLRNKEIKTAMWRLFVKIYFLQKLEG

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 61B.

Table 61B. Comparison of the NOV61 protein sequences.

```
NOV61a  MICSAINLHLLLAVKMIHPVWILAPREQGLFLLIYLAVLVGNLLIIAVITLDQHLHTPMY
NOV61b  ----------------------------------------------TGSVITLDQHLHTPMY

NOV61a  FFLKNLSVLDLCYISVTVPKSIRNSLTRRSSISYLGCVAQVYFFSAFASAELAFLTVMSY
NOV61b  FFLKNLSVLDLCYISVTVPKSIRNSLTRRSSISYLGCVAQVYFFSAFASAELAFLTVMSY
```

```
NOV61a  DRYVAICHPLQYRAVMTSGGCYQMAVTTWLSCFSYAAVHTGNMFREHVCRSSVIHQFFRD
NOV61b  DRYVAICHPLQYRAVMTSGGCYQMAVTTWLSCFSYAAVHTGNMFREHVCRSNVIHQFFRD

NOV61a  IPHVLALVSCEVFFVEFLTLALSSCLVLGCFILMMISYFQIFSTVLRIPSGQSRAKAFST
NOV61b  IPHVLALVSCEVFFVEFLTLALSSCLVLGCFILMMISYFQIFSTVLRIPSGQSRAKAFST

NOV61a  CSPQLIVIMLFLTTGLFAALGPIAKALSIQDLVIALTYTVLPPFLNPIIYSLRNKEIKTA
NOV61b  CSPQLIVIMLFLTTGLFAALGPIAKALSIQDLVIALTYTVLPPFLNPIIYSLRNKEIKTA

NOV61a  MWRLFVKIYFLQK---
NOV61b  MWRLFVKIYFLQKLEG

NOV61a  (SEQ ID NO:  934)
NOV61b  (SEQ ID NO:  936)
```

Further analysis of the NOV61a protein yielded the following properties shown in Table 61C.

Table 61C. Protein Sequence Properties NOV61a

SignalP analysis:	Cleavage site between residues 48 and 49

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 14;  peak value   6.18
      PSG score:   1.78

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.68
      possible cleavage site: between 41 and 42

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   7
      INTEGRAL    Likelihood = -2.39   Transmembrane    1 -  17
      INTEGRAL    Likelihood =-11.20   Transmembrane   33 -  49
      INTEGRAL    Likelihood = -0.22   Transmembrane  101 - 117
      INTEGRAL    Likelihood = -1.22   Transmembrane  183 - 199
      INTEGRAL    Likelihood = -8.86   Transmembrane  200 - 216
      INTEGRAL    Likelihood = -7.54   Transmembrane  245 - 261
      INTEGRAL    Likelihood = -0.00   Transmembrane  273 - 289
      PERIPHERAL  Likelihood =  2.01 (at 62)
      ALOM score: -11.20  (number of TMSs: 7)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 8
      Charge difference:  0.5   C( 2.0) - N( 1.5)
      C > N: C-terminal side will be inside
```

```
>>>Caution: Inconsistent mtop result with signal peptide
>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  1.74
      Hyd Moment(95):  4.81    G content:       0
      D/E content:     1       S/T content:     1
      Score: -4.54

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 99  RRS|SI

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.0%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1
```

```
COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       11.1 %: mitochondrial
       11.1 %: vacuolar
       11.1 %: vesicles of secretory system
       11.1 %: Golgi

>> prediction for CG56667-01 is end (k=9)
```

A search of the NOV61a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 61D.

5

Table 61D. Geneseq Results for NOV61a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV61a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAG72316	Human olfactory receptor polypeptide, SEQ ID NO: 1997 - Homo sapiens, 177 aa. [WO200127158-A2, 19-APR-2001]	47..219 1..173	149/173 (86%) 151/173 (87%)	2e-80
ABB77320	Human G-protein coupled receptor SEQ ID NO 4 - Homo sapiens, 309 aa. [WO200198323-A2, 27-DEC-2001]	30..304 28..301	148/275 (53%) 189/275 (67%)	8e-78
AAU95571	Human olfactory and pheromone G protein-coupled receptor #58 - Homo sapiens, 309 aa. [WO200224726-A2, 28-MAR-2002]	30..304 28..301	148/275 (53%) 189/275 (67%)	8e-78
ABP51618	Human G protein coupled receptor SEQ ID NO:118 - Homo sapiens, 308 aa. [WO200250276-A2, 27-JUN-2002]	30..304 28..300	149/275 (54%) 191/275 (69%)	8e-78
ABG76784	Human G-protein coupled receptor (GPCR) protein #18 - Homo sapiens, 309 aa. [WO200259313-A2, 01-AUG-2002]	30..304 28..301	148/275 (53%) 189/275 (67%)	8e-78

In a BLAST search of public sequence databases, the NOV61a protein was found to have homology to the proteins shown in the BLASTP data in Table 61E.

Table 61E. Public BLASTP Results for NOV61a				
Protein Accession Number	**Protein/Organism/Length**	**NOV61a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8NHC8	Seven transmembrane helix receptor - Homo sapiens (Human), 309 aa.	20..307 18..308	164/291 (56%) 200/291 (68%)	3e-81
Q8VF67	Olfactory receptor MOR218-3 - Mus musculus (Mouse), 315 aa.	30..305 32..307	146/276 (52%) 182/276 (65%)	1e-78
CAD57961	Sequence 117 from Patent WO0250276 - Homo sapiens (Human), 308 aa.	30..304 28..300	149/275 (54%) 191/275 (69%)	2e-77
Q8NHC5	Seven transmembrane helix receptor - Homo sapiens (Human), 309 aa.	30..304 28..301	148/275 (53%) 189/275 (67%)	2e-77
Q8VEU5	Olfactory receptor MOR218-8 - Mus musculus (Mouse), 320 aa.	30..313 27..310	143/284 (50%) 185/284 (64%)	4e-75

PFam analysis indicates that the NOV61a protein contains the domains shown in the Table 61F.

Table 61F. Domain Analysis of NOV61a			
Pfam Domain	**NOV61a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
7tm_1	41..290	53/278 (19%) 174/278 (63%)	3.9e-31

Example 62.

The NOV62 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 62A.

Table 62A. NOV62 Sequence Analysis		
NOV62a, CG56868-01 DNA Sequence	SEQ ID NO: 937	2422 bp
	ORF Start: ATG at 51	ORF Stop: TAG at 2382
GAACTCCTTTTCTCAAGCACTTCTGCTCTCCTCTACCAGAATCACTCAGAATGCTTCCCGGGTGTATA		

TTCTTGATGATTTTACTCATTCCTCAGGTTAAAGAAAAGTTCATCCTTGGAGTAGAGGGTCAACAACT
GGTTCGTCCTAAAAAGCTTCCTCTGATACAGAAGCGAGATACTGGACACACCCATGATGATGACATAC
TGAAAACGTATGAAGAAGAATTGTTGTATGAAATAAAACTAAATAGAAAAACCTTAGTCCTTCATCTT
CTAAGATCCAGGGAGTTCCTAGGCTCAAATTACAGTGAAACATTCTACTCCATGAAAGGAGAAGCGTT
CACCAGGCATCCTCAGATCATGGAACACTGTTACTATAAAGGAAACATCCTAAATGAAAAGAATTCTG
TTGCCAGCATCAGTACTTGTGACGGGTTGAGGGGATTCTTCAGAATAAACGACCAAAGATACCTCATT
GAACCAGTGAAATACTCAGATGAGGGAGAACATTTGGTGTTCAAATATAACCTGAGGGTGCCGTATGG
TGCCAATTATTCCTGTACAGAGCTTAATTTTACCAGAAAAACTGTTCCAGGGGATAATGAATCTGAAG
AAGACTCCAAAATAAAAGGCATCCATGATGAAAAGTATGTTGAATTGTTCATTGTTGCTGATGATACT
GTGTATCGCAGAAATGGTCATCCTCACAATAAACTAAGGAACCGAATTTGGGGAATGGTCAATTTTGT
CAACATGATTTATAAAACCTTAAACATCCATGTGACGTTGGTTGGCATTGAAATATGGACACATGAAG
ATAAAATAGAACTATATTCAAATATAGAAACTACCTTATTGCGTTTTTCATTTTGGCAAGAAAAGATC
CTTAAAACACGGAAGGATTTTGATCATGTTGTATTACTCAGTGGGAAGTGGCTCTACTCACATGTGCA
AGGAATTTCTTATCCAGGGGGTATGTGCCTGCCCTATTATTCCACCAGTATCATTAAGGATCTTTTAC
CTGACACAAACATAATTGCAAACAGAATGGCACATCAACTGGGGCATAACCTTGGGATGCAGCATGAC
GAGTTCCCATGCACCTGTCCTTCAGGAAAATGCGTGATGGACAGTGATGGAAGCATTCCTGCACTGAA
ATTCAGTAAATGCAGCCAAAACCAATACCACCAGTACTTGAAGGATTATAAGCCAACATGCATGCTCA
ACATTCCATTTCCTTACAATTTTCATGATTTCCAATTTTGTGGAAACAAGAAGTTGGATGAGGGTGAA
GAGTGTGACTGTGGCCCTGCTCAGGAGTGTACTAATCCTTGCTGTGATGCACACACATGTGTACTGAA
GCCAGGATTTACTTGTGCAGAAGGAGAATGCTGTGAATCTTGTCAGATAAAAAAAGCAGGGTCCATAT
GCAGACCGGCGAAAGATGAATGTGATTTTCCTGAGATGTGCACTGGCCACTCGCCTGCCTGTCCTAAG
GACCAGTTCAGGGTCAATGGATTTCCTTGCAAGAACTCAGAAGGCTACTGTTTCATGGGGAAATGTCC
AACTCGTGAGGATCAGTGCTCTGAACTATTTGATGATGAGGCAATAGAGAGTCATGATATCTGCTACA
AGATGAATACAAAAGGAAATAAATTTGGATACTGCAAAAACAAGGAAAACAGATTTCTTCCCTGTGAG
GAGAAGGATGTCAGATGTGGAAAGATCTACTGCACTGGAGGGGAGCTTTCCTCTCTCCTTGGAGAAGA
CAAGACTTATCACCTTAAGGATCCCCAGAAGAATGCTACTGTCAAATGCAAAACTATTTTTTTATACC
ATGATTCTACAGACATTGGCCTGGTGGCGTCAGGAACAAAATGTGGAGAGGGAATGGTGTGCAACAAT
GGTGAATGTCTAAACATGGAAAAGGTCTATATCTCAACCAATTGCCCCTCTCAGTGCAATGAAAATCC
TGTGGATGGCCACGGACTCCAGTGCCACTGTGAGGAAGGACAGGCACCTGTAGCCTGTGAAGAAACCT
TACATGTTACCAGTATCACCATCTTGGTTGTTGTGCTTGTCCTGGTTATTGTCGGTATCGGAGTTCTT
ATACTATTAGTTCGTTACCGAAAATGTATCAAGTTGAAGCAAGTTCAGAGCCCACCTACAGAAACCCT
GGGAGTGGAGAACAAAGGATACTTTGGTGATGAGCAGCAGATAAGGACTGAGCCAATCCTGCCAGAAA
TTCATTTCCTAAATCAGAGAACTCCAGAATCCTTGGAAAGCCTGCCCACTAGTTTTTCAAGTCCCCAC
TACATCACACTGAAACCTGCAAGTAAAGATTCAAGAGGAATCGCAGATCCCAATCAAAGTGCCAAGTG
GTAGGTTACCCTGACAGATAGTACCTCCCTTTTTTATTTTTC

NOV62a, CG56868-01 Protein Sequence	SEQ ID NO: 938	777 aa	MW at 88341.2kD

MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDILKTYEEELLYEIKLNRK
TLVLHLLRSREFLGSNYSETFYSMKGEAFTRHPQIMEHCYYKGNILNEKNSVASISTCDGLRGFFRIN
DQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKTVPGDNESEEDSKIKGIHDEKYVELF
IVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYKTLNIHVTLVGIEIWTHEDKIELYSNIETTLLRFS
FWQEKILKTRKDFDHVVLLSGKWLYSHVQGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHN
LGMQHDEFPCTCPSGKCVMDSDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNK
KLDEGEECDCGPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGH
SPACPKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGYCKNKEN
RFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDSTDIGLVASGTKCGE
GMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPVACEETLHVTSITILVVVLVLVI
VGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFGDEQQIRTEPILPEIHFLNQRTPESLESLPT
SFSSPHYITLKPASKDSRGIADPNQSAKW

NOV62b, 276580332 DNA Sequence	SEQ ID NO: 939	247 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGTACCGATGAGGGTGAAGAGTGTGACTGTGGCCCTGCTCAGGAGTGTACTAATCCTTGCTGTG
ATGCACACACATGTGTACTGAAGCCAGGATTTACTTGTGCAGAAGGAGAATGCTGTGAATCTTGTCAG
ATAAAAAAAGCAGGGTCCATATGCAGACCGGCGAAAGATGAATGTGATTTTCCTGAGATGTGCACTGG
CCACTCGCCTGCCTGTCCTAAGGACCAGTTCAGGGTCGACGGC

NOV62b, 276580332 Protein Sequence	SEQ ID NO: 940	82 aa	MW at 8696.6kD
TGTDEGEECDCGPAQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTG HSPACPKDQFRVDG			

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 62B.

Table 62B. Comparison of the NOV62 protein sequences.

```
NOV62a  MLPGCIFLMILLIPQVKEKFILGVEGQQLVRPKKLPLIQKRDTGHTHDDDILKTYEEELL
NOV62b  ------------------------------------------------------------

NOV62a  YEIKLNRKTLVLHLLRSREFLGSNYSETFYSMKGEAFTRHPQIMEHCYYKGNILNEKNSV
NOV62b  ------------------------------------------------------------

NOV62a  ASISTCDGLRGFFRINDQRYLIEPVKYSDEGEHLVFKYNLRVPYGANYSCTELNFTRKTV
NOV62b  ------------------------------------------------------------

NOV62a  PGDNESEEDSKIKGIHDEKYVELFIVADDTVYRRNGHPHNKLRNRIWGMVNFVNMIYKTL
NOV62b  ------------------------------------------------------------

NOV62a  NIHVTLVGIEIWTHEDKIELYSNIETTLLRFSFWQEKILKTRKDFDHVVLLSGKWLYSHV
NOV62b  ------------------------------------------------------------

NOV62a  QGISYPGGMCLPYYSTSIIKDLLPDTNIIANRMAHQLGHNLGMQHDEFPCTCPSGKCVMD
NOV62b  ------------------------------------------------------------

NOV62a  SDGSIPALKFSKCSQNQYHQYLKDYKPTCMLNIPFPYNFHDFQFCGNKKLDEGEECDCGP
NOV62b  ----------------------------------------------TGTDEGEECDCGP

NOV62a  AQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGHSPAC
NOV62b  AQECTNPCCDAHTCVLKPGFTCAEGECCESCQIKKAGSICRPAKDECDFPEMCTGHSPAC

NOV62a  PKDQFRVNGFPCKNSEGYCFMGKCPTREDQCSELFDDEAIESHDICYKMNTKGNKFGYCK
NOV62b  PKDQFRVDG------------------------------------------------------

NOV62a  NKENRFLPCEEKDVRCGKIYCTGGELSSLLGEDKTYHLKDPQKNATVKCKTIFLYHDSTD
NOV62b  ------------------------------------------------------------

NOV62a  IGLVASGTKCGEGMVCNNGECLNMEKVYISTNCPSQCNENPVDGHGLQCHCEEGQAPVAC
NOV62b  ------------------------------------------------------------

NOV62a  EETLHVTSITILVVVLVLVIVGIGVLILLVRYRKCIKLKQVQSPPTETLGVENKGYFGDE
NOV62b  ------------------------------------------------------------

NOV62a  QQIRTEPILPEIHFLNQRTPESLESLPTSFSSPHYITLKPASKDSRGIADPNQSAKW
NOV62b  ------------------------------------------------------------

NOV62a  (SEQ ID NO:  938)
NOV62b  (SEQ ID NO:  940)
```

Further analysis of the NOV62a protein yielded the following properties shown in Table 62C.

Table 62C. Protein Sequence Properties NOV62a

SignalP analysis:	Cleavage site between residues 19 and 20

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 16;  peak value  12.97
      PSG score:   8.57

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -5.58
      possible cleavage site: between 18 and 19

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   2
      INTEGRAL    Likelihood = -2.18   Transmembrane    1 -  17
      INTEGRAL    Likelihood =-18.52   Transmembrane  671 - 687
      PERIPHERAL  Likelihood =  4.35 (at 235)
      ALOM score: -18.52  (number of TMSs: 2)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 8
      Charge difference:  0.0   C( 1.0) - N( 1.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  3.80
      Hyd Moment(95):  1.48    G content:       1
      D/E content:     1       S/T content:     0
      Score: -6.33

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: RPKK (4) at   31
      pat7: none
      bipartite: none
      content of basic residues:  11.3%
      NLS Score: -0.22

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
```

```
      KKXX-like motif in the C-terminus: QSAK

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:
      Bacterial regulatory proteins, gntR family signature (PS00043):
*** found ***
         EEELLYEIKLNRKTLVLHLLRS at 56

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 55.5

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       34.8 %: nuclear
       30.4 %: endoplasmic reticulum
       21.7 %: mitochondrial
        4.3 %: vesicles of secretory system
        4.3 %: cytoplasmic
        4.3 %: peroxisomal

>> prediction for CG56868-01 is nuc (k=23)
```

A search of the NOV62a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 62D.

Table 62D. Geneseq Results for NOV62a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV62a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB47567	Protease PRTS-9 - Homo sapiens, 776 aa. [WO200171004-A2, 27-SEP-2001]	1..776 1..776	771/776 (99%) 773/776 (99%)	0.0
AAU77409	Human NOV4b protein, homologue of ADAM proteins - Homo sapiens, 779 aa. [WO200206329-A2, 24-JAN-2002]	1..777 1..779	776/780 (99%) 776/780 (99%)	0.0
AAU77408	Human NOV4a protein, homologue of ADAM proteins - Homo sapiens, 778 aa. [WO200206329-A2, 24-JAN-2002]	1..777 1..778	766/779 (98%) 774/779 (99%)	0.0
AAU16950	Human novel secreted protein, SEQ ID 191 - Homo sapiens, 695 aa. [WO200155441-A2, 02-AUG-2001]	1..677 17..693	664/677 (98%) 675/677 (99%)	0.0
AAB64744	Gene 16 human secreted protein homologous amino acid sequence #138 - Macaca fascicularis, 375 aa. [WO200077237-A1, 21-DEC-2000]	336..710 1..375	353/375 (94%) 364/375 (96%)	0.0

In a BLAST search of public sequence databases, the NOV62a protein was found to have homology to the proteins shown in the BLASTP data in Table 62E.

Table 62E. Public BLASTP Results for NOV62a				
Protein Accession Number	**Protein/Organism/Length**	**NOV62a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9H2U9	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Sperm maturation-related glycoprotein GP-83) - Homo sapiens (Human), 754 aa.	1..776 1..754	741/776 (95%) 751/776 (96%)	0.0
Q28475	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Epididymal apical protein I) (EAP I) - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 776 aa.	1..776 1..776	717/776 (92%) 748/776 (95%)	0.0
O35227	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) - Mus musculus (Mouse), 788 aa.	1..774 1..771	523/774 (67%) 638/774 (81%)	0.0
AAH43207	Hypothetical protein - Homo sapiens (Human), 561 aa (fragment).	212..736 29..553	524/525 (99%) 525/525 (99%)	0.0
Q63180	ADAM 7 precursor (A disintegrin and metalloproteinase domain 7) (Epididymal apical protein I) (EAP I) - Rattus norvegicus (Rat), 789 aa.	1..774 1..772	520/774 (67%) 632/774 (81%)	0.0

PFam analysis indicates that the NOV62a protein contains the domains shown in the Table 62F.

Table 62F. Domain Analysis of NOV62a			
Pfam Domain	**NOV62a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Pep_M12B_propep	73..188	38/119 (32%) 100/119 (84%)	6.7e-45
Reprolysin	199..394	88/203 (43%) 173/203 (85%)	5.5e-104
disintegrin	411..486	41/76 (54%) 53/76 (70%)	1.6e-25

Example 63.

The NOV63 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 63A.

5

Table 63A. NOV63 Sequence Analysis

NOV63a, CG56870-01 DNA Sequence	SEQ ID NO: 941	1268 bp
	ORF Start: ATG at 71	ORF Stop: TAA at 1196

ACTTCTTTCTTTTCTGTTTCAGAGTTACTGATTTATTCTTGAGATTCCTCTACTCTCGTTATCTGACC
TC**ATG**GATGAACTTCAGGATGTTCAGCTCACAGAGATCAAACCACTTCTAAATGATAAGAATGGTACA
AGAAACTTCCAGGACTTTGACTGTCAGGAACATGATATAGAAACAACTCATGGTGTGGTCCACGTCAC
TATAAGAGGCTTACCCAAAGGAAACAGACCAGTTATACTAACATATCATGACATTGGCCTCAACCGTA
AATCCTGTTTCAATGCATTCTTTAACTTTGAGGATATGCAAGAGATCACCCAGCACTTTGCTGTCTGT
CATGTGGATGCCCCAGGCCAGCAGGAAGGTGCACCCTCTTTCCCAACAGGGTATCAGTACCCCACAAT
GGATGAGCTGGCTGAAATGCTGCCTCCTGTTCTTACCCACCTAAGCCTGAAAAGCATCATTGGAATTG
GAGTTGGAGCTGGAGCTTACATCCTCAGCAGATTTGCACTCAACCATCCAGAGCTTGTGGAAGGCCTT
GTGCTCATTAATGTTGACCCTTGCGCTAAAGGCTGGATTGACTGGGCAGCTTCCAAACTCTCTGGCCT
GACAACCAATGTTGTGGACATTATTTTGGCTCATCACTTTGGGCAGGAAGAGTTACAGGCCAACCTGG
ACCTGATCCAAACCTACAGAATGCATATTGCCCAAGACATCAACCAAGACAACCTGCAGCTCTTCTTG
AATTCCTACAATGGGCGCAGAGACCTGGAGATCGAAAGACCCATACTGGGCCAAAATGATAACAAATC
AAAAACATTAAAGTGTTCTACTTTACTGGTGGTAGGGGACAATTCGCCTGCAGTTGAGGCTGTGGTCG
AATGCAATTCCCGCCTGAACCCTATAAATACAACTTTGCTAAAGATGGCGGACTGTGGGGGACTGCCC
CAGGTAGTTCAGCCTGGGAAGCTCACCGAGGCCTTCAAGTACTTTTTGCAGGGAATGGGCTACGTCCC
GTCTGCCAGCATGACTCGGCTCGCCCGATCACGAACCCACTCAACCTCGAGTAGCCTCGGCTCTGGAG
AAAGTCCCTTCAGCCGGTCTGTCACCAGCAATCAGTCAGATGGAACTCAAGAATCCTGTGAGTCCCCT
GATGTCCTGGACAGACACCAGACCATGGAGGTGTCCTGC**TAA**GCAGATGCTCCTCCCCTGGACCATTG
CAAGTCCATCCTTCAAATGACCACTCCATAATATAACATTTCAT

NOV63a, CG56870-01 Protein Sequence	SEQ ID NO: 942	375 aa	MW at 41413.3kD

MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQEHDIETTHGVVHVTIRGLPKGNRPVILTYHDIGLNRK
SCFNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDELAEMLPPVLTHLSLKSIIGIG
VGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSGLTTNVVDIILAHHFGQEELQANLD
LIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERPILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVE
CNSRLNPINTTLLKMADCGGLPQVVQPGKLTEAFKYFLQGMGYVPSASMTRLARSRTHSTSSSLGSGE
SPFSRSVTSNQSDGTQESCESPDVLDRHQTMEVSC

NOV63b, CG56870-06 DNA Sequence	SEQ ID NO: 943	1041 bp
	ORF Start: ATG at 120	ORF Stop: TAA at 969

ACTTCTTTCTTTTCTGTTTCAGAGTTACTGATTTATTCTTGAGATTCCTCTACTCTCGTTATCTGACC
TCATGGATGAACTTCAGGATGTTCAGCTCACAGAGATCAAACCACTTCTAA**ATG**ATAAGAATGGTACA
AGAAACTTCCAGGACTTTGACTGTCAGGTATCAGTACCCCACAATGGATGAGCTGGCTGAAATGCTGC
CTCCTGTTCTTACCCACCTAAGCCTGAAAAGCATCATTGGAATTGGAGTTGGAGCTGGAGCTTACATC
CTCAGCAGATTTGCACTCAACCATCCAGAGCTTGTGGAAGGCCTTGTGCTCATTAATGTTGACCCTTG
CGCTAAAGGCTGGATTGACTGGGCAGCTTCCAAACTCTCTGGCCTGACAACCAATGTTGTGGACATTA
TTTTGGCTCATCACTTTGGGCAGGAAGAGTTACAGGCCAACCTGGACCTGATCCAAACCTACAGAATG
CATATTGCCCAAGACATCAACCAAGACAACCTGCAGCTCTTCTTGAATTCCTACAATGGGCGCAGAGA
CCTGGAGATCGAAAGACCCATACTGGGCCAAAATGATAACAAATCAAAAACATTAAAGTGTTCTACTT
TACTGGTGGTAGGGGACAATTCGCCTGCAGTTGAGGCTGTGGTCGAATGCAATTCCCGCCTGAACCCT
ATAAATACAACTTTGCTAAAGATGGCGGACTGTGGGGGACTGCCCCAGGTAGTTCAGCCTGGGAAGCT
CACCGAGGCCTTCAAGTACTTTTTGCAGGGAATGGGCTACGTCCCGTCTGCCAGCATGACTCGGCTCG
CCCGATCACGAACCCACTCAACCTCGAGTAGCCTCGGCTCTGGAGAAAGTCCCTTCAGCCGGTCTGTC
ACCAGCAATCAGTCAGATGGAACTCAAGAATCCTGTGAGTCCCCTGATGTCCTGGACAGACACCAGAC
CATGGAGGTGTCCTGC**TAA**GCAGATGCTCCTCCCCTGGACCATTGCAAGTCCATCCTTCAAATGACCA

CTCCATAATATAACATTTCAT

NOV63b, CG56870-06 Protein Sequence	SEQ ID NO: 944	283 aa	MW at 31109.1kD

MIRMVQETSRTLTVRYQYPTMDELAEMLPPVLTHLSLKSIIGIGVGAGAYILSRFALNHPELVEGLVL
INVDPCAKGWIDWAASKLSGLTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNS
YNGRRDLEIERPILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADCGGLPQV
VQPGKLTEAFKYFLQGMGYVPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQSDGTQESCESPDV
LDRHQTMEVSC

NOV63c, 276585681 DNA Sequence	SEQ ID NO: 945	994 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCAAGCTTACAAGAAACTTCCAGGACTTTGACTGTCAGGAACATGATATAGAAACAACTCATGGTG
TGGTCCACGTCACTATAAGAGGCTTACCCAAAGGAAACAGACCAGTTATACTAACATATCATGACATT
GGCCTCAACCATAAATCCTGTTTCAATGCATTCTTTAACTTTGAGGATATGCAAGAGATCACCCAGCA
CTTTGCTGTCTGTCATGTGGATGCCCCAGGCCAGCAGGAAGGTGCACCCTCTTTCCCAACAGGGTATC
AGTACCCCACAATGGATGAGCTGGCTGAAGTGCTGCCTCCTGTTCTTACCCACCTAAGCCTGAAAAGC
ATCATTGGAATTGGAGTTGGAGCTGGAGCTTACATCCTCAGCAGATTTGCACTCAACCATCCAGAGCT
TGTGGAAGGCCTTGTGCTCATTAATGTTGACCCTTGCGCTAAAGGCTGGATTGACTGGGCAGCTTCCA
AACTCTCTGGCCTGACAACCAATGTTGTGGACATTATTTTGGCTCATCACTTTGGGCAGGAAGAGTTA
CAGGCCAACCTGGACCTGATCCAAACCTACAGAATGCATATTGCCCAAGACATCAACCAAGACAACCT
GCAGCTCTTCTTGAATTCCTACAATGGACGCAGAGACCTGGAGATCGAAAGACCCATACTGGGCCAAA
ATGATAACAAATCAAAAACATTAAAGTGTTCTACTTTACTGGTGGTAGGGGACAATTCGCCTGCAGTT
GAGGCTGTGGTCGAATGCAATTCCCGCCTGAACCCTATAAATACAACTTTGCTAAAGATGGCGGACTG
TGGGGGACTGCCCCAGGTAGTTCAGCCTGGGAAGCTCACCGAGGCCTTCAAGTACTTTTTGCAGGGAA
TGGGCTACATACCATCTGCCAGCATGACTCGGCTCGCCCGATCACGAACCCACTCAACCTCGAGTAGC
CTCGGCTCTGGAGAAAGTCCCTTCAGCCGGTCTGTCGACGGC

NOV63c, 276585681 Protein Sequence	SEQ ID NO: 946	331 aa	MW at 36428.9kD

TKLTRNFQDFDCQEHDIETTHGVVHVTIRGLPKGNRPVILTYHDIGLNHKSCFNAFFNFEDMQEITQH
FAVCHVDAPGQQEGAPSFPTGYQYPTMDELAEVLPPVLTHLSLKSIIGIGVGAGAYILSRFALNHPEL
VEGLVLINVDPCAKGWIDWAASKLSGLTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNL
QLFLNSYNGRRDLEIERPILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADC
GGLPQVVQPGKLTEAFKYFLQGMGYIPSASMTRLARSRTHSTSSSLGSGESPFSRSVDG

NOV63d, CG56870-02 DNA Sequence	SEQ ID NO: 947	1175 bp
	ORF Start: ATG at 16	ORF Stop: TAA at 1141

TCGTTATCTGACCTC**ATG**GATGAACTTCAGGATGTTCAGCTCACAGAGATCAAACCACTTCTAAATGA
TAAGAATGGTACAAGAAACTTCCAGGACTTTGACTGTCAGGAACATGATATAGAAACAACTCATGGTG
TGGTCCACGTCACTATAAGAGGCTTACCCAAAGGAAACAGACCAGTTATACTAACATATCATGACATT
GGCCTCAACCATAAATCCTGTTTCAATGCATTCTTTAACTTTGAGGATATGCAAGAGATCACCCAGCA
CTTTGCTGTCTGTCATGTGGATGCCCCAGGCCAGCAGGAAGGTGCACCCTCTTTCCCAACAGGGTATC
AGTACCCCACAATGGATGAGCTGGCTGAAGTGCTGCCTCCTGTTCTTACCCACCTAAGCCTGAAAAGC
ATCATTGGAATTGGAGTTGGAGCTGGAGCTTACATCCTCAGCAGATTTGCACTCAACCATCCAGAGCT
TGTGGAAGGCCTTGTGCTCATTAATGTTGACCCTTGCGCTAAAGGCTGGATTGACTGGGCAGCTTCCA
AACTCTCTGGCCTGACAACCAATGTTGTGGACATTATTTTGGCTCATCACTTTGGGCAGGAAGAGTTA
CAGGCCAACCTGGACCTGATCCAAACCTACAGAATGCATATTGCCCAAGACATCAACCAAGACAACCT
GCAGCTCTTCTTGAATTCCTACAATGGACGCAGAGACCTGGAGATCGAAAGACCCATACTGGGCCAAA
ATGATAACAAATCAAAAACATTAAAGTGTTCTACTTTACTGGTGGTAGGGGACAATTCGCCTGCAGTT
GAGGCTGTGGTCGAATGCAATTCCCGCCTGAACCCTATAAATACAACTTTGCTAAAGATGGCGGACTG
TGGGGGACTGCCCCAGGTAGTTCAGCCTGGGAAGCTCACCGAGGCCTTCAAGTACTTTTTGCAGGGAA
TGGGCTACATACCATCTGCCAGCATGACTCGGCTCGCCCGATCACGAACCCACTCAACCTCGAGTAGC
CTCGGCTCTGGAGAAAGTCCCTTCAGCCGGTCTGTCACCAGCAATCAGTCAGATGGAACTCAAGAATC
CTGTGAGTCCCCTGATGTCCTGGACAGACACCAGACCATGGAGGTGTCCTGCT**AA**GCAGATGCTCCTC
CCCTGGACCATTGCAAGTC

NOV63d, CG56870-02 Protein Sequence	SEQ ID NO: 948	375 aa	MW at 41376.2kD

MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQEHDIETTHGVVHVTIRGLPKGNRPVILTYHDIGLNHK
SCFNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDELAEVLPPVLTHLSLKSIIGIG
VGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSGLTTNVVDIILAHHFGQEELQANLD
LIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERPILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVE
CNSRLNPINTTLLKMADCGGLPQVVQPGKLTEAFKYFLQGMGYIPSASMTRLARSRTHSTSSSLGSGE
SPFSRSVTSNQSDGTQESCESPDVLDRHQTMEVSC

NOV63e, CG56870-03 DNA Sequence	SEQ ID NO: 949	1232 bp
	ORF Start: ATG at 71	ORF Stop: TAA at 1160

ACTTCTTTCTTTTCTGTTTCAGAGTTACTGATTTATTCTTGAGATTCCTCTACTCTCGTTATCTGACC
TCATGGATGAACTTCAGGATGTTCAGCTCACAGAGATCAAACCACTTCTAAATGATAAGGAACATGAT
ATAGAAACAACTCATGGTGTGGTCCACGTCACTATAAGAGGCTTACCCAAAGGAAACAGACCAGTTAT
ACTAACATATCATGACATTGGCCTCAACCATAAATCCTGTTTCAATGCATTCTTTAACTTTGAGGATA
TGCAAGAGATCACCCAGCACTTTGCTGTCTGTCATGTGGATGCCCCAGGCCAGCAGGAAGGTGCACCC
TCTTTCCCAACAGGGTATCAGTACCCCACAATGGATGAGCTGGCTGAAATGCTGCCTCCTGTTCTTAC
CCACCTAAGCCTGAAAAGCATCATTGGAATTGGAGTTGGAGCTGGAGCTTACATCCTCAGCAGATTTG
CACTCAACCATCCAGAGCTTGTGGAAGGCCTTGTGCTCATTAATGTTGACCCTTGCGCTAAAGGCTGG
ATTGACTGGGCAGCTTCCAAACTCTCTGGCCTGACAACCAATGTTGTGGACATTATTTTGGCTCATCA
CTTTGGGCAGGAAGAGTTACAGGCCAACCTGGACCTGATCCAAACCTACAGAATGCATATTGCCCAAG
ACATCAACCAAGACAACCTGCAGCTCTTCTTGAATTCCTACAATGGGCGCAGAGACCTGGAGATCGAA
AGACCCATACTGGGCCAAAATGATAACAAATCAAAAACATTAAAGTGTTCTACTTTACTGGTGGTAGG
GGACAATTCGCCTGCAGTTGAGGCTGTGGTCGAATGCAATTCCCGCCTGAACCCTATAAATACAACTT
TGCTAAAGATGGCGGACTGTGGGGGACTGCCCCAGGTAGTTCAGCCTGGGAAGCTCACCGAGGCCTTC
AAGTACTTTTTGCAGGGAATGGGCTACGTCCCGTCTGCCAGCATGACTCGGCTCGCCCGATCACGAAC
CCACTCAACCTCGAGTAGCCTCGGCTCTGGAGAAAGTCCCTTCAGCCGGTCTGTCACCAGCAATCAGT
CAGATGGAACTCAAGAATCCTGTGAGTCCCCTGATGTCCTGGACAGACACCAGACCATGGAGGTGTCC
TGCTAAGCAGATGCTCCTCCCCTGGACCATTGCAAGTCCATCCTTCAAATGACCACTCCATAATATAA
CATTTCAT

NOV63e, CG56870-03 Protein Sequence	SEQ ID NO: 950	363 aa	MW at 39967.8kD

MDELQDVQLTEIKPLLNDKEHDIETTHGVVHVTIRGLPKGNRPVILTYHDIGLNHKSCFNAFFNFEDM
QEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDELAEMLPPVLTHLSLKSIIGIGVGAGAYILSRFA
LNHPELVEGLVLINVDPCAKGWIDWAASKLSGLTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQD
INQDNLQLFLNSYNGRRDLEIERPILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTL
LKMADCGGLPQVVQPGKLTEAFKYFLQGMGYVPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQS
DGTQESCESPDVLDRHQTMEVSC

NOV63f, CG56870-04 DNA Sequence	SEQ ID NO: 951	1220 bp
	ORF Start: ATG at 71	ORF Stop: TAA at 1148

ACTTCTTTCTTTTCTGTTTCAGAGTTACTGATTTATTCTTGAGATTCCTCTACTCTCGTTATCTGACC
TCATGGATGAACTTCAGGATGTTCAGCTCACAGAGATCAAACCACTTCTAAATGATAAGAATGGTACA
AGAAACTTCCAGGACTTTGACTGTCAGGAACATGATATAGAAACAACTCATGGTGTGGTCCACGTCAC
TATAAGAGGCTTACCCAAAGGAAACAGACCAGTTATACTAACATATCATGACATTGGCCTCAACCGTA
AATCCTGTTTCAATGCATTCTTTAACTTTGAGGATATGCAAGAGATCACCCAGCACTTTGCTGTCTGT
CATGTGGATGCCCCAGGCCAGCAGGAAGGTGCACCCTCTTTCCCAACAGGGTATCAGTACCCCACAAT
GGATGAGCTGGCTGAAATGCTGCCTCCTGTTCTTACCCACCTAAGCCTGAAAAGCATCATTGGAATTG
GAGTTGGAGCTGGAGCTTACATCCTCAGCAGATTTGCACTCAACCATCCAGAGCTTGTGGAAGGCCTT
GTGCTCATTAATGTTGACCCTTGCGCTAAAGGCTGGATTGACTGGGCAGCTTCCAAACTCTCTGGCCT
GACAACCAATGTTGTGGACATTATTTTGGCTCATCACTTTGGGCAGGAAGAGTTACAGGCCAACCTGG
ACCTGATCCAAACCTACAGAATGCATATTGCCCAAGACATCAACCAAGACAACCTGCAGCTCTTCTTG
AATTCCTACAATGGACGCAGAGACCTGGAGATCGAAAGACCCATACTGGGCCAAAATGATAACAAATC
AAAAACATTAAAGTGTTCTACTTTACTGGTGGTAGGGGACAATTCGCCTGCAGTTGAGGCTGTGATGG
CGGACTGTGGGGGACTGCCCCAGGTAGTTCAGCCTGGGAAGTTCACCGAGGCCTTCAAGTACTTTTTG
CAGGGAATGGGCTACACACCATCTGCCAGCATGACTCGGCTCGCCCGATCACGAACCCACTCAACCTC

GAGTAGCCTCGGCTCTGGAGAAAGTCCCTTCAGCCGGTCTGTCACCAGCAATCAGTCAGATGGAACTC
AAGAATCCTGTGAGTCCCCTGATGTCCTGGACAGACACCAGACCATGGAGGTGTCCTGC**TAA**GCAGAT
GCTCCTCCCCTGGACCATTGCAAGTCCATCCTTCAAATGACCACTCCATAATATAACATTTCAT

NOV63f, CG56870-04 Protein Sequence	SEQ ID NO: 952	359 aa	MW at 39652.2kD

MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQEHDIETTHGVVHVTIRGLPKGNRPVILTYHDIGLNRK
SCFNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDELAEMLPPVLTHLSLKSIIGIG
VGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSGLTTNVVDIILAHHFGQEELQANLD
LIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERPILGQNDNKSKTLKCSTLLVVGDNSPAVEAVMA
DCGGLPQVVQPGKFTEAFKYFLQGMGYTPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQSDGTQ
ESCESPDVLDRHQTMEVSC

NOV63g, CG56870-05 DNA Sequence	SEQ ID NO: 953	970 bp
	ORF Start: ATG at 1	ORF Stop: TAA at 898

ATGGATGAACTTCAGGATGTTCAGCTCACAGAGATCAAACCACTTCTAAATGATAAGAATGGTACAAG
AAACTTCCAGGACTTTGACTGTCAGTATCAGTACCCCACAATGGATGAGCTGGCTGAAATGCTGCCTC
CTGTTCTTACCCACCTAAGCCTGAAAAGCATCATTGGAATTGGAGTTGGAGCTGGAGCTTACATCCTC
AGCAGATTTGCACTCAACCATCCAGAGCTTGTGGAAGGCCTTGTGCTCATTAATGTTGACCCTTGCGC
TAAAGGCTGGATTGACTGGGCAGCTTCCAAACTCTCTGGCCTGACAACCAATGTTGTGGACATTATTT
TGGCTCATCACTTTGGGCAGGAAGAGTTACAGGCCAACCTGGACCTGATCCAAACCTACAGAATGCAT
ATTGCCCAAGACATCAACCAAGACAACCTGCAGCTCTTCTTGAATTCCTACAATGGGCGCAGAGACCT
GGAGATCGAAAGACCCATACTGGGCCAAAATGATAACAAATCAAAAACATTAAAGTGTTCTACTTTAC
TGGTGGTAGGGGACAATTCGCCTGCAGTTGAGGCTGTGGTCGAATGCAATTCCCGCCTGAACCCTATA
AATACAACTTTGCTAAAGATGGCGGACTGTGGGGGACTGCCCCAGGTAGTTCAGCCTGGGAAGCTCAC
CGAGGCCTTCAAGTACTTTTTGCAGGGAATGGGCTACGTCCCGTCTGCCAGCATGACTCGGCTCGCCC
GATCACGAACCCACTCAACCTCGAGTAGCCTCGGCTCTGGAGAAAGTCCCTTCAGCCGGTCTGTCACC
AGCAATCAGTCAGATGGAACTCAAGAATCCTGTGAGTCCCCTGATGTCCTGGACAGACACCAGACCAT
GGAGGTGTCCTGC**TAA**GCAGATGCTCCTCCCCTGGACCATTGCAAGTCCATCCTTCAAATGACCACTC
CATAATATAACATTTCAT

NOV63g, CG56870-05 Protein Sequence	SEQ ID NO: 954	299 aa	MW at 32956.9kD

MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQYQYPTMDELAEMLPPVLTHLSLKSIIGIGVGAGAYIL
SRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSGLTTNVVDIILAHHFGQEELQANLDLIQTYRMH
IAQDINQDNLQLFLNSYNGRRDLEIERPILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPI
NTTLLKMADCGGLPQVVQPGKLTEAFKYFLQGMGYVPSASMTRLARSRTHSTSSSLGSGESPFSRSVT
SNQSDGTQESCESPDVLDRHQTMEVSC

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 63B.

Table 63B. Comparison of the NOV63 protein sequences.

```
NOV63a  MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQEHDIETTHGVVHVTIRGLPKGNRPVILTY
NOV63b  ------------------------------------------------------------
NOV63c  ----------------TKLTRNFQDFDCQEHDIETTHGVVHVTIRGLPKGNRPVILTY
NOV63d  MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQEHDIETTHGVVHVTIRGLPKGNRPVILTY
NOV63e  ----------MDELQDVQLTEIKPLLNDKEHDIETTHGVVHVTIRGLPKGNRPVILTY
NOV63f  MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQEHDIETTHGVVHVTIRGLPKGNRPVILTY
NOV63g  ------------------------------------------------------------

NOV63a  HDIGLNRKSCFNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDELAEML
NOV63b  -------------------------MIRMVQETS-----RTLTVRYQYPTMDELAEML
NOV63c  HDIGLNHKSCFNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDELAEVL
NOV63d  HDIGLNHKSCFNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDELAEVL
```

```
NOV63e  HDIGLNHKSCFNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDELAEML
NOV63f  HDIGLNRKSCFNAFFNFEDMQEITQHFAVCHVDAPGQQEGAPSFPTGYQYPTMDELAEML
NOV63g  ----------------MDELQDVQLTEIKPLLNDKNGTRNFQDFDCQYQYPTMDELAEML

NOV63a  PPVLTHLSLKSIIGIGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSG
NOV63b  PPVLTHLSLKSIIGIGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSG
NOV63c  PPVLTHLSLKSIIGIGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSG
NOV63d  PPVLTHLSLKSIIGIGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSG
NOV63e  PPVLTHLSLKSIIGIGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSG
NOV63f  PPVLTHLSLKSIIGIGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSG
NOV63g  PPVLTHLSLKSIIGIGVGAGAYILSRFALNHPELVEGLVLINVDPCAKGWIDWAASKLSG

NOV63a  LTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERP
NOV63b  LTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERP
NOV63c  LTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERP
NOV63d  LTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERP
NOV63e  LTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERP
NOV63f  LTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERP
NOV63g  LTTNVVDIILAHHFGQEELQANLDLIQTYRMHIAQDINQDNLQLFLNSYNGRRDLEIERP

NOV63a  ILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADCGGLPQVVQPG
NOV63b  ILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADCGGLPQVVQPG
NOV63c  ILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADCGGLPQVVQPG
NOV63d  ILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADCGGLPQVVQPG
NOV63e  ILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADCGGLPQVVQPG
NOV63f  ILGQNDNKSKTLKCSTLLVVGDNSPAVEAV----------------MADCGGLPQVVQPG
NOV63g  ILGQNDNKSKTLKCSTLLVVGDNSPAVEAVVECNSRLNPINTTLLKMADCGGLPQVVQPG

NOV63a  KLTEAFKYFLQGMGYVPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQSDGTQESCE
NOV63b  KLTEAFKYFLQGMGYVPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQSDGTQESCE
NOV63c  KLTEAFKYFLQGMGYIPSASMTRLARSRTHSTSSSLGSGESPFSRSVDG-----------
NOV63d  KLTEAFKYFLQGMGYIPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQSDGTQESCE
NOV63e  KLTEAFKYFLQGMGYVPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQSDGTQESCE
NOV63f  KFTEAFKYFLQGMGYTPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQSDGTQESCE
NOV63g  KLTEAFKYFLQGMGYVPSASMTRLARSRTHSTSSSLGSGESPFSRSVTSNQSDGTQESCE

NOV63a  SPDVLDRHQTMEVSC
NOV63b  SPDVLDRHQTMEVSC
NOV63c  ---------------
NOV63d  SPDVLDRHQTMEVSC
NOV63e  SPDVLDRHQTMEVSC
NOV63f  SPDVLDRHQTMEVSC
NOV63g  SPDVLDRHQTMEVSC

NOV63a  (SEQ ID NO:  942)
NOV63b  (SEQ ID NO:  944)
NOV63c  (SEQ ID NO:  946)
NOV63d  (SEQ ID NO:  948)
NOV63e  (SEQ ID NO:  950)
NOV63f  (SEQ ID NO:  952)
NOV63g  (SEQ ID NO:  954)
```

Further analysis of the NOV63a protein yielded the following properties shown in Table 63C.

Table 63C. Protein Sequence Properties NOV63a	
SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 0;  neg.chg 4
      H-region:  length 1;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1): -14.03
      possible cleavage site: between 48 and 49

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   0
      PERIPHERAL  Likelihood =  4.67 (at 174)
      ALOM score:   0.47  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  8.01
      Hyd Moment(95):  7.59    G content:       0
      D/E content:     2       S/T content:     0
      Score: -6.61

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.5%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
```

```
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 55.5

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       39.1 %: cytoplasmic
       39.1 %: nuclear
       13.0 %: mitochondrial
        4.3 %: Golgi
        4.3 %: endoplasmic reticulum

>> prediction for CG56870-01 is cyt (k=23)
```

A search of the NOV63a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 63D.

Table 63D. Geneseq Results for NOV63a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV63a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB94494	Human protein sequence SEQ ID NO:15186 - Homo sapiens, 363 aa. [EP1074617-A2, 07-FEB-2001]	1..375 1..363	360/375 (96%) 361/375 (96%)	0.0
AAG64392	Human reducing agent and tunicamycin-responsive protein 40 - Homo sapiens, 363 aa. [WO200155375-A1, 02-AUG-2001]	1..375 1..363	360/375 (96%) 361/375 (96%)	0.0
AAM94019	Human stomach cancer expressed polypeptide SEQ ID NO 108 - Homo sapiens, 363 aa. [WO200109317-A1, 08-FEB-2001]	1..375 1..363	360/375 (96%) 361/375 (96%)	0.0
ABG10563	Novel human diagnostic protein #10554 - Homo sapiens, 382 aa. [WO200175067-A2, 11-OCT-2001]	68..374 1..307	283/323 (87%) 287/323 (88%)	e-155
AAU31598	Novel human secreted protein #2089 - Homo sapiens, 395 aa. [WO200179449-A2, 25-OCT-2001]	68..374 1..307	282/323 (87%) 286/323 (88%)	e-154

In a BLAST search of public sequence databases, the NOV63a protein was found to have homology to the proteins shown in the BLASTP data in Table 63E.

Table 63E. Public BLASTP Results for NOV63a				
Protein Accession Number	**Protein/Organism/Length**	**NOV63a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9UGV2	NDRG3 protein - Homo sapiens (Human), 375 aa.	1..375 1..375	373/375 (99%) 374/375 (99%)	0.0
Q9QYF9	NDRG3 protein (Ndr3 protein) - Mus musculus (Mouse), 375 aa.	1..375 1..375	358/375 (95%) 368/375 (97%)	0.0
Q8VCV2	Similar to N-myc downstream regulated 3 - Mus musculus (Mouse), 388 aa.	1..375 1..388	359/388 (92%) 368/388 (94%)	0.0
Q96SM2	Hypothetical protein FLJ14759 - Homo sapiens (Human), 363 aa.	1..375 1..363	360/375 (96%) 361/375 (96%)	0.0
AAH44139	Similar to N-myc downstream regulated 3 - Brachydanio rerio (Zebrafish) (Danio rerio), 371 aa.	1..375 2..371	290/375 (77%) 324/375 (86%)	e-169

PFam analysis indicates that the NOV63a protein contains the domains shown in the Table 63F.

Table 63F. Domain Analysis of NOV63a			
Pfam Domain	**NOV63a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
abhydrolase	87..307	46/237 (19%) 142/237 (60%)	0.0014
Ndr	32..317	181/301 (60%) 275/301 (91%)	5.6e-186

Example 64.

The NOV64 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 64A.

Table 64A. NOV64 Sequence Analysis		
NOV64a, CG57109-01 DNA Sequence	SEQ ID NO: 955	2536 bp
	ORF Start: ATG at 150	ORF Stop: TAG at 2094
GGGACACTGACATGGACTGAAGGAGTAGAAAAGAAGCCTTGGGCTCTCCCAGATGGAAGAATGACCGT		
GTGAGGAAACTGTTTAACCTCAAGGGCAGGGAAATCAGGAGCGTCTCTGATTTCTTCAGGGAAGGGGA		

TGCTTTCATAGCTATGGGCAAAGAACCACTGACACTGAAGAGCATTCAGGTGGCTGTAGAAGAACTGT
ACCCCAACAAAGCCCGGGCCCTGACACTGGCCCAGCACAGCCGTGCCCCTTCTCCAAGGCTGAGGAGC
AGGCTGTTTAGCAAGGCTCTGAAAGGAGACCACCGCTGTGGGGAGACCGAGACCCCCAAGAGCTGCAG
CGAAGTTGCAGGATGCAAGGCAGCCATGAGGCACCAGGGGAAGATCCCCGAGGAGCTTTCACTAGATG
ACAGAGCGAGGACCCAGAAGAAGTGGGGGAGGGGGAAATGGGAGCCAGAACCCAGTAGCAAGCCCCCC
AGGGAAGCCACTCTGGAAGAGAGGCACGCAAGGGGAGAGAAGCATCTTGGGGTGGAGATTGAAAAGAC
CTCGGGTGAAATTATCAGATGCGAGAAGTGCAAGAGAGAGAGGGAGCTTCAGCAGAGCCTGGAGCGTG
AGAGGCTTTCTCTGGGGACCAGTGAGCTGGATATGGGGAAGGGCCCAATGTATGATGTGGAGAAGCTG
GTGAGGACCAGAAGCTGCAGGAGGTCTCCCGAGGCAAATCCTGCAAGTGGGGAGGAAGGGTGGAAGGG
TGACAGCCACAGGAGCAGCCCCAGGAATCCCACTCAAGAGCTGAGGAGACCCAGCAAGAGCATGGACA
AGAAAGAGGACAGAGGCCCAGAGGATCAAGAAAGCCATGCTCAGGGAGCAGCCAAGGCCAAGAAGGAC
CTTGTGGAAGTTCTTCCTGTCACAGAGGAGGGGCTGAGGGAGGTGAAGAAGGACACCAGGCCCATGAG
CAGGAGCAAACATGGTGGCTGGCTCCTGAGAGAGCACCAGGCGGGCTTTGAGAAGCTCCGCAGGACCC
GAGGAGAAGAGAAGGAGGCAGAGAAGGAGAAAAAGCCATGTATGTCTGGAGGCAGAAGGATGACTCTC
AGAGATGACCAACCTGCAAAGCTAGAAAAGGAGCCCAAGACGAGGCCAGAAGAGAACAAGCCAGAGCG
GCCCAGCGGTCGGAAGCCACGGCCCATGGGCATCATTGCCGCCAATGTGGAAAAGCATTATGAGACTG
GCCGGGTCATTGGGGATGGGAACTTTGCTGTCGTGAAGGAGTGCAGACACCGCGAGACCAGGCAGGCC
TATGCGATGAAGATCATTGACAAGTCCAGACTCAAGGGCAAGGAGGACATGGTGGACAGTGAGATCTT
GATCATCCAGAGCCTCTCTCACCCCAACATCGTGAAATTGCATGAAGTCTACGAAACAGACATGGAAA
TCTACCTGATCCTGGAGTACGTGCAGGGAGGAGACCTTTTTGACGCCATCATAGAAAGTGTGAAGTTC
CCGGAGCCCGATGCTGCCCTCATGATCATGGACTTATGCAAAGCCCTCGTCCACATGCACGACAAGAG
CATTGTCCACCGGGACCTCAAGCCGGAAAACCTTTTGGTTCAGCGAAATGAGGACAAATCTACTACCT
TGAAATTGGCTGATTTTGGACTTGCAAAGCATGTGGTGAGACCTATATTTACTGTGTGTGGGACCCCA
ACTTACGTAGCTCCCGAAATTCTTTCTGAGAAAGGTTATGGACTGGAGGTGGACATGTGGGCGGCTGG
CGTGATCCTCTATATCCTGCTGTGTGGCTTTCCCCCATTCCGCAGCCCTGAGAGGGACCAGGACGAGC
TCTTTAACATCATCCAGCTGGGCCACTTTGAGTTCCTCCCCCCTTACTGGGACAATATCTCTGATGCT
GCTAAAGATCTGGTGAGCCGGTTGCTGGTGGTAGACCCCAAAAAGCGCTACACAGCTCATCAGGTTCT
TCAGCACCCCTGGATCGAAACAGCTGGCAAGACCAATACAGTGAAACGACAGAAGCAGGTGTCCCCCA
GCAGCGAGGGTCACTTCCGGAGCCAGCACAAGAGGGTTGTGGAGCAGGTATCATAGTCACCACCTTGG
GAATCTGTCCAGCCCCCAGTTCTGCTCAAGGACAGAGAAAAGGATAGAAGTTTGAGAGAAAAACAATG
AAAGAGGCTTCTTCACATAATTGGTGAATCAGAGGGAGAGACACTGAGTATATTTTAAAGCATATTAA
AAAAATTAAGTCAATGTTAAATGTCACAACATATTTTTAGATTTGTATATTTAAAGCCTTTAATACAT
TTTTGGGGGGTAAGCATTGTCATCAGTGAGGAATTTTGGTAATAATGATGTGTTTTGCTTCCCCTTTG
TAACCAAGTTTATTCTGTACTACAGGAGTGGTGCTTACCAGGGTCTAAACTCCCCCTGTGAGATTAAT
AAGGTGCATTGTGGTCTTTCTGTGTTAATAAAATGTGCTCTGAATAACAGAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAGG

NOV64a, CG57109-01 Protein Sequence	SEQ ID NO: 956	648 aa	MW at 73813.6kD

MGKEPLTLKSIQVAVEELYPNKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAG
CKAAMRHQGKIPEELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEI
IRCEKCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKGDSHR
SSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREVKKDTRPMSRSKH
GGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAKLEKEPKTRPEENKPERPSGR
KPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQAYAMKIIDKSRLKGKEDMVDSEILIIQS
LSHPNIVKLHEVYETDMEIYLILEYVQGGDLFDAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHR
DLKPENLLVQRNEDKSTTLKLADFGLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILY
ILLCGFPPFRSPERDQDELFNIIQLGHFEFLPPYWDNISDAAKDLVSRLLVVDPKKRYTAHQVLQHPW
IETAGKTNTVKRQKQVSPSSEGHFRSQHKRVVEQVS

NOV64b, 260457400 DNA Sequence	SEQ ID NO: 957	793 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCTATGAGACTGGCCGGGTCATTGGGGATGGGAACTTTGCTGTCGTGAAGGAGTGCAGAC
ACCGCGAGACCAGGCAGGCCTATGCGATGAAGATCATTGACAAGTCCAGACTCAAGGGCAAGGAGGAC
ATGGTGGACAGTGAGATCTTGATCATCCAGAGCCTCTCTCACCCCAACATCGTGAAATTGCATGAAGT
CTACGAAACAGACATGGAAATCTACCTGATCCTGGAGTACGTGCAGGGAGGAGACCTTTTTGACGCCA
TCATAGAAAGTGTGAAGTTCCCGGAGCCCGATGCTGCCCTCATGATCATGGACTTATGCAAAGCCCTC
GTCCACATGCACGACAAGAGCATTGTCCACCGGGACCTCAAGCCGGAAAACCTTTTGGTTCAGCGAAA
TGAGGACAAATCTACTACCTTGAAATTGGCTGATTTTGGACTTGCAAAGCATGTGGTGAGACCTATAT

TTACTGTGTGTGGGACCCCAACTTACGTAGCTCCCGAAATTCTTTCTGAGAAAGGTTATGGACTGGAG
GTGGACATGTGGGCTGCTGGCGTGATCCTCTATATCCTGCTGTGTGGCTTTCCCCCATTCCGCAGCCC
TGAGAGGGACCAGGACGAGCTCTTTAACATCATCCAGCTGGGCCACTTTGAGTTCCTCCCCCCTTACT
GGGACAATATCTCTGATGCTGCTAAAGATCTGGTGAGCCGGTTGCTGGTGGTAGACCCCAAAAAGCGC
TACACAGCTCATCAGGTTCTTCAGCACCCCTGGATCCTCGAGGGC

NOV64b, 260457400 Protein Sequence	SEQ ID NO: 958	264 aa	MW at 30106.4kD

TGSYETGRVIGDGNFAVVKECRHRETRQAYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEV
YETDMEIYLILEYVQGGDLFDAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRN
EDKSTTLKLADFGLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSP
ERDQDELFNIIQLGHFEFLPPYWDNISDAAKDLVSRLLVVDPKKRYTAHQVLQHPWILEG

NOV64c, 260457409 DNA Sequence	SEQ ID NO: 959	625 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCTATGAGACTGGCCGGGTCATTGGGGATGGGAACTTTGCTGTCGTGAAGGAGTGCAGAC
ACCGCGAGACCAGGCAGGCCTATGCGATGAAGATCATTGACAAGTCCAGACTCAAGGGCAAGGAGGAC
ATGGTGGACAGTGAGATCTTGATCATCCAGAGCCTCTCTCACCCCAACATCGTGAAATTGCATGAAGT
CTACGAAACAGACATGGAAATCTACCTGATCCTGGAGTACGTGCAGGGAGGAGACCTTTTTGACGCCA
TCATAGAAAGTGTGAAGTTCCCGGAGCCCGATGCTGCCCTCATGATCATGGACTTATGCAAAGCCCTC
GTCCACATGCACGACAAGAGCATTGTCCACCGGGACCTCAAGCCGGAAAACCTTTTGGTTCAGCGAAA
TGAGGACAAATCTACTACCTTGAAATTGGCTGATTTTGGACTTGCAAAGCATGTGGTGAGACCTATAT
TTACTGTGTGTGGGACCCCAACTTACGTAGCTCCCGAAATTCTTTCTGAGAAAGCTGCTAAAGATCTG
GTGAGCCGGTTGCTGGTGGTAGACCCCAAAAAGCGCTACACAGCTCATCAGGTTCTTCAGCACCCCTG
GATCCTCGAGGGC

NOV64c, 260457409 Protein Sequence	SEQ ID NO: 960	208 aa	MW at 23605.1kD

TGSYETGRVIGDGNFAVVKECRHRETRQAYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEV
YETDMEIYLILEYVQGGDLFDAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRN
EDKSTTLKLADFGLAKHVVRPIFTVCGTPTYVAPEILSEKAAKDLVSRLLVVDPKKRYTAHQVLQHPW
ILEG

NOV64d, CG57109-05 DNA Sequence	SEQ ID NO: 961	2133 bp
	ORF Start: ATG at 90	ORF Stop: TAG at 2034

TTGACCGTGTGAGGAAACTGTTTAACCTCAAGGGCAGGGAAATCAGGAGCGTCTCTGATTTCTTCAGG
GAAGGGGATGCTTTCATAGCTATGGGCAAAGAACCACTGACACTGAAGAGCATTCAGGTGGCTGTAGA
AGAACTGTACCCCAACAAAGCCCGGGCCCTGACACTGGCCCAGCACAGCCGTGCCCCTTCTCCAAGGC
TGAGGAGCAGGCTGTTTAGCAAGGCTCTGAAAGGAGACCACCGCTGTGGGGAGACCGAGACCCCCAAG
AGCTGCAGCGAAGTTGCAGGATGCAAGGCAGCCATGAGGCACCAGGGGAAGATCCCCGAGGAGCTTTC
ACTAGATGACAGAGCGAGGACCCAGAAGAAGTGGGGGAGGGGGAAATGGGAGCCAGAACCCAGTAGCA
AGCCCCCCAGGGAAGCCACTCTGGAAGAGAGGCACGCAAGGGGAGAGAAGCATCTTGGGGTGGAGATT
GAAAAGACCTCGGGTGAAATTATCAGATGCGAGAAGTGCAAGAGAGAGAGGGAGCTCCAGCAGAGCCT
GGAGCGTGAGAGGCTTTCTCTGGGGACCAGTGAGCTGGATATGGGGAAGGGCCCAATGTATGATGTGG
AGAAGCTGGTGAGGACCAGAAGCTGCAGGAGGTCTCCCGAGGCAAATCCTGCAAGTGGGGAGGAAGGG
TGGAAGGGTGACAGCCACAGGAGCAGCCCCAGGAATCCCACTCAAGAGCTGAGGAGACCCAGCAAGAG
CATGGACAAGAAAGAGGACAGAGGCCCAGAGGATCAAGAAAGCCATGCTCAGGGAGCGGCCAAGGCCA
AGAAGGACCTTGTGGAAGTTCTTCCTGTCACAGAGGAGGGGCTGAGAGAGGTGAAGAAGGACACCAGG
CCCATGAGCAGGAGCAAACATGGTGGCTGGCTCCTGAGAGAGCACCAGGCGGGCTTTGAGAAGCTCCG
CAGGACCCGAGGAGAAGAGAAGGAGGCAGAGAAGGAGAAAAAGCCATGTATGTCTGGAGGCAGAAGGA
TGACTCTCAGAGACGACCAACCTGCAAAGCTAGAAAAGGAGCCCAAGACGAGGCCAGAAGAGAACAAG
CCAGAGCGGCCCAGCGGTCGGAAGCCACGGCCCATGGGCATCATTGCCGCCAATGTGGAAAAGCATTA
TGAGACTGGCCGGGTCATTGGGGATGGGAACTTTGCTGTCGTGAAGGAGTGCAGACACCGCGAGACCA
GGCAGGCCTATGCGATGAAGATCATTGACAAGTCCAGACTCAAGGGCAAGGAGGACATGGTGGACAGT
GAGATCTTGATCATCCAGAGCCTCTCTCACCCCAACATCGTGAAATTGCATGAAGTCTACGAAACAGA
CATGGAAATCTACCTGATCCTGGAGTACGTGCAGGGAGGAGACCTTTTTGACGCCATCATAGAAAGTG
TGAAGTTCCCGGAGCCCGATGCTGCCCTCATGATCATGGACTTATGCAAAGCCCTCGTCCACATGCAC
GACAAGAGCATTGTCCACCGGGACCTCAAGCCGGAAAACCTTTTGGTTCAGCGAAATGAGGACAAATC

TACTACCTTGAAATTGGCTGATTTTGGACTTGCAAAGCATGTGGTGAGACCTATATTTACTGTGTGTG
GGACCCCAACTTACGTAGCTCCCGAAATTCTTTCTGAGAAAGGTTATGGACTGGAGGTGGACATGTGG
GCTGCTGGCGTGATCCTCTATATCCTGCTGTGTGGCTTTCCCCCATTCCGCAGCCCTGAGAGGGACCA
GGACGAGCTCTTTAACATCATCCAGCTGGGCCACTTTGAGTTCCTCCCCCCTTACTGGGACAATATCT
CTGATGCTGCTAAAGATCTGGTGAGCCGGTTGCTGGTGGTAGACCCCAAAAAGCGCTACACAGCTCAT
CAGGTTCTTCAGCACCCCTGGATCGAAACAGCTGGCAAGACCAATACAGTGAAACGACAGAAGCAGGT
GTCCCCCAGCAGCGAGGGTCACTTCCGGAGCCAGCACAAGAGGGTTGTGGAGCAGGTATCATAGTCAC
CACCTTGGGAATCTGTCCAGCCCCCAGTTCTGCTCAAGGACAGAGAAAAGGATAGAAGTTTGAGAGAA
AAACAATGAAAGAGGCTTCTTCACA

NOV64d, CG57109-05 Protein Sequence	SEQ ID NO: 962	648 aa	MW at 73813.6kD

MGKEPLTLKSIQVAVEELYPNKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAG
CKAAMRHQGKIPEELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEI
IRCEKCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKGDSHR
SSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREVKKDTRPMSRSKH
GGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAKLEKEPKTRPEENKPERPSGR
KPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQAYAMKIIDKSRLKGKEDMVDSEILIIQS
LSHPNIVKLHEVYETDMEIYLILEYVQGGDLFDAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHR
DLKPENLLVQRNEDKSTTLKLADFGLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILY
ILLCGFPPFRSPERDQDELFNIIQLGHFEFLPPYWDNISDAAKDLVSRLLVVDPKKRYTAHQVLQHPW
IETAGKTNTVKRQKQVSPSSEGHFRSQHKRVVEQVS

NOV64e, 267253965 DNA Sequence	SEQ ID NO: 963	1966 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCACCATGGGCAAAGAACCACTGACACTGAAGAGCATTCAGGTGGCTGTAGAAGAACTGT
ACCCCAACAAAGCCCGGGCCCTGACACTGGCCCAGCACAGCCGTGCCCCTTCTCCAAGGCTGAGGAGC
AGGCTGTTTAGCAAGGCTCTGAAAGGAGACCACCGCTGTGGGGAGACCGAGACCCCCAAGAGCTGCAG
CGAAGTTGCAGGATGCAAGGCAGCCATGAGGCACCAGGGGAAGATCCCCGAGGAGCTTTCACTAGATG
ACAGAGCGAGGACCCAGAAGAAGTGGGGGAGGGGGAAATGGGAGCCAGAACCCAGTAGCAAGCCCCCC
AGGGAAGCCACTCTGGAAGAGAGGCACGCAAGGGGAGAGAAGCATCTTGGGGTGGAGATTGAAAAGAC
CTCGGGTGAAATTATCAGATGCGAGAAGTGCAAGAGAGAGAGGGAGCTCCAGCAGAGCCTGGAGCGTG
AGAGGCTTTCTCTGGGGACCAGTGAGCTGGATATGGGGAAGGGCCCAATGTATGATGTGGAGAAGCTG
GTGAGGACCAGAAGCTGCAGGAGGTCTCCCGAGGCAAATCCTGCAAGTGGGGAGGAAGGGTGGAAGGG
TGACAGCCACAGGAGCAGCCCCAGGAATCCCACTCAAGAGCTGAGGAGACCCAGCAAGAGCATGGACA
AGAAAGAGGACAGAGGCCCAGAGGATCAAGAAAGCCATGCTCAGGGAGCGGCCAAGGCCAAGAAGGAC
CTTGTGGAAGTTCTTCCTGTCACAGAGGAGGGGCTGAGAGAGGTGAAGAAGGACACCAGGCCCATGAG
CAGGAGCAAACATGGTGGCTGGCTCCTGAGAGAGCACCAGGCGGGCTTTGAGAAGCTCCGCAGGACCC
GAGGAGAAGAGAAGGAGGCAGAGAAGGAGAAAAAGCCATGTATGTCTGGAGGCAGAAGGATGACTCTC
AGAGACGACCAACCTGCAAAGCTAGAAAAGGAGCCCAAGACGAGGCCAGAAGAGAACAAGCCAGAGCG
GCCCAGCGGTCGGAAGCCACGGCCCATGGGCATCATTGCCGCCAATGTGGAAAAGCATTATGAGACTG
GCCGGGTCATTGGGGATGGGAACTTTGCTGTCGTGAAGGAGTGCAGACACCGCGAGACCAGGCAGGCC
TATGCGATGAAGATCATTGACAAGTCCAGACTCAAGGGCAAGGAGGACATGGTGGACAGTGAGATCTT
GATCATCCAGAGCCTCTCTCACCCCAACATCGTGAAATTGCATGAAGTCTACGAAACAGACATGGAAA
TCTACCTGATCCTGGAGTACGTGCAGGGAGGAGACCTTTTTGACGCCATCATAGAAAGTGTGAAGTTC
CCGGAGCCCGATGCTGCCCTCATGATCATGGACTTATGCAAAGCCCTCGTCCACATGCACGACAAGAG
CATTGTCCACCGGGACCTCAAGCCGGAAAACCTTTTGGTTCAGCGAAATGAGGACAAATCTACTACCT
TGAAATTGGCTGATTTTGGACTTGCAAAGCATGTGGTGAGACCTATATTTACTGTGTGTGGGACCCCA
ACTTACGTAGCTCCCGAAATTCTTTCTGAGAAAGGTTATGGACTGGAGGTGGACATGTGGGCTGCTGG
CGTGATCCTCTATATCCTGCTGTGTGGCTTTCCCCCATTCCGCAGCCCTGAGAGGGACCAGGACGAGC
TCTTTAACATCATCCAGCTGGGCCACTTTGAGTTCCTCCCCCCTTACTGGGACAATATCTCTGATGCT
GCTAAAGATCTGGTGAGCCGGTTGCTGGTGGTAGACCCCAAAAAGCGCTACACAGCTCATCAGGTTCT
TCAGCACCCCTGGATCGAAACAGCTGGCAAGACCAATACAGTGAAACGACAGAAGCAGGTGTCCCCCA
GCAGCGAGGGTCACTTCCGGAGCCAGCACAAGAGGGTTGTGGAGCAGGTATCACTCGAGGGC

NOV64e, 267253965 Protein Sequence	SEQ ID NO: 964	655 aa	MW at 74459.2kD

TGSTMGKEPLTLKSIQVAVEELYPNKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCS
EVAGCKAAMRHQGKIPEELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKT

SGEIIRCEKCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKG
DSHRSSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREVKKDTRPMS
RSKHGGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAKLEKEPKTRPEENKPER
PSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQAYAMKIIDKSRLKGKEDMVDSEIL
IIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLFDAIIESVKFPEPDAALMIMDLCKALVHMHDKS
IVHRDLKPENLLVQRNEDKSTTLKLADFGLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAG
VILYILLCGFPPFRSPERDQDELFNIIQLGHFEFLPPYWDNISDAAKDLVSRLLVVDPKKRYTAHQVL
QHPWIETAGKTNTVKRQKQVSPSSEGHFRSQHKRVVEQVSLEG

NOV64f, 267254000 DNA Sequence	SEQ ID NO: 965	1621 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCACCATGGGCAAAGAACCACTGACACTGAAGAGCATTCAGGTGGCTGTAGAAGAACTGT
ACCCCAACAAAGCCCGGGCCCTGACACTGGCCCAGCACAGCCGTGCCCCTTCTCCAAGGCTGAGGAGC
AGGCTGTTTAGCAAGGCTCTGAAAGGAGACCACCGCTGTGGGGAGACCGAGACCCCCAAGAGCTGCAG
CGAAGTTGCAGGATGCAAGGCAGCCATGAGGCACCAGGGGAAGATCCCCGAGGAGCTTTCACTAGATG
ACAGAGCGAGGACCCAGAAGAAGTGGGGGAGGGGGAAATGGGAGCCAGAACCCAGTAGCAAGCCCCCC
AGGGAAGCCACTCTGGAAGAGAGGCACGCAAGGGGAGAGAAGCATCTTGGGGTGGAGATTGAAAAGAC
CTCGGGTGAAATTATCAGATGCGAGAAGTGCAAGAGAGAGAGGGAGCTCCAGCAGAGCCTGGAGCGTG
AGAGGCTTTCTCTGGGGACCAGTGAGCTGGATATGGGGAAGGGCCCAATGTATGATGTGGAGAAGCTG
GTGAGGACCAGAAGCTGCAGGAGGTCTCCCGAGGCAAATCCTGCAAGTGGGGAGGAAGGGTGGAAGGG
TGACAGCCACAGGAGCAGCCCCAGGAATCCCACTCAAGAGCTGAGGAGACCCAGCAAGAGCATGGACA
AGAAAGAGGACAGAGGCCCAGAGGATCAAGAAAGCCATGCTCAGGGAGCGGCCAAGGCCAAGAAGGAC
CTTGTGGAAGTTCTTCCTGTCACAGAGGAGGGGCTGAGAGAGGTGAAGAAGGACACCAGGCCCATGAG
CAGGAGCAAACATGGTGGCTGGCTCCTGAGAGAGCACCAGGCGGGCTTTGAGAAGCTCCGCAGGACCC
GAGGAGAAGAGAAGGAGGCAGAGAAGGAGAAAAAGCCATGTATGTCTGGAGGCAGAAGGATGACTCTC
AGAGACGACCAACCTGCAAAGCTAGAAAAGGAGCCCAAGACGAGGCCAGAAGAGAACAAGCCAGAGCG
GCCCAGCGGTCGGAAGCCACGGCCCATGGGCATCATTGCCGCCAATGTGGAAAAGCATTATGAGACTG
GCCGGGTCATTGGGGATGGGAACTTTGCTGTCGTGAAGGAGTGCAGACACCGCGAGACCAGGCAGGCC
TATGCGATGAAGATCATTGACAAGTCCAGACTCAAGGGCAAGGAGGACATGGTGGACAGTGAGATCTT
GATCATCCAGAGCCTCTCTCACCCCAACATCGTGAAATTGCATGAAGTCTACGAAACAGACATGGAAA
TCTACCTGATCCTGGAGTACGTGCAGGGAGGAGACCTTTTTGACGCCATCATAGAAAGTGTGAAGTTC
CCGGAGCCCGATGCTGCCCTCATGATCATGGACTTATGCAAAGCCCTCGTCCACATGCACGACAAGAG
CATTGTCCACCGGGACCTCAAGCCGGAAAACCTTTTGGTTCAGCGAAATGAGGACAAATCTACTACCT
TGAAATTGGCTGATTTTGGACTTGCAAAGCATGTGGTGAGACCTATATTTACTGTGTGTGGGACCCCA
ACTTACGTAGCTCCCGAAATTCTTTCTGAGAAAGGTTATGGACTGGAGCTCGAGGGC

NOV64f, 267254000 Protein Sequence	SEQ ID NO: 966	540 aa	MW at 61179.2kD

TGSTMGKEPLTLKSIQVAVEELYPNKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCS
EVAGCKAAMRHQGKIPEELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKT
SGEIIRCEKCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKG
DSHRSSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREVKKDTRPMS
RSKHGGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAKLEKEPKTRPEENKPER
PSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQAYAMKIIDKSRLKGKEDMVDSEIL
IIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLFDAIIESVKFPEPDAALMIMDLCKALVHMHDKS
IVHRDLKPENLLVQRNEDKSTTLKLADFGLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLELEG

NOV64g, 267253987 DNA Sequence	SEQ ID NO: 967	793 bp
	ORF Start: at 1	ORF Stop: at 793

ACCGGATCCTATGAGACTGGCCGGGTCATTGGGGATGGGAACTTTGCTGTCGTGAAGGAGTGCAGACA
CCGCGAGACCAGGCAGGCCTATGCGATGAAGATCATTGACAAGTCCAGACTCAAGGGCAAGGAGGACA
TGGTGGACAGTGAGATCTTGATCATCCAGAGCCTCTCTCACCCCAACATCGTGAAATTGCATGAAGTC
TACGAAACAGACATGGAAATCTACCTGATCCTGGAGTACGTGCAGGGAGGAGACCTTTTTGACGCCAT
CATAGAAAGTGTGAAGTTCCCGGAGCCCGATGCTGCCCTCATGATCATGGACTTATGCAAAGCCCTCG
TCCACATGCACGACAAGAGCATTGTCCACCGGGACCTCAAGCCGGAAAACCTTTTGGTTCAGCGAAAT
GAGGACAAATCTACTACCTTGAAATTGGCTGATTTTGGACTTGCAAAGCATGTGGTGAGACCTATATT
TACTGTGTGTGGGACCCCAACTTACGTAGCTCCCGAAATTCTTTCTGAGAAAGGTTATGGACTGGAGG
TGGACATGTGGGCTGCTGGCGTGATCCTCTATATCCTGCTGTGTGGCTTTCCCCCATTCCGCAGCCCT

GAGAGGGACCAGGACGAGCTCTTTAACATCATCCAGCTGGGCCACTTTGAGTTCCTCCCCCCTTACTG
GGACAATATCTCTGATGCTGCTAAAGATCTGGTGAGCCGGTTGCTGGTGGTAGACCCCAAAAAGCGCT
ACACAGCTCATCAGGTTCTTCAGCACCCCTGGATCCTCGAGGGCC

NOV64g, 267253987 Protein Sequence	SEQ ID NO: 968	264 aa	MW at 30106.4kD

TGSYETGRVIGDGNFAVVKECRHRETRQAYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEV
YETDMEIYLILEYVQGGDLFDAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRN
EDKSTTLKLADFGLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSP
ERDQDELFNIIQLGHFEFLPPYWDNISDAAKDLVSRLLVVDPKKRYTAHQVLQHPWILEG

NOV64h, CG57109-02 DNA Sequence	SEQ ID NO: 969	2808 bp
	ORF Start: ATG at 151	ORF Stop: TAG at 2659

TTTACAGAGTCAGGCCTCACCGTGAGAGGGCTCCTGTATTAGTCCCTTTTCATGCTGCTTATAGAGAC
ATACCTGAGACTGGGCAATTTGCAGAGAAAGGTTTTCTTGGACTTACAGTTAGTTCCACGTGGCTGGG
GAAGCCTCACAATC**ATG**GCGGAAGGCAAGGAAGGGCAAGTCCCATCTTACATGGATGGCAGCAGGCAA
AGAGAGAATGAGGAAGATGCAAAAGCGGAAACCCCTGATGTAACCATCAGATCTTATGAGATTTATTC
ACTACCATGGAACAGACAGCAAGGCGTATGTGACCATTCTCTAGAATATTTAAGCTCGAGAATCTCAG
AGCGGAAGCTGCAAGGCTCCTGGCTGCCTGCCAGCCGAGGGAATCTGGAGAAACCATTCCTGGGGCCG
CGTGGCCCCGTCGTGCCCTTGTTCTGCCCTCGGAATGGCCTTCACTCAGCACATCCTGAGAACAGCCC
TCTGAAGCCCAGGGTCGTGACCGTAGTGAAGCTGGGTGGGCAGCGCCCCCGAAAGATCACTCTGCTCC
TCAACAGGCGATCAGTGCAGACGTTCGAGCAGCTCTTAGCTGACATCTCAGAAGCCTTGGGCTCTCCC
AGATGGAAGAATGACCGTGTGAGGAAACTGTTTAACCTCAAGGGCAGGGAAATCAGGAGCGTCTCTGA
TTTCTTCAGGGAAGGGGATGCTTTCATAGCTATGGGCAAAGAACCACTGACACTGAAGAGCATTCAGG
TGGCTGTAGAAGAACTGTACCCCAACAAAGCCCGGGCCCTGACACTGGCCCAGCACAGCCGTGCCCCT
TCTCCAAGGCTGAGGAGCAGGCTGTTTAGCAAGGCTCTGAAAGGAGACCACCGCTGTGGGGAGACCGA
GACCCCCAAGAGCTGCAGCGAAGTTGCAGGATGCAAGGCAGCCATGAGGCACCAGGGGAAGATCCCCG
AGGAGCTTTCACTAGATGACAGAGCGAGGACCCAGAAGAAGTGGGGGAGGGGGAAATGGGAGCCAGAA
CCCAGTAGCAAGCCCCCCAGGGAAGCCACTCTGGAAGAGAGGCACGCAAGGGGAGAGAAGCATCTTGG
GGTGGAGATTGAAAAGACCTCGGGTGAAATTATCAGATGCGAGAAGTGCAAGAGAGAGAGGGAGCTCC
AGCAGAGCCTGGAGCGTGAGAGGCTTTCTCTGGGGACCAGTGAGCTGGATATGGGGAAGGGCCCAATG
TATGATGTGGAGAAGCTGGTGAGGACCAGAAGCTGCAGGAGGTCTCCCGAGGCAAATCCTGCAAGTGG
GGAGGAAGGGTGGAAGGGTGACAGCCACAGGAGCAGCCCCAGGAATCCCACTCAAGAGCTGAGGAGAC
CCAGCAAGAGCATGGACAAGAAAGAGGACAGAGGCCCAGAGGATCAAGAAAGCCATGCTCAGGGAGCA
GCCAAGGCCAAGAAGGACCTTGTGGAAGTTCTTCCTGTCACAGAGGAGGGGCTGAGGGAGGTGAAGAA
GGACACCAGGCCCATGAGCAGGAGCAAACATGGTGGCTGGCTCCTGAGAGAGCACCAGGCGGGCTTTG
AGAAGCTCCGCAGGACCCGAGGAGAAGAGAAGGAGGCAGAGAAGGAGAAAAAGCCATGTATGTCTGGA
GGCAGAAGGATGACTCTCAGAGATGACCAACCTGCAAAGCTAGAAAAGGAGCCCAAGACGAGGCCAGA
AGAGAACAAGCCAGAGCGGCCCAGCGGTCGGAAGCCACGGCCCATGGGCATCATTGCCGCCAATGTGG
AAAAGCATTATGAGACTGGCCGGGTCATTGGGGATGGGAACTTTGCTGTCGTGAAGGAGTGCAGACAC
CGCGAGACCAGGCAGGCCTATGCGATGAAGATCATTGACAAGTCCAGACTCAAGGGCAAGGAGGACAT
GGTGGACAGTGAGATCTTGATCATCCAGAGCCTCTCTCACCCCAACATCGTGAAATTGCATGAAGTCT
ACGAAACAGACATGGAAATCTACCTGATCCTGGAGTACGTGCAGGGAGGAGACCTTTTTGACGCCATC
ATAGAAAGTGTGAAGTTCCCGGAGCCCGATGCTGCCCTCATGATCATGGACTTATGCAAAGCCCTCGT
CCACATGCACGACAAGAGCATTGTCCACCGGGACCTCAAGCCGGAAAACCTTTTGGTTCAGCGAAATG
AGGACAAATCTACTACCTTGAAATTGGCTGATTTTGGACTTGCAAAGCATGTGGTGAGACCTATATTT
ACTGTGTGTGGGACCCCAACTTACGTAGCTCCCGAAATTCTTTCTGAGAAAGGTTATGGACTGGAGGT
GGACATGTGGGCTGCTGGCGTGATCCTCTATATCCTGCTGTGTGGCTTTCCCCCATTCCGCAGCCCTG
AGAGGGACCAGGACGAGCTCTTTAACATCATCCAGCTGGGCCACTTTGAGTTCCTCCCCCCTTACTGG
GACAATATCTCTGATACAGCTGCTAAAGATCTGGTGAGCCGGTTGCTGGTGGTAGACCCCAAAAAGCG
CTACACAGCTCATCAGGTTCTTCAGCACCCCTGGATCGAAACAGCTGGCAAGACCAATACAGTGAAAC
GACAGAAGCAGGTGTCCCCCAGCAGCGAGGGTCACTTCCGGAGCCAGCACAAGAGGGTTGTGGAGCAG
GTATCA**TAG**TCACCACCTTGGGAATCTGTCCAGCCCCCAGTTCTGCTCAAGGACAGAGAAAAGGATAG
AAGTTTGAGAGAAAAACAATGAAAGAGGCTTCTTCACATAATTGGTGAATCAGAGGGAGAGACACTGA
GTATATTTTAAAGCATATTA

NOV64h, CG57109-02 Protein Sequence	SEQ ID NO: 970	836 aa	MW at 95152.6kD

MAEGKEGQVPSYMDGSRQRENEEDAKAETPDVTIRSYEIYSLPWNRQQGVCDHSLEYLSSRISERKLQ

GSWLPASRGNLEKPFLGPRGPVVPLFCPRNGLHSAHPENSPLKPRVVTVVKLGGQRPRKITLLLNRRS
VQTFEQLLADISEALGSPRWKNDRVRKLFNLKGREIRSVSDFFREGDAFIAMGKEPLTLKSIQVAVEE
LYPNKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAGCKAAMRHQGKIPEELSL
DDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEIIRCEKCKRERELQQSLE
RERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKGDSHRSSPRNPTQELRRPSKSM
DKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREVKKDTRPMSRSKHGGWLLREHQAGFEKLRR
TRGEEKEAEKEKKPCMSGGRRMTLRDDQPAKLEKEPKTRPEENKPERPSGRKPRPMGIIAANVEKHYE
TGRVIGDGNFAVVKECRHRETRQAYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEVYETDM
EIYLILEYVQGGDLFDAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKST
TLKLADFGLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSPERDQD
ELFNIIQLGHFEFLPPYWDNISDTAAKDLVSRLLVVDPKKRYTAHQVLQHPWIETAGKTNTVKRQKQV
SPSSEGHFRSQHKRVVEQVS

NOV64i, CG57109-03 DNA Sequence	SEQ ID NO: 971	3016 bp
	ORF Start: ATG at 52	ORF Stop: TAG at 2617

TTTACAGAGTCAGGCCTCACCGTGAGAGGGCTCCTGTATTAGTCCCTTTTC**ATG**CTGCTTATAGAGAC
ATACCTGAGACTGGGCAATTTGCAGAGAAAGGTTTTCTTGGACTTACAGTTAGTTCCACGTGGCTGGG
GAAGCCTCACAATCATGGCGGAAGGCAAGGAAGGGCAAGTCCCATCTTACATGGATGGCAGCAGGCAA
AGAGAGAATGAGGAAGATGCAAAAGCGGAAACCCCTGATGTAACCATCAGATCTTATGAGATTTATTC
ACTACCATGGAACAGACAGCAAGGCGTATGTGACCATTCTCTAGAATATTTAAGCTCGAGAATCTCAG
AGCGGAAGCTGCAAGGCTCCTGGCTGCCTGCCAGCCGAGGGAATCTGGAGAAACCATTCCTGGGGCCG
CGTGGCCCCGTCGTGCCCTTGTTCTGCCCTCGGAATGGCCTTCACTCAGCACATCCTGAGAACAGCCC
TCTGAAGCCCAGGGTCGTGACCGTAGTGAAGCTGGGTGGGCAGCGCCCCCGAAAGATCACTCTGCTCC
TCAACAGGCGATCAGTGCAGACGTTCGAGCAGCTCTTAGCTGACATCTCAGAAGCCTTGGGCTCTCCC
AGATGGAAGAATGACCGTGTGAGGAAACTGTTTAACCTCAAGGGCAGGGAAATCAGGAGCGTCTCTGA
TTTCTTCAGGGAAGGGGATGCTTTCATAGCTATGGGCAAAGAACCACTGACACTGAAGAGCATTCAGG
TGGCTGTAGAAGAACTGTACCCCAACAAAGCCCGGGCCCTGACACTGGCCCAGCACAGCCGTGCCCCT
TCTCCAAGGCTGAGGAGCAGGCTGTTTAGCAAGGCTCTGAAAGGAGACCACCGCTGTGGGGAGACCGA
GACCCCCAAGAGCTGCAGCGAAGTTGCAGGATGCAAGGCAGCTATGAGGCACCAGGGGAAGATCCCCG
AGGAGCTTTCACTAGATGACAGAGCGAGGACCCAGAAGAAGTGGGGGAGGGGGAAATGGGAGCCAGAA
CCCAGTAGCAAGCCCCCCAGGGAAGCCACTCTGGAAGAGAGGCACGCAAGGGGAGAGAAGCATCTTGG
GGTGGAGATTGAAAAGACCTCGGGTGAAATTATCAGATGCGAGAAGTGCAAGAGAGAGAGGGAGCTTC
AGCAGAGCCTGGAGCGTGAGAGGCTTTCTCTGGGGACCAGTGAGCTGGATATGGGGAAGGGCCCAATG
TATGATGTGGAGAAGCTGGTGAGGACCAGAAGCTGCAGGAGGTCTCCCGAGGCAAATCCTGCAAGTGG
GGAGGAAGGGTGGAAGGGTGACAGCCACAGGAGCAGCCCCAGGAATCCCACTCAAGAGCTGAGGAGAC
CCAGCAAGAGCATGGACAAGAAAGAGGACAGAGGCCCAGAGGATCAAGAAAGCCATGCTCAGGGAGCA
GCCAAGGCCAAGAAGGACCTTGTGGAAGTTCTTCCTGTCACAGAGGAGGGGCTGAGGGAGGTGAAGAA
GGACACCAGGCCCATGAGCAGGAGCAAACATGGTGGCTGGCTCCTGAGAGAGCACCAGGCGGGCTTTG
AGAAGCTCCGCAGGACCCGAGGAGAAGAGAAGGAGGCAGAGAAGGAGAAAAAGCCATGTATGTCTGGA
GGCAGAAGGATGACTCTCAGAGATGACCAACCTGCAAAGCTAGAAAAGGAGCCCAAGACGAGGCCAGA
AGAGAACAAGCCAGAGCGGCCCAGCGGTCGGAAGCCACGGCCCATGGGCATCATTGCCGCCAATGTGG
AAAAGCATTATGAGACTGGCCGGGTCATTGGGGATGGGAACTTTGCTGTCGTGAAGGAGTGCAGACAC
CGCGAGACCAGGCAGGCCTATGCGATGAAGATCATTGACAAGTCCAGACTCAAGGGCAAGGAGGACAT
GGTGGACAGTGAGATCTTGCATGAAGTCTACGAAACAGACATGGAAATCTACCTGATCCTGGAGTACG
TGCAGGGAGGAGACCTTTTTGACGCCATCATAGAAAGTGTGAAGTTCCCGGAGCCCGATGCTGCCCTC
ATGATCATGGACTTATGCAAAGCCCTCGTCCACATGCACGACAAGAGCATTGTCCACCGGGACCTCAA
GCCGGAAAACCTTTTGGTTCAGCGAAATGAGGACAAATCTACTACCTTGAAATTGGCTGATTTTGGAC
TTGCAAAGCATGTGGTGAGACCTATATTTACTGTGTGTGGGACCCCAACTTACGTAGCTCCCGAAATT
CTTTCTGAGAAAGGTTATGGACTGGAGGTGGACATGTGGGCTGCTGGCGTGATCCTCTATATCCTGCT
GTGTGGCTTTCCCCCATTCCGCAGCCCTGAGAGGGACCAGGACGAGCTCTTTAACATCATCCAGCTGG
GCCACTTTGAGTTCCTCCCCCCTTACTGGGACAATATCTCTGATGCTGCTAAAGATCTGGTGAGCCGG
TTGCTGGTGGTAGACCCCAAAAAGCGCTACACAGCTCATCAGGTTCTTCAGCACCCCTGGATCGAAAC
AGCTGGCAAGACCAATACAGTGAAACGACAGAAGCAGGTGTCCCCCAGCAGCGAGGGTCACTTCCGGA
GCCAGCACAAGAGGGTTGTGGAGCAGGTATCA**TAG**TCACCACCTTGGGAATCTGTCCAGCCCCCAGTT
CTGCTCAAGGACAGAGAAAAGGATAGAAGTTTGAGAGAAAAACAATGAAAGAGGCTTCTTCACATAAT
TGGTGAATCAGAGGGAGAGACACTGAGTATATTTTAAAGCATATTAAAAAAATTAAGTCAATGTTAAA
TGTCACAACATATTTTTAGATTTGTATATTTAAAGCCTTTAATACATTTTTGGGGGGTAAGCATTGTC
ATCAGTGAGGAATTTTGGTAATAATGATGTGTTTTGCTTCCCCTTTGTAACCAAGTTTATTCTGTACT
ACAGGAGTGGTGCTTACCAGGGTCTAAACTCCCCCTGTGAGATTAATAAGGTGCACTGTGGTCTTTCT

GTGTTAATAAAATGTGCTCTGAAT

NOV64i, CG57109-03 Protein Sequence	SEQ ID NO: 972	855 aa	MW at 97447.3kD

MLLIETYLRLGNLQRKVFLDLQLVPRGWGSLTIMAEGKEGQVPSYMDGSRQRENEEDAKAETPDVTIR
SYEIYSLPWNRQQGVCDHSLEYLSSRISERKLQGSWLPASRGNLEKPFLGPRGPVVPLFCPRNGLHSA
HPENSPLKPRVVTVVKLGGQRPRKITLLLNRRSVQTFEQLLADISEALGSPRWKNDRVRKLFNLKGRE
IRSVSDFFREGDAFIAMGKEPLTLKSIQVAVEELYPNKARALTLAQHSRAPSPRLRSRLFSKALKGDH
RCGETETPKSCSEVAGCKAAMRHQGKIPEELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHAR
GEKHLGVEIEKTSGEIIRCEKCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPE
ANPASGEEGWKGDSHRSSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEG
LREVKKDTRPMSRSKHGGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAKLEKE
PKTRPEENKPERPSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQAYAMKIIDKSRL
KGKEDMVDSEILHEVYETDMEIYLILEYVQGGDLFDAIIESVKFPEPDAALMIMDLCKALVHMHDKSI
VHRDLKPENLLVQRNEDKSTTLKLADFGLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGV
ILYILLCGFPPFRSPERDQDELFNIIQLGHFEFLPPYWDNISDAAKDLVSRLLVVDPKKRYTAHQVLQ
HPWIETAGKTNTVKRQKQVSPSSEGHFRSQHKRVVEQVS

NOV64j, CG57109-04 DNA Sequence	SEQ ID NO: 973	2433 bp
	ORF Start: ATG at 52	ORF Stop: TAG at 2284

TTTACAGAGTCAGGCCTCACCGTGAGAGGGCTCCTGTATTAGTCCCTTTTC**ATG**CTGCTTATAGAGAC
ATACCTGAGACTGGGCAATTTGCAGAGAAAGGTTTTCTTGGACTTACAGTTAGTTCCACGTGGCTGGG
GAAGCCTCACAATCATGGCGGAAGGCAAGGAAGGGCAAGTCCCATCTTACATGGATGGCAGCAGGCAA
AGAGAGAATGAGGAAGATGCAAAAGCGGAAACCCCTGATGTAACCATCAGATCTTATGAGATTTATTC
ACTACCATGGAACAGACAGCAAGGCGTATGTGACCATTCTCTAGAATATTTAAGCTCGAGAATCTCAG
AGCGGAAGCTGCAAGGCTCCTGGCTGCCTGCCAGCCGAGGGAATCTGGAGAAACCATTCCTGGGGCCG
CGTGGCCCCGTCGTGCCCTTGTTCTGCCCTCGGAATGGCCTTCACTCAGCACATCCTGAGAACAGCCC
TCTGAAGCCCAGGGTCGTGACCGTAGTGAAGCTGGGTGGGCAGCGCCCCCGAAAGATCACTCTGCTCC
TCAACAGGCGATCAGTGCAGACGTTCGAGCAGCTCTTAGCTGACATCTCAGAAGCCTTGGGCTCTCCC
AGATGGAAGAATGACCGTGTGAGGAAACTGTTTAACCTCAAGGGCAGGGAAATCAGGAGCGTCTCTGA
TTTCTTCAGGGAAGGGGATGCTTTCATAGCTATGGGCAAAGAACCACTGACACTGAAGAGCATTCAGG
TGGCTGTAGAAGAACTGTACCCCAACAAAGCCCGGGCCCTGACACTGGCCCAGCACAGCCGTGCCCCT
TCTCCAAGGCTGAGGAGCAGGCTGTTTAGCAAGGCTCTGAAAGGAGACCACCGCTGTGGGGAGACCGA
GACCCCCAAGAGCTGCAGCGAAGTTGCAGGATGCAAGGCAGCCATGAGGCACCAGGGGAAGATCCCCG
AGGAGCTTTCACTAGATGACAGAGCGAGGACCCAGAAGAAGTGGGGGAGGGGGAAATGGGAGCCAGAA
CCCAGTAGCAAGCCCCCCAGGGAAGCCACTCTGGAAGAGAGGCACGCAAGGGGAGAGAAGCATCTTGG
GGTGGAGATTGAAAAGACCTCGGGTGAAATTATCAGATGCGAGAAGTGCAAGAGAGAGAGGGAGCTCC
AGCAGAGCCTGGAGCGTGAGAGGCTTTCTCTGGGGACCAGTGAGCTGGATATGGGGAAGGGCCCAATG
TATGATGTGGAGAAAAAGCCATGTATGTCTGGAGGCAGAAGGATGACTCTCAGAGATGACCAACCTGC
AAAGCTAGAAAAGGAGCCCAAGACGAGGCCAGAAGAGAACAAGCCAGAGCGGCCCAGCGGTCGGAAGC
CACGGCCCATGGGCATCATTGCCGCCAATGTGGAAAAGCATTATGAGACTGGCCGGGTCATTGGGGAT
GGGAACTTTGCTGTCGTGAAGGAGTGCAGACACCGCGAGACCAGGCAGGCCTATGCGATGAAGATCAT
TGACAAGTCCAGACTCAAGGGCAAGGAGGACATGGTGGACAGTGAGATCTTGATCATCCAGAGCCTCT
CTCACCCCAACATCGTGAAATTGCATGAAGTTTACGAAACAGACATGGAAATCTACCTGATCCTGGAG
TACGTGCAGGGAGGAGACCTTTTTGACGCCATCATAGAAAGTGTGAAGTTCCCGGAGCCCGATGCTGC
CCTCATGATCATGGACTTATGCAAAGCCCTCGTCCACATGCACGACAAGAGCATTGTCCACCGGGACC
TCAAGCCGGAAAACCTTTTGGTTCAGCGAAATGAGGACAAATCTACTACCTTGAAATTGGCTGATTTT
GGACTTGCAAAGCATGTGGTGAGACCTATATTTACTGTGTGTGGGACCCCAACTTACGTAGCTCCCGA
AATTCTTTCTGAGAAAGGTTATGGACTGGAGGTGGACATGTGGGCTGCTGGCGTGATCCTCTATATCC
TGCTGTGTGGCTTTCCCCCATTCCGCAGCCCTGAGAGGGACCAGGACGAGCTCTTTAACATCATCCAG
CTGGGCCACTTTGAGTTCCTCCCCCCTTACTGGGACAATATCTCTGATACAGCTGCTAAAGATCTGGT
GAGCCGGTTGCTGGTGGTAGACCCCAAAAAGCGCTACACAGCTCATCAGGTTCTTCAGCACCCCTGGA
TCGAAACAGCTGGCAAGACCAATACAGTGAAACGACAGAAGCAGGTGTCCCCCAGCAGCGAGGGTCAC
TTCCGGAGCCAGCACAAGAGGGTTGTGGAGCAGGTATCA**TAG**TCACCACCTTGGGAATCTGTCCAGCC
CCCAGTTCTGCTCAAGGACAGAGAAAAGGATAGAAGTTTGAGAGAAAAACAATGAAAGAGGCTTCTTC
ACATAATTGGTGAATCAGAGGGAGAGACACTGAGTATATTTTAAAGCATATTA

NOV64j, CG57109-04 Protein Sequence	SEQ ID NO: 974	744 aa	MW at 84729.4kD

MLLIETYLRLGNLQRKVFLDLQLVPRGWGSLTIMAEGKEGQVPSYMDGSRQRENEEDAKAETPDVTIR
SYEIYSLPWNRQQGVCDHSLEYLSSRISERKLQGSWLPASRGNLEKPFLGPRGPVVPLFCPRNGLHSA
HPENSPLKPRVVTVVKLGGQRPRKITLLLNRRSVQTFEQLLADISEALGSPRWKNDRVRKLFNLKGRE
IRSVSDFFREGDAFIAMGKEPLTLKSIQVAVEELYPNKARALTLAQHSRAPSPRLRSRLFSKALKGDH
RCGETETPKSCSEVAGCKAAMRHQGKIPEELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHAR
GEKHLGVEIEKTSGEIIRCEKCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKKPCMSGGRRMTL
RDDQPAKLEKEPKTRPEENKPERPSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQA
YAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLFDAIIESVKF
PEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADFGLAKHVVRPIFTVCGTP
TYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSPERDQDELFNIIQLGHFEFLPPYWDNISDT
AAKDLVSRLLVVDPKKRYTAHQVLQHPWIETAGKTNTVKRQKQVSPSSEGHFRSQHKRVVEQVS

NOV64k, CG57109-06 DNA Sequence	SEQ ID NO: 975	2720 bp
	ORF Start: ATG at 149	ORF Stop: TGA at 1826

GGACACTGACATGGACTGAAGGAGTAGAAAAGAAGCCTTGGGCTCTCCCAGATGGAAGAATGACCGTG
TGAGGAAACTGTTTAACCTCAAGGGCAGGGAAATCAGGAGCGTCTCTGATTTCTTCAGGGAAGGGGAT
GCTTTCATAGCT**ATG**GGCAAAGAACCACTGACACTGAAGAGCATTCAGGTGGCTGTAGAAGAACTGTA
CCCCAACAAAGCCCGGGCCCTGACACTGGCCCAGCACAGCCGTGCCCCTTCTCCAAGGCTGAGGAGCA
GGCTGTTTAGCAAGGCTCTGAAAGGAGACCACCGCTGTGGGGAGACCGAGACCCCCAAGAGCTGCAGC
GAAGTTGCAGGATGCAAGGCAGCTATGAGGCACCAGGGGAAGATCCCCGAGGAGCTTTCACTAGATGA
CAGAGCGAGGACCCAGAAGAAGTGGGGGAGGGGGAAATGGGAGCCAGAACCCAGTAGCAAGCCCCCCA
GGGAAGCCACTCTGGAAGAGAGGCACGCAAGGGGAGAGAAGCATCTTGGGGTGGAGATTGAAAAGACC
TCGGGTGAAATTATCAGATGCGAGAAGTGCAAGAGAGAGAGGGAGCTTCAGCAGAGCCTGGAGCGTGA
GAGGCTTTCTCTGGGGACCAGTGAGCTGGATATGGGGAAGGGCCCAATGTATGATGTGGAGAAGCTGG
TGAGGACCAGAAGCTGCAGGAGGTCTCCCGAGGCAAATCCTGCAAGTGGGGAGGAAGGGTGGAAGGGT
GACAGCCACAGGAGCAGCCCCAGGAATCCCACTCAAGAGCTGAGGAGACCCAGCAAGAGCATGGACAA
GAAAGAGGACAGAGGCCCAGAGGATCAAGAAAGCCATGCTCAGGGAGCAGCCAAGGCCAAGAAGGACC
TTGTGGAAGTTCTTCCTGTCACAGAGGAGGGGCTGAGGGAGGTGAAGAAGGACACCAGGCCCATGAGC
AGGAGCAAACATGGTGGCTGGCTCCTGAGAGAGCACCAGGCGGGCTTTGAGAAGCTCCGCAGGACCCG
AGGAGAAGAGAAGGAGGCAGAGAAGGAGAAAAAGCCATGTATGTCTGGAGGCAGAAGGATGACTCTCA
GAGATGACCAACCTGCAAAGCTAGAAAAGGAGCCCAAGACGAGGCCAGAAGAGAACAAGCCAGAGCGG
CCCAGCGGTCGGAAGCCACGGCCCATGGGCATCATTGCCGCCAATGTGGAAAAGCATTATGAGACTGG
CCGGGTCATTGGGGATGGGAACTTTGCTGTCGTGAAGGAGTGCAGACACCGCGAGACCAGGCAGGCCT
ATGCGATGAAGATCATTGACAAGTCCAGACTCAAGGGCAAGGAGGACATGGTGGACAGTGAGATCTTG
CATGAAGTCTACGAAACAGACATGGAAATCTACCTGATCCTGGAGTACGTGCAGGGAGGAGACCTTTT
TGACGCCATCATAGAAAGTGTGAAGTTCCCGGAGCCCGATGCTGCCCTCATGATCATGGACTTATGCA
AAGCCCTCGTCCACATGCACGACAAGAGCATTGTCCACCGGGACCTCAAGCCGGAAAACCTTTTGGTT
CAGCGAAATGAGGACAAATCTACTACCTTGAAATTGGCTGATTTTGGACTTGCAAAGCATGTGGTGAG
ACCTATATTTACTGTGTGTGGGACCCCAACTTACGTAGCTCCCGAAATTCTTTCTGAGAAAGGTAAGT
GTTACACATCGATCTGTGGGACTCTAGTTCCCTTACTAACAAATGTTTCATTTGTCATATTTACTAGT
TTTCAATATGGAATAAATCTCAGAGAACTGACACTTAGGCTTGGATTTGGACTTCAA**TGA**AAATATTT
GAAGTAGGCTTGACCAAAGCATGAGAGCTTTCTCCTCATTAGGGCTGCCCTTGTTACAGTCAATGGAT
CAGTGTGTGTGCATGTGTGTGTGTGTGTGTGTTTATTGTGTTTAGGCAGGACAGTGAGATGAAAGA
TGATGGAGAATTGGGTGGGGAACTCAAGCAAGCAAGGCTACTTGACCCAAGGCTATCTCTAATAGGAG
AGAATTGAAGCAGTCCTTATGGTACTTGGTTTAAAAATTTCTTCACCAACCTTGCATTTAAAGGAAAA
GGATCCCATTTTCCTCCATGAACTCTATGAATATTTATTACCTACTTGTATATTATGCAAGGATCCAA
TGAGCTGTTTTAGTGACAAACTTTCTAAAACATTTAAAAAGGAAATAATAGTTATGATATGGCTCTAA
ATATATGAATGACTATTTGACTACTGTGGCACTCCAGGAGAAACCAATTTACCCACCATTGGTAAGAT
GGGAAGACTCTCATTGGGTTGCAGGGTTGGTGACAGGGAGAAGGGATGGACGGATAGTTTCCCAGCAG
CAGAAGCATCAGGATAATTAAGATGAGGAGATGGCCAGGGATGGTGGCTCATGCCTGTAATCCCAGCT
CTTTGGGAGGCTGAGGCAGGTGGATCACCTGAGGTCAGGCGTTTGAAACCATCCTGGCCAACATGGTG
AAACCCTGTCTGTACTAAAACTACAAAAAAATTAGCTGGGCGTGGTGGCACATGCCTGTAATCTGAGC
TACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCGGGGAGGTGGAGTTTGCAGTGAGCCAAGATCG
TGCCATTGCACTCCAGCCTGGGCAACAAGAGTGAAACTTCATCTCAAAAAAAAAAAAAAAAAAAAAC

NOV64k, CG57109-06 Protein Sequence	SEQ ID NO: 976	559 aa	MW at 63436.9kD

MGKEPLTLKSIQVAVEELYPNKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAG
CKAAMRHQGKIPEELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEI
IRCEKCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKGDSHR
SSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREVKKDTRPMSRSKH
GGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAKLEKEPKTRPEENKPERPSGR
KPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQAYAMKIIDKSRLKGKEDMVDSEILHEVY
ETDMEIYLILEYVQGGDLFDAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNE
DKSTTLKLADFGLAKHVVRPIFTVCGTPTYVAPEILSEKGKCYTSICGTLVPLLTNVSFVIFTSFQYG
INLRELTLRLGFGLQ

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 64B.

Table 64B. Comparison of the NOV64 protein sequences.

```
NOV64a  ------------------------------------------------------------
NOV64b  ------------------------------------------------------------
NOV64c  ------------------------------------------------------------
NOV64d  ------------------------------------------------------------
NOV64e  ------------------------------------------------------------
NOV64f  ------------------------------------------------------------
NOV64g  ------------------------------------------------------------
NOV64h  ---------------------------------MAEGKEGQVPSYMDGSRQRENEEDAKA
NOV64i  MLLIETYLRLGNLQRKVFLDLQLVPRGWGSLTIMAEGKEGQVPSYMDGSRQRENEEDAKA
NOV64j  MLLIETYLRLGNLQRKVFLDLQLVPRGWGSLTIMAEGKEGQVPSYMDGSRQRENEEDAKA
NOV64k  ------------------------------------------------------------

NOV64a  ------------------------------------------------------------
NOV64b  ------------------------------------------------------------
NOV64c  ------------------------------------------------------------
NOV64d  ------------------------------------------------------------
NOV64e  ------------------------------------------------------------
NOV64f  ------------------------------------------------------------
NOV64g  ------------------------------------------------------------
NOV64h  ETPDVTIRSYEIYSLPWNRQQGVCDHSLEYLSSRISERKLQGSWLPASRGNLEKPFLGPR
NOV64i  ETPDVTIRSYEIYSLPWNRQQGVCDHSLEYLSSRISERKLQGSWLPASRGNLEKPFLGPR
NOV64j  ETPDVTIRSYEIYSLPWNRQQGVCDHSLEYLSSRISERKLQGSWLPASRGNLEKPFLGPR
NOV64k  ------------------------------------------------------------

NOV64a  ------------------------------------------------------------
NOV64b  ------------------------------------------------------------
NOV64c  ------------------------------------------------------------
NOV64d  ------------------------------------------------------------
NOV64e  ------------------------------------------------------------
NOV64f  ------------------------------------------------------------
NOV64g  ------------------------------------------------------------
NOV64h  GPVVPLFCPRNGLHSAHPENSPLKPRVVTVVKLGGQRPRKITLLLNRRSVQTFEQLLADI
NOV64i  GPVVPLFCPRNGLHSAHPENSPLKPRVVTVVKLGGQRPRKITLLLNRRSVQTFEQLLADI
NOV64j  GPVVPLFCPRNGLHSAHPENSPLKPRVVTVVKLGGQRPRKITLLLNRRSVQTFEQLLADI
NOV64k  ------------------------------------------------------------

NOV64a  ---------------------------------------MGKEPLTLKSIQVAVEELYP
```

```
NOV64b  ------------------------------------------------------------
NOV64c  ------------------------------------------------------------
NOV64d  ----------------------------------------MGKEPLTLKSIQVAVEELYP
NOV64e  ----------------------------------TGSTMGKEPLTLKSIQVAVEELYP
NOV64f  ----------------------------------TGSTMGKEPLTLKSIQVAVEELYP
NOV64g  ------------------------------------------------------------
NOV64h  SEALGSPRWKNDRVRKLFNLKGREIRSVSDFFREGDAFIAMGKEPLTLKSIQVAVEELYP
NOV64i  SEALGSPRWKNDRVRKLFNLKGREIRSVSDFFREGDAFIAMGKEPLTLKSIQVAVEELYP
NOV64j  SEALGSPRWKNDRVRKLFNLKGREIRSVSDFFREGDAFIAMGKEPLTLKSIQVAVEELYP
NOV64k  ----------------------------------------MGKEPLTLKSIQVAVEELYP

NOV64a  NKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAGCKAAMRHQGKIP
NOV64b  ------------------------------------------------------------
NOV64c  ------------------------------------------------------------
NOV64d  NKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAGCKAAMRHQGKIP
NOV64e  NKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAGCKAAMRHQGKIP
NOV64f  NKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAGCKAAMRHQGKIP
NOV64g  ------------------------------------------------------------
NOV64h  NKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAGCKAAMRHQGKIP
NOV64i  NKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAGCKAAMRHQGKIP
NOV64j  NKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAGCKAAMRHQGKIP
NOV64k  NKARALTLAQHSRAPSPRLRSRLFSKALKGDHRCGETETPKSCSEVAGCKAAMRHQGKIP

NOV64a  EELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEIIRCE
NOV64b  ------------------------------------------------------------
NOV64c  ------------------------------------------------------------
NOV64d  EELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEIIRCE
NOV64e  EELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEIIRCE
NOV64f  EELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEIIRCE
NOV64g  ------------------------------------------------------------
NOV64h  EELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEIIRCE
NOV64i  EELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEIIRCE
NOV64j  EELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEIIRCE
NOV64k  EELSLDDRARTQKKWGRGKWEPEPSSKPPREATLEERHARGEKHLGVEIEKTSGEIIRCE

NOV64a  KCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKG
NOV64b  ------------------------------------------------------------
NOV64c  ------------------------------------------------------------
NOV64d  KCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKG
NOV64e  KCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKG
NOV64f  KCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKG
NOV64g  ------------------------------------------------------------
NOV64h  KCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKG
NOV64i  KCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKG
NOV64j  KCKRERELQQSLERERLSLGTSELDMGKGPMYDVE------------------------
NOV64k  KCKRERELQQSLERERLSLGTSELDMGKGPMYDVEKLVRTRSCRRSPEANPASGEEGWKG

NOV64a  DSHRSSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREV
NOV64b  ------------------------------------------------------------
NOV64c  ------------------------------------------------------------
NOV64d  DSHRSSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREV
NOV64e  DSHRSSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREV
NOV64f  DSHRSSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREV
NOV64g  ------------------------------------------------------------
```

```
NOV64h  DSHRSSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREV
NOV64i  DSHRSSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREV
NOV64j  ------------------------------------------------------------
NOV64k  DSHRSSPRNPTQELRRPSKSMDKKEDRGPEDQESHAQGAAKAKKDLVEVLPVTEEGLREV

NOV64a  KKDTRPMSRSKHGGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAK
NOV64b  ------------------------------------------------------------
NOV64c  ------------------------------------------------------------
NOV64d  KKDTRPMSRSKHGGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAK
NOV64e  KKDTRPMSRSKHGGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAK
NOV64f  KKDTRPMSRSKHGGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAK
NOV64g  ------------------------------------------------------------
NOV64h  KKDTRPMSRSKHGGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAK
NOV64i  KKDTRPMSRSKHGGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAK
NOV64j  ------------------------------------KKPCMSGGRRMTLRDDQPAK
NOV64k  KKDTRPMSRSKHGGWLLREHQAGFEKLRRTRGEEKEAEKEKKPCMSGGRRMTLRDDQPAK

NOV64a  LEKEPKTRPEENKPERPSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQ
NOV64b  ------------------------------TGSYETGRVIGDGNFAVVKECRHRETRQ
NOV64c  ------------------------------TGSYETGRVIGDGNFAVVKECRHRETRQ
NOV64d  LEKEPKTRPEENKPERPSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQ
NOV64e  LEKEPKTRPEENKPERPSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQ
NOV64f  LEKEPKTRPEENKPERPSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQ
NOV64g  ------------------------------TGSYETGRVIGDGNFAVVKECRHRETRQ
NOV64h  LEKEPKTRPEENKPERPSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQ
NOV64i  LEKEPKTRPEENKPERPSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQ
NOV64j  LEKEPKTRPEENKPERPSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQ
NOV64k  LEKEPKTRPEENKPERPSGRKPRPMGIIAANVEKHYETGRVIGDGNFAVVKECRHRETRQ

NOV64a  AYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLF
NOV64b  AYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLF
NOV64c  AYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLF
NOV64d  AYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLF
NOV64e  AYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLF
NOV64f  AYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLF
NOV64g  AYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLF
NOV64h  AYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLF
NOV64i  AYAMKIIDKSRLKGKEDMVDSEI-------------LHEVYETDMEIYLILEYVQGGDLF
NOV64j  AYAMKIIDKSRLKGKEDMVDSEILIIQSLSHPNIVKLHEVYETDMEIYLILEYVQGGDLF
NOV64k  AYAMKIIDKSRLKGKEDMVDSEIL-------------HEVYETDMEIYLILEYVQGGDLF

NOV64a  DAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADF
NOV64b  DAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADF
NOV64c  DAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADF
NOV64d  DAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADF
NOV64e  DAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADF
NOV64f  DAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADF
NOV64g  DAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADF
NOV64h  DAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADF
NOV64i  DAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADF
NOV64j  DAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADF
NOV64k  DAIIESVKFPEPDAALMIMDLCKALVHMHDKSIVHRDLKPENLLVQRNEDKSTTLKLADF

NOV64a  GLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSPERDQ
```

```
NOV64b  GLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSPERDQ
NOV64c  GLAKHVVRPIFTVCGTPTYVAPEILSEK------------------------------
NOV64d  GLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSPERDQ
NOV64e  GLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSPERDQ
NOV64f  GLAKHVVRPIFTVCGTPTYVAPEILSEKG--------------------------------
NOV64g  GLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSPERDQ
NOV64h  GLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSPERDQ
NOV64i  GLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSPERDQ
NOV64j  GLAKHVVRPIFTVCGTPTYVAPEILSEKGYGLEVDMWAAGVILYILLCGFPPFRSPERDQ
NOV64k  GLAKHVVRPIFTVCGTPTYVAPEILSEKGK-----------CYTSICG-----------

NOV64a  DELFNIIQLGHFEFLPPYWDNISD-AAKDLVSRLLVVDPKKRYTAHQVLQHPWIETAGKT
NOV64b  DELFNIIQLGHFEFLPPYWDNISD-AAKDLVSRLLVVDPKKRYTAHQVLQHPWILEG---
NOV64c  ---------------------AAK----DLVSRLLVVDPKKRYTAHQVLQHPWILEG---
NOV64d  DELFNIIQLGHFEFLPPYWDNISD-AAKDLVSRLLVVDPKKRYTAHQVLQHPWIETAGKT
NOV64e  DELFNIIQLGHFEFLPPYWDNISD-AAKDLVSRLLVVDPKKRYTAHQVLQHPWIETAGKT
NOV64f  ----------------------------YGLELEG----------------------
NOV64g  DELFNIIQLGHFEFLPPYWDNISD-AAKDLVSRLLVVDPKKRYTAHQVLQHPWILEG---
NOV64h  DELFNIIQLGHFEFLPPYWDNISDTAAKDLVSRLLVVDPKKRYTAHQVLQHPWIETAGKT
NOV64i  DELFNIIQLGHFEFLPPYWDNISD-AAKDLVSRLLVVDPKKRYTAHQVLQHPWIETAGKT
NOV64j  DELFNIIQLGHFEFLPPYWDNISDTAAKDLVSRLLVVDPKKRYTAHQVLQHPWIETAGKT
NOV64k  --------T-----LVPLLTNVSFVIFTSFQYGINLRELTLRLGFGLQ------------

NOV64a  NTVKRQKQVSPSSEGHFRSQHKRVVEQVS---
NOV64b  --------------------------------
NOV64c  --------------------------------
NOV64d  NTVKRQKQVSPSSEGHFRSQHKRVVEQVS---
NOV64e  NTVKRQKQVSPSSEGHFRSQHKRVVEQVSLEG
NOV64f  --------------------------------
NOV64g  --------------------------------
NOV64h  NTVKRQKQVSPSSEGHFRSQHKRVVEQVS---
NOV64i  NTVKRQKQVSPSSEGHFRSQHKRVVEQVS---
NOV64j  NTVKRQKQVSPSSEGHFRSQHKRVVEQVS---
NOV64k  --------------------------------

NOV64a  (SEQ ID NO:  956)
NOV64b  (SEQ ID NO:  958)
NOV64c  (SEQ ID NO:  960)
NOV64d  (SEQ ID NO:  962)
NOV64e  (SEQ ID NO:  964)
NOV64f  (SEQ ID NO:  966)
NOV64g  (SEQ ID NO:  968)
NOV64h  (SEQ ID NO:  970)
NOV64i  (SEQ ID NO:  972)
NOV64j  (SEQ ID NO:  974)
NOV64k  (SEQ ID NO:  976)
```

Further analysis of the NOV64a protein yielded the following properties shown in Table 64C.

Table 64C. Protein Sequence Properties NOV64a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 9;  pos.chg 2;  neg.chg 1
      H-region:  length 6;  peak value  -5.22
      PSG score:  -9.62

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -7.85
      possible cleavage site: between 34 and 35

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -3.29   Transmembrane  534 - 550
      PERIPHERAL  Likelihood =  4.72 (at 452)
      ALOM score:  -3.29  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 541
      Charge difference:  1.0   C(-2.0) - N(-3.0)
      C > N: C-terminal side will be inside

>>> Single TMS is located near the C-terminus

>>> membrane topology: type Nt  (cytoplasmic tail 1 to 533)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  9.28
      Hyd Moment(95):  7.07    G content:       1
      D/E content:     2       S/T content:     2
      Score: -6.37

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-10 motif at 52  SRL FS

NUCDISC: discrimination of nuclear localization signals
      pat4: RKPR (4) at  340
      pat4: PKKR (4) at  598
      pat7: PSGRKPR (3) at  337
      pat7: PKKRYTA (5) at  598
      bipartite: none
      content of basic residues:  17.9%
      NLS Score:  0.72

KDEL: ER retention motif in the C-terminus: none
```

```
ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: too long tail

Dileucine motif in the tail: found
      LL at 276
      LL at 483

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 89

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

---------------------------

Final Results (k = 9/23):

       30.4 %: nuclear
       26.1 %: cytoplasmic
       13.0 %: Golgi
        8.7 %: mitochondrial
        8.7 %: vesicles of secretory system
        8.7 %: endoplasmic reticulum
        4.3 %: peroxisomal

>> prediction for CG57109-01 is nuc (k=23)
```

A search of the NOV64a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 64D.

Table 64D. Geneseq Results for NOV64a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV64a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE24133	Human kinase (PKIN)-4 protein - Homo sapiens, 835 aa. [WO200233099-A2, 25-APR-2002]	1..648 188..835	648/648 (100%) 648/648 (100%)	0.0
ABB82471	Human serine/threonine protein kinase - Homo sapiens, 835 aa. [WO200270678-A2, 12-SEP-2002]	1..648 188..835	648/648 (100%) 648/648 (100%)	0.0
AAO15419	Novel human kinase protein 2 - Homo sapiens, 817 aa. [WO200242438-A2, 30-MAY-2002]	1..648 170..817	647/648 (99%) 648/648 (99%)	0.0
ABB82474	Human serine/threonine protein kinase fragment - Homo sapiens, 751 aa. [WO200270678-A2, 12-SEP-2002]	1..484 188..671	484/484 (100%) 484/484 (100%)	0.0
AAY42696	Rat serine-threonine protein kinase PK80 sequence - Rattus norvegicus, 733 aa. [WO9950395-A1, 07-OCT-1999]	1..648 98..733	495/649 (76%) 535/649 (82%)	0.0

In a BLAST search of public sequence databases, the NOV64a protein was found to have homology to the proteins shown in the BLASTP data in Table 64E.

Table 64E. Public BLASTP Results for NOV64a				
Protein Accession Number	**Protein/Organism/Length**	**NOV64a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9C098	Hypothetical protein KIAA1765 - Homo sapiens (Human), 608 aa (fragment).	41..648 1..608	608/608 (100%) 608/608 (100%)	0.0
Q8BWQ5	Hypothetical eukaryotic protein kinase containing protein - Mus musculus (Mouse), 619 aa.	1..632 1..619	467/633 (73%) 510/633 (79%)	0.0
O15075	Serine/threonine-protein kinase DCAMKL1 (EC 2.7.1.-) (Doublecortin- like and CAM kinase-like 1) - Homo sapiens (Human), 740 aa.	348..641 382..674	155/294 (52%) 212/294 (71%)	3e-87
Q8BQN2	Double cortin and calcium/calmodulin-dependent protein kinase-like 1 - Mus musculus (Mouse), 452 aa.	348..641 75..367	154/294 (52%) 212/294 (71%)	1e-86
Q8CCN4	Double cortin and calcium/calmodulin-dependent protein kinase-like 1 - Mus musculus (Mouse), 449 aa.	348..641 91..383	154/294 (52%) 212/294 (71%)	1e-86

PFam analysis indicates that the NOV64a protein contains the domains shown in the Table 64F.

Table 64F. Domain Analysis of NOV64a			
Pfam Domain	**NOV64a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
pkinase	356..613	109/301 (36%) 220/301 (73%)	8.1e-102

Example 65.

The NOV65 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 65A.

Table 65A. NOV65 Sequence Analysis

NOV65a, CG57399-04 DNA Sequence	SEQ ID NO: 977	947 bp
	ORF Start: ATG at 7	ORF Stop: TGA at 937

GACCTG**ATG**AAGAATGACACGAGGATACACTTTCAGGAAGACTGGAAGATAATAACCCTGTTTATAGG
CGGCAATGACCTCTGTGATTTCTGCAATGATCTGGTCCACTATTCTCCCCAGAACTTCACAGACAACA
TTGGAAAGGCCCTGGACATCCTCCATGCTGAGGTTCCTCGGGCATTTGTGAACCTGGTGACGGTGCTT
GAGATCGTCAACCTGAGGGAGCTGTACCAGGAGAAAAAAGTCTACTGCCCAAGGATGATCCTCAGGTC
TCTGTGTCCCTGTGTCCTGAAGTTTGATGATAACTCAACAGAACTTGCTACCCTCATCGAATTCAACA
AGAAGTTTCAGGAGAAGACCCACCAACTGATTGAGAGTGGGCGATATGACACAAGGGAAGATTTTACT
GTGGTTGTGCGGCCGTTCTTTGAAAACGTGGACATGCCAAAGACCTCGGAAGGATTGCCTGACAACTC
TTTCTTCGCTCCTGACTGTTTCCACTTCAGCAGCAAGTCTCACTCCCGAGCAGCCAGTGCTCTCTGGA
ACAATATGCTGGAGCCTGTTGGCCAGAAGACGACTCGTCATAAGTTTGAAAACAAGATCAATATCACA
TGTCCGAACCAGGTCCAGCCGTTTCTGAGGACCTACAAGAACAGCATGCAGGGTCATGGGACCTGGCT
GCCATGCAGGGACAGAGCCCCTTCTGCCTTGCACCCTACCTCAGTGCATGCCCTGAGACCTGCAGACA
TCCAAGTTGTGGCTGCTCTGGGGGATTCTCTGACCGCTGGCAATGGAATTGGCTCCAAACCAGACGAC
CTCCCCGATGTCACCACACAGTATCGGGGACTGTCATACAGAGAAAGTAAACCAGGGTTCTTATCAGA
CTCCTGGGTCAGCAAATCCAACAGGAAATGCACCAGAAAAGCACCAAATCCC**TGA**ATCTACAC

NOV65a, CG57399-04 Protein Sequence	SEQ ID NO: 978	310 aa	MW at 35268.7kD

MKNDTRIHFQEDWKIITLFIGGNDLCDFCNDLVHYSPQNFTDNIGKALDILHAEVPRAFVNLVTVLEI
VNLRELYQEKKVYCPRMILRSLCPCVLKFDDNSTELATLIEFNKKFQEKTHQLIESGRYDTREDFTVV
VRPFFENVDMPKTSEGLPDNSFFAPDCFHFSSKSHSRAASALWNNMLEPVGQKTTRHKFENKINITCP
NQVQPFLRTYKNSMQGHGTWLPCRDRAPSALHPTSVHALRPADIQVVAALGDSLTAGNGIGSKPDDLP
DVTTQYRGLSYRESKPGFLSDSWVSKSNRKCTRKAPNP

NOV65b, CG57399-01 DNA Sequence	SEQ ID NO: 979	4268 bp
	ORF Start: ATG at 1	ORF Stop: TAG at 4258

ATGACCTGGGACACAGCTCTCTGGACCTCAGTTTTTCTGATTGGGCTCCTTCCTACCCTTGGTTTCGC
TAATTGCATCCTCCAGACTTCTGGTAAAATGTGTACTTTAAGAGGTAGATACCCCCAGCCCCCACAAC
CACCTCTCTGCTTGTCTCCCCTAGTCCACCAGCTCCGACCAGCAGACATCAAAGTGGTGGCCGCCCTG
GGTAATGATGAAACCTTCCAGGAAAGTGGTGCAGGGCAGCTAAGTGAGCCTGACCCCAGGCAGTGGTC
CTGGCCACAGGCCTGCTTGCCTGGGGTAAAAAAGGAAATGCAAGATGTGGTAGGTGAGAGAACGCCGA
GCCGTCGCCGCAGCCTCCGCCGCCGAGAAGCCCTTGTTCCCGCTGCTGGGAAGGAGAGTCTGTGCCGA
CAAGATATTTTCATTTCCTTGTTGGAAATTATCAAGCATTTTCCTCCCTCCCCTCAGGACATCAACCT
GGAGAAAGACTGGAAGCTGGTCACACTCTTCATTGGGGTCAACGACTTGTGTCATTACTGTCCACTTG
TTCAGGGCCCCGTTATAGACCTGGGTGGGATGGATACCCTCCACTCCCTGCAGCTCCCAAGGGCTTTC
GTCAACGTGGTGGAGGTCATGGAGCTGGCTAGCCTGTACCAGGGCCAAGGCGGGAAATGTGCCATGCT
GGCAGCTCAGGAAGCCTGGAACAGCCTCCTGGCCTCCAGCAGGTACAGTGAGCAGGAGTCCTTCACCG
TGGTTTTCCAGCCTTTCTTCTATGAGACCACCCCATCTGACCCCCGACTCCAGGATTCTACCACGCTG
GCCTGGCATCTCTGGAATAGGATGATGGAGCCAGCAGGAGAGAAAGATGAGCCATTGAGTGTAAAACA
CGGGAGGCCAATGAAGTGTCCCTCTCAGGAGAGCCCCTATCTGTTCAGCTACAGAAACAGCAACTACC
TGACCAGACTGCAGAAACCCCAAGACAAGCTTGTAAGAGAAGGAGCGGAAATCAGATGTCCTGACAAA
GACCCCTCCGATACGGTTCCCACCTCAGTTCATAGGCTGAAGCCGGCTGACATCAACGTAATTGGAGC
CCTGGGTGACTCTCTCACGGCAGGCAATGGGGCCGGGTCCACACCTGGGAACGTCTTGGACGTCTTGA
CTCAGTACCGAGGCCTGTCCTGGAGCGTCGGCGGAGATGAGAACATCGGCACCGTTACCACCCTGGCA
GACATCCTCCGGGAATTCAACCCTTCCCTGAAGGGCTTCTCTGTTGGCACTGGGAAAGAAACCAGTCC
TAATGCCTTCTTAAACCAGGCTGTGGCAGGAGGCCGAGCTGAGCAGGCCAGGAGGCTGGTGGACCTGA
TGAAGAATGACACGAGGATACACTTTCAGGAAGACTGGAAGATAATAACCCTGTTTATAGGCGGCAAT
GACCTCTGTGATTTCTGCAATGATCTGGTACACTATTCTCCCCAGAACTTCACAGACAACATTGGAAA
GGCCCTGGACATCCTCCATGCTGAGTCTCAGGTTCCTCGGGCATTTGTGAACCTGGTGACGGTGCTTG
AGATCGTCAACCTGAGGGAGCTGTACCAGGAGAAAAAAGTCTACTGCCCAAGGATGATCCTCAGGTCA
CTGTGTCCCTGTGTCCTGAAGTTTGATGATAACTCAACAGAACTTGCTACCCTCATCGAATTCAACAA
GAAGTTTCAGGAGAAGACCCACCAACTGATTGAGAGTGGGCGATATGACACAAGGGAAGATTTTACTG
TGGTTGTGCAGCCGTTCTTTGAAAACGTGGACATGCCAAAGACCCAGGAAGGATTGCCTGACAACTCT
TTCTTCGCTCCTGACTGTTTCCACTTCAGCAGCAAGTCTCACTCCCGAGCAGCCAGTGCTCTCTGGAA

CAATATGCTGGAGCCTGTTGGCCAGAAGACGACTCGTCATAAGTTTGAAAACAAGATCAATATCACAT
GTCCGTCACAGGTCCAGCCGTTTCTGAGGACCTACAAGAACAGCATGCAGGGTCATGGGACCTGGCTG
CCATGCAGGGACAGAGCCCCTTCTGCCTTGCACCCTACCTCAGTGCATGCCCTGAGACCTGCAGACAT
CCAAGTTGTGGCTGCTCTGGGGGATTCTCTGACCGCTGGCAATGGAATTGGCTCCAAACCAGACGACC
TCCCCGATGTCACCACACAGTATCGGGGACTGTCATACAGTGCAGGAGGGGACGGCTCCCTGGAGAAT
GTGACCACCTTACCTAGTTCTATCCTTCGGGAGTTTAACAGAAACCTCACAGGCTACGCCGTGGGCAC
GGGTGATGCCAATGACACGAATGCATTCCTCAATCAAGCTGTTCCCGGAGCAAAGGCTAGGGATCTTA
TGAGCCAAGTCCAAACTCTGATGCAGAAGATGAAAGATGATCATAGAGTAAATTTCCATGAAGACTGG
AAGGTCATCACAGTGCTGATCGGAGGCAGCGATTTATGTGACTACTGCACAGATTCGAATCTGTATTC
TGCAGCCAACTTTGTTCACCATCTCCGCAATGCCTTGGACGTCCTGCATAGAGAGGTGCCCAGAGTCC
TGGTCAACCTCGTGGACTTCCTGAACCCCACTATCATGCGGCAGGTGTTCCTGGGAAACCCAGACAAG
TGCCCAGTGCAGCAGGCCAGCGTTTTGTGTAACTGCGTTCTGACCCTGCGGGAGAACTCCCAAGAGCT
AGCCAGGCTGGAGGCCTTCAGCCGAGCCTACCAGAGCAGCATGCGCGAGCTGGTGGGGTCAGGCCGCT
ATGACACGCAGGAGGACTTCTCTGTGGTGCTGCAGCCCTTCTTCCAGAACATCCAGCTCCCTGTCCTG
CAGGATGGGCTCCCAGATACGTCCTTCTTTGCCCCAGACTGCATCCACCCAAATCAGAAATTCCACTC
CCAGCTGGCCAGAGCCCTTTGGACCAATATGCTTGAACCACTTGGAAGCAAAACAGAGACCCTGGACC
TGAGAGCAGAGATGCCCATCACCTGTCCCACTCAGAATGAGCCCTTCCTGAGAACCCCTCGGAATAGT
AACTACACGTACCCCATCAAGCCAGCCATTGAGAACTGGGGCAGTGACTTCCTGTGTACAGAGTGGAA
GGCTTCCAATAGTGTTCCAACCTCTGTCCACCAGCTCCGACCAGCAGACATCAAAGTGGTGGCCGCCC
TGGGTGACTCTCTGACTACAGCAGTGGGAGCTCGACCAAACAACTCCAGTGACCTACCCACATCTTGG
AGGGGACTCTCTTGGAGCATTGGAGGGGATGGGAACTTGGAGACTCACACCACACTGCCCAGTATTCT

GAAGAAGTTCAACCCTTACCTCCTTGGCTTCTCTACCAGCACCTGGGAGGGGACAGCAGGACTAAATG
TGGCAGCGGAAGGGGCCAGAGCTAGGAGGGACATGCCAGCCCAGGCCTGGGACCTGGTAGAGCGAATG
AAAAACAGCCCCATACACTTTCAGGAAGACTGGAAGATAATAACCCTGTTTATAGGCGGCAATGACCT
CTGTGATTTCTGCAATGATCTGGTAGGTGAATATGTTCAGCACATCCAACAGGCCCTGGACATCCTCT
CTGAGGAGCTCCCAAGGGCTTTCGTCAACGTGGTGGAGGTCATGGAGCTGGCTAGCCTGTACCAGGGC
CAAGGCGGGAAATGTGCCATGCTGGCAGCTCAGAACAACTGCACTTGCCTCAGACACTCGCAAAGCTC
CCTGGAGAAGCAAGAACTGAAGAAAGTGAACTGGAACCTCCAGCATGGCATCTCCAGTTTCTCCTACT
GGCACCAATACACACAGCGTGAGGACTTTGCGGTTGTGGTGCAGCCTTTCTTCCAAAACACACTCACC
CCACTGAACAGAGGGGACACTGACCTCACCTTCTTCTCCGAGGACTGTTTTCACTTCTCAGACCGCGG
GCATGCCGAGATGGCCATCGCACTCTGGAACAACATGCTGGAACCAGTGGGCCGCAAGACTACCTCCA
ACAACTTCACCCACAGCCGAGCCAAACTCAAGTGCCCCTCTCCTGTGAGTCCTTACCTCTACACCCTG
CGGAACAGCCGATTGCTCCCAGACCAGGCTGAAGAAGCCCCCGAGGTGCTCTACTGGGCTGTCCCAGT
GGCAGCGGGAGTCGGCCTTGTGGTGGGCATCATCGGGACAGTGGTCTGGAGGTGCAGGAGAGGTGGCC
GGAGGGAAGATCCTCCAATGAGCCTGCGCACTGTGGCCCTCT**TAG**GCCCGGGG

NOV65b, CG57399-01 Protein Sequence	SEQ ID NO: 980	1419 aa	MW at 158435.1kD

MTWDTALWTSVFLIGLLPTLGFANCILQTSGKMCTLRGRYPQPPQPPLCLSPLVHQLRPADIKVVAAL
GNDETFQESGAGQLSEPDPRQWSWPQACLPGVKKEMQDVVGERTPSRRRSLRRREALVPAAGKESLCR
QDIFISLLEIIKHFPPSPQDINLEKDWKLVTLFIGVNDLCHYCPLVQGPVIDLGGMDTLHSLQLPRAF
VNVVEVMELASLYQGQGGKCAMLAAQEAWNSLLASSRYSEQESFTVVFQPFFYETTPSDPRLQDSTTL
AWHLWNRMMEPAGEKDEPLSVKHGRPMKCPSQESPYLFSYRNSNYLTRLQKPQDKLVREGAEIRCPDK
DPSDTVPTSVHRLKPADINVIGALGDSLTAGNGAGSTPGNVLDVLTQYRGLSWSVGGDENIGTVTTLA
DILREFNPSLKGFSVGTGKETSPNAFLNQAVAGGRAEQARRLVDLMKNDTRIHFQEDWKIITLFIGGN
DLCDFCNDLVHYSPQNFTDNIGKALDILHAESQVPRAFVNLVTVLEIVNLRELYQEKKVYCPRMILRS
LCPCVLKFDDNSTELATLIEFNKKFQEKTHQLIESGRYDTREDFTVVVQPFFENVDMPKTQEGLPDNS
FFAPDCFHFSSKSHSRAASALWNNMLEPVGQKTTRHKFENKINITCPSQVQPFLRTYKNSMQGHGTWL
PCRDRAPSALHPTSVHALRPADIQVVAALGDSLTAGNGIGSKPDDLPDVTTQYRGLSYSAGGDGSLEN
VTTLPSSILREFNRNLTGYAVGTGDANDTNAFLNQAVPGAKARDLMSQVQTLMQKMKDDHRVNFHEDW
KVITVLIGGSDLCDYCTDSNLYSAANFVHHLRNALDVLHREVPRVLVNLVDFLNPTIMRQVFLGNPDK
CPVQQASVLCNCVLTLRENSQELARLEAFSRAYQSSMRELVGSGRYDTQEDFSVVLQPFFQNIQLPVL
QDGLPDTSFFAPDCIHPNQKFHSQLARALWTNMLEPLGSKTETLDLRAEMPITCPTQNEPFLRTPRNS
NYTYPIKPAIENWGSDFLCTEWKASNSVPTSVHQLRPADIKVVAALGDSLTTAVGARPNNSSDLPTSW
RGLSWSIGGDGNLETHTTLPSILKKFNPYLLGFSTSTWEGTAGLNVAAEGARARRDMPAQAWDLVERM
KNSPIHFQEDWKIITLFIGGNDLCDFCNDLVGEYVQHIQQALDILSEELPRAFVNVVEVMELASLYQG
QGGKCAMLAAQNNCTCLRHSQSSLEKQELKKVNWNLQHGISSFSYWHQYTQREDFAVVVQPFFQNTLT
PLNRGDTDLTFFSEDCFHFSDRGHAEMAIALWNNMLEPVGRKTTSNNFTHSRAKLKCPSPVSPYLYTL
RNSRLLPDQAEEAPEVLYWAVPVAAGVGLVVGIIGTVVWRCRRGGRREDPPMSLRTVAL

NOV65c, CG57399-02 DNA Sequence	SEQ ID NO: 981	1624 bp
	ORF Start: ATG at 311	ORF Stop: TGA at 1241

GCCGGCTGACATCAATGTAATTGGAGCCCTGGGTGACTCTCTCACGGCAGGCAATGGGGCCGGGTCCA
CACCTGGGAACGTCTTGGACGTCTTGACTCAGTACCGAGGCCTGTCCTGGAGCGTCGGCGGAGATGAG
AACATCGGCACCGTTACCACCCTGGCGAACATCCTCCGGGAATTCAACCCTTCCCTGAAGGGCTTCTC
TGTTGGCACTGGGAAAGAAACCAGTCCTAATGCCTTCTTAAACCAGGCTGTGGCAGGAGGCCGAGCTG
AGGATCTACCTGTCCAGGCCAGGAGGCTGGTGGACCTG**ATG**AAGAATGACACGAGGATACACTTTCAG
GAAGACTGGAAGATAATAACCCTGTTTATAGGCGGCAATGACCTCTGTGATTTCTGCAATGATCTGGT
CCACTATTCTCCCCAGAACTTCACAGACAACATTGGAAAGGCCCTGGACATCCTCCATGCTGAGGTTC
CTCGGGCATTTGTGAACCTGGTGACGGTGCTTGAGATCGTCAACCTGAGGGAGCTGTACCAGGAGAAA
AAAGTCTACTGCCCAAGGATGATCCTCAGGTCTCTGTGTCCCTGTGTCCTGAAGTTTGATGATAACTC
AACAGAACTTGCTACCCTCATCGAATTCAACAAGAAGTTTCAGGAGAAGACCCACCAACTGATTGAGA
GTGGGCGATATGACACAAGGGAAGATTTTACTGTGGTTGTGCAGCCGTTCTTTGAAAACGTGGACATG
CCAAAGACCTCGGAAGGATTGCCTGACAACTCTTTCTTCGCTCCTGACTGTTTCCACTTCAGCAGCAA
GTCTCACTCCCGAGCAGCCAGTGCTCTCTGGAACAATATGCTGGAGCCTGTTGGCCAGAAGACGACTC
GTCATAAGTTTGAAAACAAGATCAATATCACATGTCCGAACCAGGTCCAGCCGTTTCTGAGGACCTAC
AAGAACAGCATGCAGGGTCATGGGACCTGGCTGCCATGCAGGGACAGAGCCCCTTCTGCCTTGCACCC
TACCTCAGTGCATGCCCTGAGACCTGCAGACATCCAAGTTGTGGCTGCTCTGGGGGATTCTCTGACCG
CTGGCAATGGAATTGGCTCCAAACCAGACGACCTCCCCGATGTCACCACACAGTATCGGGGACTGTCA
TACAGAGAAAGTAAACCAGGGTTCTTATCAGACTCCTGGGTCAGCAAATCCAACAGGAAATGCACCAG
AAAAGCACCAAATCCC**TGA**ATCTTCACCTCCCCGCTTGCATGTATACGTGTACACGTGGTGTTCCTAC
GTCTCTGTTTACTGTCTTTATGTGTTTATTCATGTTGTCTTGTAGTCACACAGCTGCCTTTACATATA
TGTACACATCTGCACAGAAAACCTCTGAAACCCATCGCACACTTCGAGAGGCCATAACCAAGACACAA
TCACAATCAGCCATGTCTTGAAAGATTAGCAATTCGACAAGAGGAAAGGGTGAGAAAGGGCATCCCGA
ACACGGAAGTGGAGAAGCTCAGGGTGTGTCAGGCGAGCGGTTGCGTGTAGATATTCTCAAGTTTCTTT
CTCTCCTAATAAAGTTCTCATTCCTGTAGGCTTCAAAGTAAGTGGCGAGTAGCTCAGAAT

NOV65c, CG57399-02 Protein Sequence	SEQ ID NO: 982	310 aa	MW at 35240.6kD

MKNDTRIHFQEDWKIITLFIGGNDLCDFCNDLVHYSPQNFTDNIGKALDILHAEVPRAFVNLVTVLEI
VNLRELYQEKKVYCPRMILRSLCPCVLKFDDNSTELATLIEFNKKFQEKTHQLIESGRYDTREDFTVV
VQPFFENVDMPKTSEGLPDNSFFAPDCFHFSSKSHSRAASALWNNMLEPVGQKTTRHKFENKINITCP
NQVQPFLRTYKNSMQGHGTWLPCRDRAPSALHPTSVHALRPADIQVVAALGDSLTAGNGIGSKPDDLP
DVTTQYRGLSYRESKPGFLSDSWVSKSNRKCTRKAPNP

NOV65d, CG57399-03 DNA Sequence	SEQ ID NO: 983	4425 bp
	ORF Start: ATG at 16	ORF Stop: TAG at 4285

CTGGAGCATTCTGGC**ATG**GGGCTGCGGCCAGGCATTTTCCTCCTGGAGCTGCTGCTGCTTCTGGGGCA
AGGTACCCCTCAGATCCATACCTCTCCTAGAAAGAGTACATTGGAAGGGCAGCTATGGCCAGAGACAG
TTCACTCTCTGAAGCCTTCTGATATTAAATTTGTGGCAGCCATTGGCAATCTGGAAATTGTGCCAGAC
CCAGGGACGGGCGATCTGGAGAAGCAAGACGAAAGGCCACAGCAGGTGTGCATGGGAGTGATGACAGT
CCTTTCAGACATCATCAGATATTTCAGTCCTTCTGTTCCAATGCCTGTGTGCCACACTGGAAAGAGAG
TCATACCCCACGATGGTGCTGAGGACTTGTGGATTCAGGCTCAAGAACTGGTGAGAAACATGAAAGAG
AACCAACTTGACTTTCAATTTGACTGGAAGCTCATCAATGTGTTCTTCAGTAATGCAAGCCAGTGTTA
CCTGTGCCCCTCTGCTCAACAGAATGGGCTTGCGGCGGGCGGCGTGGATGAGCTGATGGGGGTGCTGG
ACTACCTGCAGCAGGAGGTGCCCAGAGCATTTGTAAACCTGGTGGACCTCTCTGAGGTTGCAGAGGTC
TCTCGTCAGTATCACGGCACTTGGCTCAGCCCTGCACCAGAGCCCTGTAATTGCTCAGAGGAGACCAC
CCGGCTGGCCAAGGTGGTGATGCAGTGGTCTTATCAGGAAGCCTGGAACAGCCTCCTGGCCTCCAGCA
GGTACAGTGAGCAGGAGTCCTTCACCGTGGTTTTCCAGCCTTTCTTCTATGAGACCACCCCATCTGAC
CCCCGACTCCAGGATTCTACCACGCTGGCCTGGCATCTCTGGAATAGGATGATGGAGCCAGCAGGAGA
GAAAGATGAGCCATTGAGTGTAAAACACGGGAGGCCAATGAAGTGTCCCTCTCAGGAGAGCCCCTATC
TGTTCAGCTACAGAAACAGCAACTACCTGACCAGACTGCAGAAACCCCAAGACAAGCTTGAGGTAAGA
GAAGGAGCGGAAATCAGATGTCCTGACAAAGACCCCTCCGATACGGTTCCCACCTCAGTTCATAGGCT
GAAGCCGGCTGACATCAACGTAATTGGAGCCCTGGGTGACTCTCTCACGGCAGGCAATGGGGCCGGGT
CCACACCTGGGAACGTCTTGGACGTCTTGACTCAGTACCGAGGCCTGTCCTGGAGCGTCGGCGGAGAT
GAGAACATCGGCACCGTTACCACCCTGGCGGACATCCTCCGGGAATTCAACCCTTCCCTGAAGGGCTT
CTCTGTTGGCACTGGGAAAGAAACCAGTCCTAATGCCTTCTTAAACCAGGCTGTGGCAGGAGGCCGAG
CTGAGCAGGCCAGGAGGCTGGTGGACCTGATGAAGAATGACACGAGGATACACTTTCAGGAAGACTGG

AAGATAATAACCCTGTTTATAGGCGGCAATGACCTCTGTGATTTCTGCAATGATCTGGTACACTATTC
TCCCCAGAACTTCACAGACAACATTGGAAAGGCCCTGGACATCCTCCATGCTGAGGTTCCTCGGGCAT
TTGTGAACCTGGTGACGGTGCTTGAGATCGTCAACCTGAGGGAGCTGTACCAGGAGAAAAAAGTCTAC
TGCCCAAGGATGATCCTCAGGTCACTGTGTCCCTGTGTCCTGAAGTTTGATGATAACTCAACAGAACT
TGCTACCCTCATCGAATTCAACAAGAAGTTTCAGGAGAAGACCCACCAACTGATTGAGAGTGGGCGAT
ATGACACAAGGGAAGATTTTACTGTGGTTGTGCAGCCGTTCTTTGAAAACGTGGACATGCCAAAGACC
CAGGAAGGATTGCCTGACAACTCTTTCTTCGCTCCTGACTGTTTCCACTTCAGCAGCAAGTCTCACTC
CCGAGCAGCCAGTGCTCTCTGGAACAATATGCTGGAGCCTGTTGGCCAGAAGACGACTCGTCATAAGT
TTGAAAACAAGATCAATATCACATGTCCGAACCAGGTAGAGTGGCCGTTTCTGAGGACCTACAAGAAC
AGCATGCAGGGTCATGGGACCTGGCTGCCATGCAGGGACAGAGCCCCTTCTGCCTTGCACCCTACCTC
AGTGCATGCCCTGAGACCTGCAGACATCCAAGTTGTGGCTGCTCTGGGGGATTCTCTGACCGCTGGCA
ATGGAATTGGCTCCAAACCAGACGACCTCCCCGATGTCACCACACAGTATCGGGGACTGTCATACAGT
GCAGGAGGGGACGGCTCCCTGGAGAATGTGACCACCTTACCTGATATCCTTCGGGAGTTTAACAGAAA
CCTCACAGGCTACGCCGTGGGCACGGGTGATGCCAATGACACGAATGCATTCCTCAATCAAGCTGTTC
CCGGAGCAAAGGCTAGGGATCTTATGAGCCAAGTCCAAACTCTGATGCAGAAGATGAAAGATGATCAT
AGAGTAAATTTCCATGAAGACTGGAAGGTCATCACAGTGCTGATCGGAGGCAGCGATTTATGTGACTA
CTGCACAGATTCGAATCTGTATTCTGCAGCCAACTTTGTTCACCATCTCCGCAATGCCTTGGACGTCC
TGCATAGAGAGGTGCCCAGAGTCCTGGTCAACCTCGTGGACTTCCTGAACCCCACTATCATGCGGCAG
GTGTTCCTGGGAAACCCAGACAAGTGCCCAGTGCAGCAGGCCAGCGTTTTGTGTAACTGCGTTCTGAC
CCTGCGGGAGAACTCCCAAGAGCTAGCCAGGCTGGAGGCCTTCAGCCGAGCCTACCAGAGCAGCATGC
GCGAGCTGGTGGGGTCAGGCCGCTATGACACGCAGGAGGACTTCTCTGTGGTGCTGCAGCCCTTCTTC
CAGAACATCCAGCTCCCTGTCCTGCAGGATGGGCTCCCAGATACGTCCTTCTTTGCCCCAGACTGCAT
CCACCCAAATCAGAAATTCCACTCCCAGCTGGCCAGAGCCCTTTGGACCAATATGCTTGAACCACTTG
GAAGCAAAACAGAGACCCTGGACCTGAGAGCAGAGATGCCCATCACCTGTCCCACTCAGAATGAGCCC
TTCCTGAGAACCCCTCGGAATAGTAACTACACGTACCCCATCAAGCCAGCCATTGAGAACTGGGGCAG
TGACTTCCTGTGTACAGAGTGGAAGGCTTCCAATAGTGTTCCAACCTCTGTCCACCAGCTCCGACCAG
CAGACATCAAAGTGGTGGCCGCCCTGGGTGACTCTCTGACTGTGGCAGTGGGAGCTCGACCAAACAAC
TCCAGTGACCTACCCACATCTTGGAGGGGACTCTCTTGGAGCATTGGAGGGGATGGGAACTTGGAGAC

TCACACCACACTGCCCGACATTCTGAAGAAGTTCAACCCTTACCTCCTTGGCTTCTCTACCAGCACCT
GGGAGGGGACAGCAGGACTAAATGTGGCAGCGGAAGGGGCCAGAGCTAGGGACATGCCAGCCCAGGCC
TGGGACCTGGTAGAGCGAATGAAAAACAGCCCCCAGGACATCAACCTGGAGAAAGACTGGAAGCTGGT
CACACTCTTCATTGGGGTCAACGACTTGTGTCATTACTGTGAGAATCCGGTAGGCGAATATGTTCAGC
ACATCCAACAGGCCCTGGACATCCTCTCTGAGGAGCTCCCAAGGGCTTTCGTCAACGTGGTGGAGGTC
ATGGAGCTGGCTAGCCTGTACCAGGGCCAAGGCGGGAAATGTGCCATGCTGGCAGCTCAGAACAACTG
CACTTGCCTCAGACACTCGCAAAGCTCCCTGGAGAAGCAAGAACTGAAGAAAGTGAACTGGAACCTCC
AGCATGGCATCTCCAGTTTCTCCTACTGGCACCAATACACACAGCGTGAGGACTTTGCGGTTGTGGTG
CAGCCTTTCTTCCAAAACACACTCACCCCACTGAACAGAGGGGACACTGACCTCACCTTCTTCTCCGA
GGACTGTTTTCACTTCTCAGACCGCGGGCATGCCGAGATGGCCATCGCACTCTGGAACAACATGCTGG
AACCAGTGGGCCGCAAGACTACCTCCAACAACTTCACCCACAGCCGAGCCAAACTCAAGTGCCCCTCT
CCTGAGAGCCCTTACCTCTACACCCTGCGGAACAGCCGATTGCTCCCAGACCAGGCTGAAGAAGCCCC
CGAGGTGCTCTACTGGGCTGTCCCAGTGGCAGCGGGAGTCGGCCTTGTGGTGGGCATCATCGGGACAG
TGGTCTGGAGGTGCAGGAGAGGTGGCCGGAGGGAAGATCCTCCAATGAGCCTGCGCACTGTGGCCCTC
TAGGCCCGGGGGTGGGTCCTCACCCTAAACTCCCTATAGCCACTCTCTTCACCGCCCTCTGCCCCAGC
CACTCCCGGCCACCAGGACATGCTTCAATGCCTGGTGCCATAGGAAGCCCAGGGGACAGTCACAACTT
CTTGG

NOV65d, CG57399-03 Protein Sequence	SEQ ID NO: 984	1423 aa	MW at 159352.7kD

MGLRPGIFLLELLLLLGQGTPQIHTSPRKSTLEGQLWPETVHSLKPSDIKFVAAIGNLEIVPDPGTGD
LEKQDERPQQVCMGVMTVLSDIIRYFSPSVPMPVCHTGKRVIPHDGAEDLWIQAQELVRNMKENQLDF
QFDWKLINVFFSNASQCYLCPSAQQNGLAAGGVDELMGVLDYLQQEVPRAFVNLVDLSEVAEVSRQYH
GTWLSPAPEPCNCSEETTRLAKVVMQWSYQEAWNSLLASSRYSEQESFTVVFQPFFYETTPSDPRLQD
STTLAWHLWNRMMEPAGEKDEPLSVKHGRPMKCPSQESPYLFSYRNSNYLTRLQKPQDKLEVREGAEI
RCPDKDPSDTVPTSVHRLKPADINVIGALGDSLTAGNGAGSTPGNVLDVLTQYRGLSWSVGGDENIGT
VTTLADILREFNPSLKGFSVGTGKETSPNAFLNQAVAGGRAEQARRLVDLMKNDTRIHFQEDWKIITL
FIGGNDLCDFCNDLVHYSPQNFTDNIGKALDILHAEVPRAFVNLVTVLEIVNLRELYQEKKVYCPRMI
LRSLCPCVLKFDDNSTELATLIEFNKKFQEKTHQLIESGRYDTREDFTVVVQPFFENVDMPKTQEGLP
DNSFFAPDCFHFSSKSHSRAASALWNNMLEPVGQKTTRHKFENKINITCPNQVEWPFLRTYKNSMQGH
GTWLPCRDRAPSALHPTSVHALRPADIQVVAALGDSLTAGNGIGSKPDDLPDVTTQYRGLSYSAGGDG

SLENVTTLPDILREFNRNLTGYAVGTGDANDTNAFLNQAVPGAKARDLMSQVQTLMQKMKDDHRVNFH
EDWKVITVLIGGSDLCDYCTDSNLYSAANFVHHLRNALDVLHREVPRVLVNLVDFLNPTIMRQVFLGN
PDKCPVQQASVLCNCVLTLRENSQELARLEAFSRAYQSSMRELVGSGRYDTQEDFSVVLQPFFQNIQL
PVLQDGLPDTSFFAPDCIHPNQKFHSQLARALWTNMLEPLGSKTETLDLRAEMPITCPTQNEPFLRTP
RNSNYTYPIKPAIENWGSDFLCTEWKASNSVPTSVHQLRPADIKVVAALGDSLTVAVGARPNNSSDLP
TSWRGLSWSIGGDGNLETHTTLPDILKKFNPYLLGFSTSTWEGTAGLNVAAEGARARDMPAQAWDLVE
RMKNSPQDINLEKDWKLVTLFIGVNDLCHYCENPVGEYVQHIQQALDILSEELPRAFVNVVEVMELAS
LYQGQGGKCAMLAAQNNCTCLRHSQSSLEKQELKKVNWNLQHGISSFSYWHQYTQREDFAVVVQPFFQ
NTLTPLNRGDTDLTFFSEDCFHFSDRGHAEMAIALWNNMLEPVGRKTTSNNFTHSRAKLKCPSPESPY
LYTLRNSRLLPDQAEEAPEVLYWAVPVAAGVGLVVGIIGTVVWRCRRGGRREDPPMSLRTVAL

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 65B.

Table 65B. Comparison of the NOV65 protein sequences.

```
NOV65a  ------------------------------------------------------------
NOV65b  MTWDTALWTSVFLIGLLPTLGFANCILQTSGKMCTLRGRYPQPPQPPLCLSPLVHQLRPA
NOV65c  ------------------------------------------------------------
NOV65d  ----MGLRPGIFLLELLLLLGQGTPQIHTSPRKSTLEGQLWP---------ETVHSLKPS

NOV65a  ------------------------------------------------------------
NOV65b  DIKVVAALGNDETFQESGAGQLSEPDPRQWSWPQACLPGVKKEMQDVVGERTPSRRRSLR
NOV65c  ------------------------------------------------------------
NOV65d  DIKFVAAIGNLEIVPDPGTGDLEKQDER----PQQVCMGVMTVLSDIIRYFSPSVPMPVC

NOV65a  ------------------------------------------------------------
NOV65b  RR-EALVPAAGKESLCRQDIFISLLEIIKHFPPSPQDINLEKDWKLVTLFIGVNDLCHYC
NOV65c  ------------------------------------------------------------
NOV65d  HTGKRVIPHDG-----AEDLWIQAQELVRNMKEN--QLDFQFDWKLINVFFSNASQCYLC

NOV65a  ------------------------------------------------------------
NOV65b  PLVQGPVIDLGGMDTLH-----S-LQLPRAFVNVVEVMELASLYQGQGG-----------
NOV65c  ------------------------------------------------------------
NOV65d  PSAQQNGLAAGGVDELMGVLDYLQQEVPRAFVNLVDLSEVAEVSRQYHGTWLSPAPEPCN

NOV65a  ------------------------------------------------------------
NOV65b  -------K---CAMLAAQEAWNSLLASSRYSEQESFTVVFQPFFYETTPSDPRLQDSTTL
NOV65c  ------------------------------------------------------------
NOV65d  CSEETTRLAKVVMQWSYQEAWNSLLASSRYSEQESFTVVFQPFFYETTPSDPRLQDSTTL

NOV65a  ------------------------------------------------------------
NOV65b  AWHLWNRMMEPAGEKDEPLSVKHGRPMKCPSQESPYLFSYRNSNYLTRLQKPQDK-LVRE
NOV65c  ------------------------------------------------------------
NOV65d  AWHLWNRMMEPAGEKDEPLSVKHGRPMKCPSQESPYLFSYRNSNYLTRLQKPQDKLEVRE

NOV65a  ------------------------------------------------------------
NOV65b  GAEIRCPDKDPSDTVPTSVHRLKPADINVIGALGDSLTAGNGAGSTPGNVLDVLTQYRGL
NOV65c  ------------------------------------------------------------
NOV65d  GAEIRCPDKDPSDTVPTSVHRLKPADINVIGALGDSLTAGNGAGSTPGNVLDVLTQYRGL

NOV65a  ------------------------------------------------------------
NOV65b  SWSVGGDENIGTVTTLADILREFNPSLKGFSVGTGKETSPNAFLNQAVAGGRAEQARRLV
NOV65c  ------------------------------------------------------------
```

```
NOV65d  SWSVGGDENIGTVTTLADILREFNPSLKGFSVGTGKETSPNAFLNQAVAGGRAEQARRLV

NOV65a  --MKNDTRIHFQEDWKIITLFIGGNDLCDFCNDLVHYSPQNFTDNIGKALDILHAE--VP
NOV65b  DLMKNDTRIHFQEDWKIITLFIGGNDLCDFCNDLVHYSPQNFTDNIGKALDILHAESQVP
NOV65c  --MKNDTRIHFQEDWKIITLFIGGNDLCDFCNDLVHYSPQNFTDNIGKALDILHAE--VP
NOV65d  DLMKNDTRIHFQEDWKIITLFIGGNDLCDFCNDLVHYSPQNFTDNIGKALDILHAE--VP

NOV65a  RAFVNLVTVLEIVNLRELYQEKKVYCPRMILRSLCPCVLKFDDNSTELATLIEFNKKFQE
NOV65b  RAFVNLVTVLEIVNLRELYQEKKVYCPRMILRSLCPCVLKFDDNSTELATLIEFNKKFQE
NOV65c  RAFVNLVTVLEIVNLRELYQEKKVYCPRMILRSLCPCVLKFDDNSTELATLIEFNKKFQE
NOV65d  RAFVNLVTVLEIVNLRELYQEKKVYCPRMILRSLCPCVLKFDDNSTELATLIEFNKKFQE

NOV65a  KTHQLIESGRYDTREDFTVVVRPFFENVDMPKTSEGLPDNSFFAPDCFHFSSKSHSRAAS
NOV65b  KTHQLIESGRYDTREDFTVVVQPFFENVDMPKTQEGLPDNSFFAPDCFHFSSKSHSRAAS
NOV65c  KTHQLIESGRYDTREDFTVVVQPFFENVDMPKTSEGLPDNSFFAPDCFHFSSKSHSRAAS
NOV65d  KTHQLIESGRYDTREDFTVVVQPFFENVDMPKTQEGLPDNSFFAPDCFHFSSKSHSRAAS

NOV65a  ALWNNMLEPVGQKTTRHKFENKINITCPNQVQP-FLRTYKNSMQGHGTWLPCRDRAPSAL
NOV65b  ALWNNMLEPVGQKTTRHKFENKINITCPSQVQP-FLRTYKNSMQGHGTWLPCRDRAPSAL
NOV65c  ALWNNMLEPVGQKTTRHKFENKINITCPNQVQP-FLRTYKNSMQGHGTWLPCRDRAPSAL
NOV65d  ALWNNMLEPVGQKTTRHKFENKINITCPNQVEWPFLRTYKNSMQGHGTWLPCRDRAPSAL

NOV65a  HPTSVHALRPADIQVVAALGDSLTAGNGIGSKPDDLPDVTTQYRGLSYRESKPG------
NOV65b  HPTSVHALRPADIQVVAALGDSLTAGNGIGSKPDDLPDVTTQYRGLSYSAGGDGSLENVT
NOV65c  HPTSVHALRPADIQVVAALGDSLTAGNGIGSKPDDLPDVTTQYRGLSYRESKPG------
NOV65d  HPTSVHALRPADIQVVAALGDSLTAGNGIGSKPDDLPDVTTQYRGLSYSAGGDGSLENV-

NOV65a  FLSDSWVSKSNRKCTRKAPNP----------------------------------------
NOV65b  TLPSSILREFNRNLTGYAVGTGDANDTNAFLNQAVPGAKARDLMSQVQTLMQKMKDDHRV
NOV65c  FLSDSWVSKSNRKCTRKAPNP----------------------------------------
NOV65d  TTLPDILREFNRNLTGYAVGTGDANDTNAFLNQAVPGAKARDLMSQVQTLMQKMKDDHRV

NOV65a  ----------------------------------------------------------
NOV65b  NFHEDWKVITVLIGGSDLCDYCTDSNLYSAANFVHHLRNALDVLHREVPRVLVNLVDFLN
NOV65c  ----------------------------------------------------------
NOV65d  NFHEDWKVITVLIGGSDLCDYCTDSNLYSAANFVHHLRNALDVLHREVPRVLVNLVDFLN

NOV65a  ----------------------------------------------------------
NOV65b  PTIMRQVFLGNPDKCPVQQASVLCNCVLTLRENSQELARLEAFSRAYQSSMRELVGSGRY
NOV65c  ----------------------------------------------------------
NOV65d  PTIMRQVFLGNPDKCPVQQASVLCNCVLTLRENSQELARLEAFSRAYQSSMRELVGSGRY

NOV65a  ----------------------------------------------------------
NOV65b  DTQEDFSVVLQPFFQNIQLPVLQDGLPDTSFFAPDCIHPNQKFHSQLARALWTNMLEPLG
NOV65c  ----------------------------------------------------------
NOV65d  DTQEDFSVVLQPFFQNIQLPVLQDGLPDTSFFAPDCIHPNQKFHSQLARALWTNMLEPLG

NOV65a  ----------------------------------------------------------
NOV65b  SKTETLDLRAEMPITCPTQNEPFLRTPRNSNYTYPIKPAIENWGSDFLCTEWKASNSVPT
NOV65c  ----------------------------------------------------------
NOV65d  SKTETLDLRAEMPITCPTQNEPFLRTPRNSNYTYPIKPAIENWGSDFLCTEWKASNSVPT

NOV65a  ----------------------------------------------------------
NOV65b  SVHQLRPADIKVVAALGDSLTTAVGARPNNSSDLPTSWRGLSWSIGGDGNLETHTTLPSI
```

```
NOV65c  ------------------------------------------------------------
NOV65d  SVHQLRPADIKVVAALGDSLTVAVGARPNNSSDLPTSWRGLSWSIGGDGNLETHTTLPDI

NOV65a  ------------------------------------------------------------
NOV65b  LKKFNPYLLGFSTSTWEGTAGLNVAAEGARARRDMPAQAWDLVERMKNSP--IHFQEDWK
NOV65c  ------------------------------------------------------------
NOV65d  LKKFNPYLLGFSTSTWEGTAGLNVAAEGARAR-DMPAQAWDLVERMKNSPQDINLEKDWK

NOV65a  ------------------------------------------------------------
NOV65b  IITLFIGGNDLCDFCNDLVGEYVQHIQQALDILSEELPRAFVNVVEVMELASLYQGQGGK
NOV65c  ------------------------------------------------------------
NOV65d  LVTLFIGVNDLCHYCENPVGEYVQHIQQALDILSEELPRAFVNVVEVMELASLYQGQGGK

NOV65a  ------------------------------------------------------------
NOV65b  CAMLAAQNNCTCLRHSQSSLEKQELKKVNWNLQHGISSFSYWHQYTQREDFAVVVQPFFQ
NOV65c  ------------------------------------------------------------
NOV65d  CAMLAAQNNCTCLRHSQSSLEKQELKKVNWNLQHGISSFSYWHQYTQREDFAVVVQPFFQ

NOV65a  ------------------------------------------------------------
NOV65b  NTLTPLNRGDTDLTFFSEDCFHFSDRGHAEMAIALWNNMLEPVGRKTTSNNFTHSRAKLK
NOV65c  ------------------------------------------------------------
NOV65d  NTLTPLNRGDTDLTFFSEDCFHFSDRGHAEMAIALWNNMLEPVGRKTTSNNFTHSRAKLK

NOV65a  ------------------------------------------------------------
NOV65b  CPSPVSPYLYTLRNSRLLPDQAEEAPEVLYWAVPVAAGVGLVVGIIGTVVWRCRRGGRRE
NOV65c  ------------------------------------------------------------
NOV65d  CPSPESPYLYTLRNSRLLPDQAEEAPEVLYWAVPVAAGVGLVVGIIGTVVWRCRRGGRRE

NOV65a  -----------
NOV65b  DPPMSLRTVAL
NOV65c  -----------
NOV65d  DPPMSLRTVAL

NOV65a  (SEQ ID NO:  978)
NOV65b  (SEQ ID NO:  980)
NOV65c  (SEQ ID NO:  982)
NOV65d  (SEQ ID NO:  984)
```

Further analysis of the NOV65a protein yielded the following properties shown in Table 65C.

Table 65C. Protein Sequence Properties NOV65a

SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	

```
PSG:  a new signal peptide prediction method
      N-region:  length 11;  pos.chg 2;  neg.chg 2
      H-region:  length 0;  peak value  -1.15
      PSG score:  -5.55

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1): -10.77
```

```
      possible cleavage site: between 28 and 29

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  1.43 (at 55)
      ALOM score:   1.43  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:        1       Hyd Moment(75):  3.74
      Hyd Moment(95):  6.40     G content:       0
      D/E content:      2       S/T content:     1
      Score: -6.13

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  12.3%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: KAPN

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none
```

```
checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 70.6

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       87.0 %: nuclear
        8.7 %: mitochondrial
        4.3 %: peroxisomal

>> prediction for CG57399-04 is nuc (k=23)
```

A search of the NOV65a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 65D.

Table 65D. Geneseq Results for NOV65a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV65a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABB09556	Human lipase NHL (Val 1318 variant) - Homo sapiens, 1458 aa. [WO200259328-A1, 01-AUG-2002]	1..293 491..783	284/293 (96%) 286/293 (96%)	e-170
ABB09555	Human lipase NHL (Ala 1318 variant) - Homo sapiens, 1458 aa. [WO200259328-A1, 01-AUG-2002]	1..293 491..783	284/293 (96%) 286/293 (96%)	e-170
AAE22860	Human phospholipase-like enzyme - Homo sapiens, 1216 aa. [WO200231161-A2, 18-APR-2002]	1..293 247..544	273/298 (91%) 279/298 (93%)	e-160
AAW30751	Rat phospholipase-B/lipase - Rattus rattus, 1450 aa. [JP09248190-A, 22-SEP-1997]	1..293 495..787	203/293 (69%) 236/293 (80%)	e-118
ABP53556	Human phospholipase protein SEQ ID NO:2 - Homo sapiens, 472 aa. [WO200262977-A2, 15-AUG-2002]	1..217 203..416	92/218 (42%) 130/218 (59%)	8e-44

In a BLAST search of public sequence databases, the NOV65a protein was found to have homology to the proteins shown in the BLASTP data in Table 65E.

Table 65E. Public BLASTP Results for NOV65a				
Protein Accession Number	**Protein/Organism/Length**	**NOV65a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96DP9	Hypothetical protein FLJ30866 - Homo sapiens (Human), 270 aa.	1..259 1..259	258/259 (99%) 259/259 (99%)	e-154
Q05017	Phospholipase ADRAB-B precursor (EC 3.1.-.-) - Oryctolagus cuniculus (Rabbit), 1458 aa.	1..293 491..783	238/293 (81%) 260/293 (88%)	e-140
O70320	Phospholipase B - Cavia porcellus (Guinea pig), 1463 aa.	1..291 490..780	221/291 (75%) 254/291 (86%)	e-131
O54728	Phospholipase B - Rattus norvegicus (Rat), 1450 aa.	1..293 495..787	203/293 (69%) 236/293 (80%)	e-117
Q8IUP7	Similar to phospholipase B - Homo sapiens (Human), 423 aa.	1..217 154..367	92/218 (42%) 130/218 (59%)	2e-43

PFam analysis indicates that the NOV65a protein contains the domains shown in the Table 65F.

Table 65F. Domain Analysis of NOV65a			
Pfam Domain	**NOV65a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Lipase_GDSL	1..178	39/370 (11%) 141/370 (38%)	1.3e-10

Example 66.

The NOV66 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 66A.

Table 66A. NOV66 Sequence Analysis		
NOV66a, CG57562-02 DNA Sequence	SEQ ID NO: 985	3517 bp
	ORF Start: ATG at 5	ORF Stop: TGA at 3506

AAAGATGGCGGCAGCGGCGGCGGTGGGCAACGCGGTGCCCTGCGGGGCCCGGCCTTGCGGGGTCCGGC
CTGACGGGCAGCCCAAGCCCGGGCCGCAGCCGCGCGCGCTCCTTGCCGCCGGGCCGGCGCTCATAGCG
AACGGTGACGAGCTGGTGGCTGCCGTGTGGCCGTACCGGCGGTTGGCGCTGTTGCGGCGCCTCACGGT
GCTGCCATTCGCCGGGCTGCTTTACCCGGCCTGGTTGGGTGCCGCAGCCGCTGGCTGCTGGGGCTGGG
GCAGCAGTTGGGTGCAGATCCCCGAAGCTGCGCTGCTCGTGCTTGCCACCATCTGCCTCGCGCACGCG

CTCACTGTCCTCTCGGGGCATTGGTCTGTGCACGCGCATTGCGCGCTCACCTGCACCCCGGAGTACGA
CCCCAGCAAAGCGACCTTTGTGAAGGTGGTGCCAACCCCCAACAATGGCTCCACGGAGCTCGTGGCCC
TGCACCGCAATGAGGGCGAAGACGGGCTTGAGGTGCTGTCCTTCGAATTCCAGAAGATCAAGTATTCC
TACGATGCCCTGGAGAAGAAGCAGTTTCTCCCCGTGGCCTTTCCTGTGGGAAACGCCTTCTCATACTA
TCAGAGCAACAGAGGCTTCCAGGAAGACTCAGAGATCCGAGCAGCTGAGAAGAAATTTGGGAGCAACA
AGGCCGAGATGGTGGTGCCTGACTTCTCGGAGCTTTTCAAGGAGAGAGCCACAGCCCCCTTCTTTGTA
TTTCAGGTGTTCTGTGTGGGGCTCTGGTGCCTGGATGAGTACTGGTACTACAGCGTCTTTACGCTATC
CATGCTGGTGGCGTTCGAGGCCTCGCTGGTGCAGCAGCAGATGCGGAACATGTCGGAGATCCGGAAGA
TGGGCAACAAGCCCCACATGATCCAGGTCTACCGAAGCCGCAAGTGGAGGCCCATTGCCAGTGATGAG
ATCGTACCAGGGGACATCGTCTCCATCGGCCGCTCCCCACAGGAGAACCTGGTGCCATGTGACGTGCT
TCTGCTGCGAGGCCGCTGCATCGTAGACGAGGCCATGCTCACGGGGGAGTCCGTGCCACAGATGAAGG
AGCCCATCGAAGACCTCAGCCCAGACCGGGTGCTGGACCTCCAGGCTGATTCCCGGCTGCACGTCATC
TTCGGGGGCACCAAGGTGGTGCAGCACATCCCCCCACAGAAAGCCACCACGGGCCTGAAGCCGGTTGA
CAGCGGGTGCGTGGCCTACGTCCTGCGGACCGGATTCAACACATCCCAGGGCAAGCTGCTGCGCACCA
TCCTCTTCGGGGTCAAGAGGGTGACTGCGAACAACCTGGAGACCTTCATCTTCATCCTCTTCCTCCTG
GTGTTTGCCATCGCTGCAGCTGCCTATGTATGGATTGAAGGTACCAAGGACCCCAGCCGGAACCGCTA
CAAGCTGTTTCTGGAGTGCACCCTGATCCTCACCTCGGTCGTGCCTCCTGAGCTGCCCATCGAGCTGT
CCCTGGCCGTCAACACCTCCCTCATCGCCCTGGCCAAGCTCTACATGTACTGCACAGAGCCCTTCCGG
ATCCCCTTTGCTGGCAAGGTCGAGGTGTGCTGCTTTGACAAGACGGGGACGTTGACCAGTGACAGCCT
GGTGGTGCGCGGTGTGGCCGGGCTGAGAGACGGGAAGGAGGTGACCCCAGTGTCCAGCATCCCTGTAG
AAACACACCGGGCCCTGGCCTCGTGCCACTCGCTCATGCAGCTGGACGACGGCACCCTCGTGGGTGAC
CCTCTAGAGAAGGCCATGCTGACGGCCGTGGACTGGACGCTGACCAAAGATGAGAAAGTATTCCCCCG
AAGTATTAAAACTCAGGGGCTGAAAATTCACCAGCGCTTTCATTTTGCCAGTGCCCTGAAGCGAATGT
CCGTGCTTGCCTCGTATGAGAAGCTGGGCTCCACCGACCTTTGTTACATCGCGGCCGTGAAGGGGGCC
CCCGAAACTCTGCACTCCATGGCCCGGGAGGTCAAGCGGGAGGCCCTGGAGTGCAGCCTCAAGTTCGT
CGGCTTCATTGTGGTCTCCTGCCCGCTCAAGGCTGACTCGAAGGCCGTGATCCGGGAGATCCAGAATG
CGTCCCACCGGGTGTTCATGATCACGGGAGACAACCCGCTCACTGCATGCCACGTGGCCCAGGAGCTG
CACTTCATTGAAAAGGCCCACACGCTGATCCTGCAGCCTCCCTCCGAGAAAGGCCGGCAGTGCGAGTG
GCGCTCCATTGACGGCAGCATCGTGCTGCCCCTGGCCCGGGGCTCCCCAAAGGCACTGGCCCTGGAGT
ACGCACTGTGCCTCACAGGCGACGGCTTGGCCCACCTGCAGGCCACCGACCCCCAGCAGCTGCTCCGC
CTCATCCCCCATGTGCAGGTGTTCGCCCGTGTGGCTCCCAAGCAGAAGGAGTTTGTCATCACCAGCCT
GAAGGAGCTGGGCTACGTGACCCTCATGTGTGGGGATGGCACCAACGACGTGGGCGCCCTGAAGCATG
CTGACGTGGGTGTGGCGCTCTTGGCCAATGCCCCTGAGCGGGTTGTCGAGCGGCGACGGCGGCCCCGG
GACAGCCCAACCCTGAGCAACAGTGGCATCAGAGCCACCTCCAGGACAGCCAAGCAGCGGTCGGGGCT
CCCTCCCTCCGAGGAGCAGCCAACCTCCCAGAGGGACCGCCTGAGCCAGGTGCTGCGAGACCTCGAGG
ACGAGAGTACGCCCATTGTGAAACTGGGGGATGCCAGCATCGCAGCACCCTTCACCTCCAAGCTCTCA
TCCATCCAGTGCATCTGCCACGTGATCAAGCAGGGCCGCTGCACGCTGGTGACCACGCTACAGATGTT
CAAGATCCTGGCGCTCAATGCCCTCATCCTGGCCTACAGCCAGAGCGTCCTCTACCTGGAGGGAGTCA
AGTTCAGTGACTTCCAGGCCACCCTACAGGGGCTGCTGCTGGCCGGCTGCTTCCTCTTCATCTCCCGT
TCCAAGCCCCTCAAGACCCTCTCCCGAGAACGGCCCCTGCCCAACATCTTCAACCTGTACACCATCCT
CACCGTCATGCTCCAGTTCTTTGTGCACTTCCTGAGCCTTGTCTACCTGTACCGTGAGGCCCAGGCCC
GGAGCCCCGAGAAGCAGGAGCAGTTCGTGGACTTGTACAAGGAGTTTGAGCCAAGCCTGGTCAACAGC
ACCGTCTACATCATGGCCATGGCCATGCAGATGGCCACCTTCGCCATCAATTACAAAGGCCCGCCCTT
CATGGAGAGCCTGCCCGAGAACAAGCCCCTGGTGTGGAGTCTGGCAGTTTCACTCCTGGCCATCATTG

GCCTGCTCCTCGGCTCCTCGCCCGACTTCAACAGCCAGTTTGGCCTCGTGGACATCCCTGTGGAGTTC
AAGCTGGTCATTGCCCAGGTCCTGCTCCTGGACTTCTGCCTGGCGCTCCTGGCCGACCGCGTCCTGCA
GTTCTTCCTGGGGACCCCGAAGCTGAAAGTGCCTTCC**TGA**GATGGCAGT

NOV66a, CG57562-02 Protein Sequence	SEQ ID NO: 986	1167 aa	MW at 128749.4kD

MAAAAAVGNAVPCGARPCGVRPDGQPKPGPQPRALLAAGPALIANGDELVAAVWPYRRLALLRRLTVL
PFAGLLYPAWLGAAAAGCWGWGSSWVQIPEAALLVLATICLAHALTVLSGHWSVHAHCALTCTPEYDP
SKATFVKVVPTPNNGSTELVALHRNEGEDGLEVLSFEFQKIKYSYDALEKKQFLPVAFPVGNAFSYYQ
SNRGFQEDSEIRAAEKKFGSNKAEMVVPDFSELFKERATAPFFVFQVFCVGLWCLDEYWYYSVFTLSM
LVAFEASLVQQQMRNMSEIRKMGNKPHMIQVYRSRKWRPIASDEIVPGDIVSIGRSPQENLVPCDVLL
LRGRCIVDEAMLTGESVPQMKEPIEDLSPDRVLDLQADSRLHVIFGGTKVVQHIPPQKATTGLKPVDS
GCVAYVLRTGFNTSQGKLLRTILFGVKRVTANNLETFIFILFLLVFAIAAAAYVWIEGTKDPSRNRYK
LFLECTLILTSVVPPELPIELSLAVNTSLIALAKLYMYCTEPFRIPFAGKVEVCCFDKTGTLTSDSLV
VRGVAGLRDGKEVTPVSSIPVETHRALASCHSLMQLDDGTLVGDPLEKAMLTAVDWTLTKDEKVFPRS

IKTQGLKIHQRFHFASALKRMSVLASYEKLGSTDLCYIAAVKGAPETLHSMAREVKREALECSLKFVG
FIVVSCPLKADSKAVIREIQNASHRVFMITGDNPLTACHVAQELHFIEKAHTLILQPPSEKGRQCEWR
SIDGSIVLPLARGSPKALALEYALCLTGDGLAHLQATDPQQLLRLIPHVQVFARVAPKQKEFVITSLK
ELGYVTLMCGDGTNDVGALKHADVGVALLANAPERVVERRRRPRDSPTLSNSGIRATSRTAKQRSGLP
PSEEQPTSQRDRLSQVLRDLEDESTPIVKLGDASIAAPFTSKLSSIQCICHVIKQGRCTLVTTLQMFK
ILALNALILAYSQSVLYLEGVKFSDFQATLQGLLLAGCFLFISRSKPLKTLSRERPLPNIFNLYTILT
VMLQFFVHFLSLVYLYREAQARSPEKQEQFVDLYKEFEPSLVNSTVYIMAMAMQMATFAINYKGPPFM
ESLPENKPLVWSLAVSLLAIIGLLLGSSPDFNSQFGLVDIPVEFKLVIAQVLLLDFCLALLADRVLQF
FLGTPKLKVPS

NOV66b, CG57562-01 DNA Sequence	SEQ ID NO: 987	3904 bp
	ORF Start: ATG at 68	ORF Stop: TGA at 3680

TTACCGGAAGTAAAACTTCGGAAGTGAGGCGTTCCTCTGCCCGGAAGTGAGCGCGGCGCTAGGAAAGA
TGGCGGCAGCGGCGGCGGTGGGCAACGCGGTGCCCTGCGGGGCCCGGCCTTGCGGGGTCCGGCCTGAC
GGGCAGCCCAAGCCCGGGCGCAGCCGGCGCGCGCTCCTTGCCGCCGGGCCGGCGCTCATAGCGAACGG
TGACGAGCTGGTGGCTGCCGTGTGGCCGTACCGGCGGTTGGCGCTGTTGCGGCGCCTCACGGTGCTGC
CATTCGCCGGGCTGCTTTACCCGGCCTGGTTGGGTGCCGCAGCCGCTGGCTGCTGGGGCTGGGGCAGC
AGTTGGGTGCAGATCCCCGAAGCTGCGCTGCTCGTGCTTGCCACCATCTGCCTCGCGCACGCGCTCAC
TGTCCTCTCGGGGCATTGGTCTGTGCACGCGCATTGCGCGCTCACCTGCACCCCGGAGTACGACCCCA
GCAAAGCGACCTTTGTGAAGGTGGTGCCAACCCCCAACAATGGCTCCACGGAGCTCGTGGCCCTGCAC
CGCAATGAGGGCGAAGACGGGCTTGAGGTGCTGTCCTTCGAATTCCAGAAGATCAAGTATTCCTACGA
TGCCCTGGAGAAGAAGCAGTTTCTCCCCGTGGCCTTTCCTGTGGGAAACGCCTTCTCATACTATCAGA
GCAACAGAGGCTTCCAGGAAGACTCAGAGATCCGAGCAGCTGAGAAGAAATTTGGGAGCAACAAGGCC
GAGATGGTGGTGCCTGACTTCTCGGAGCTTTTCAAGGAGAGAGCCACAGCCCCCTTCTTTGTATTTCA
GGTGTTCTGTGTGGGGCTCTGGTGCCTGGATGAGTACTGGTACTACAGCGTCTTTACGCTATCCATGC
TGGTGGCGTTCGAGGCCTCGCTGGTGCAGCAGCAGATGCGGAACATGTCGGAGATCCGGAAGATGGGC
AACAAGCCCCACATGATCCAGGTCTACCGAAGCCGCAAGTGGAGGCCCATTGCCAGTGATGAGATCGT
ACCAGGGGACATCGTCTCCATCGGCCGCTCCCCACAGGAGAACCTGGTGCCATGTGACGTGCTTCTGC
TGCGAGGCCGCTGCATCGTAGACGAGGCCATGCTCACGGGGGAGTCCGTGCCACAGATGAAGGAGCCC
ATCGAAGACCTCAGCCCAGACCGGGTGCTGGACCTCCAGGCTGATTCCCGGCTGCACGTCATCTTCGG
GGGCACCAAGGTGGTGCAGCACATCCCCCCACAGAAAGCCACCACGGGCCTGAAGCCGGTTGACAGCG
GGTGCGTGGCCTACGTCCTGCGGACCGGATTCAACACATCCCAGGGCAAGCTGCTGCGCACCATCCTC
TTCGGGGTCAAGAGGGTGACTGCGAACAACCTGGAGACCTTCATCTTCATCCTCTTCCTCCTGGTGTT
TGCCATCGCTGCAGCTGCCTATGTATGGATTGAAGGTACCAAGGACCCCAGCCGGAACCGCTACAAGC
TGTTTCTGGAGTGCACCCTGATCCTCACCTCGGTCGTGCCTCCTGAGCTGCCCATCGAGCTGTCCCTG
GCCGTCAACACCTCCCTCATCGCCCTGGCCAAGCTCTACATGTACTGCACAGAGCCCTTCCGGATCCC
CTTTGCTGGCAAGGTCGAGGTGTGCTGCTTTGACAAGACGGGGACGTTGACCAGTGACAGCCTGGTGG
TGCGCGGTGTGGCCGGGCTGAGAGACGGGAAGGAGGTGACCCCAGTGTCCAGCATCCCTGTAGAAACA
CACCGGGCCCTGGCCTCGTGCCACTCGCTCATGCAGCTGGACGACGGCACCCTCGTGGGTGACCCTCT
AGAGAAGGCCATGCTGACGGCCGTGGACTGGACGCTGACCAAAGATGAGAAAGTATTCCCCCGAAGTA
TTAAAACTCAGGGGCTGAAAATTCACCAGCGCTTTCATTTTGCCAGTGCCCTGAAGCGAATGTCCGTG
CTTGCCTCGTATGAGAAGCTGGGCTCCACCGACCTCTGCTACATCGCGGCCGTGAAGGGGGCCCCCGA
AACTCTGCACTCCATGTTCTCCCAGTGCCCGCCCGACTACCACCACATCCACACCGAGATCTCCCGGG
AAGGAGCCCGCGTCCTGGCGCTGGGGTACAAGGAGCTGGGACACCTCACTCACCAGCAGGCCCGGGAG
GTCAAGCGGGAGGCCCTGGAGTGCAGCCTCAAGTTCGTCGGCTTCATTGTGGTCTCCTGCCCGCTCAA
GGCTGACTCCAAGGCCGTGATCCGGGAGATCCAGAATGCGTCCCACCGGGTGGTCATGATCACGGGAG
ACAACCCGCTCACTGCATGCCACGTGGCCCAGGAGCTGCACTTCATTGAAAAGGCCCACACGCTGATC
CTGCAGCCTCCCTCCGAGAAAGGCCGGCAGTGCGAGTGGCGCTCCATTGACGGCAGCATCGTGCTGCC
CCTGGCCCGGGGCTCCCCAAAGGCACTGGCCCTGGAGTACGCACTGTGCCTCACAGGCGACGGCTTGG
CCCACCTGCAGGCCACCGACCCCCAGCAGCTGCTCCGCCTCATCCCCCATGTGCAGGTGTTCGCCCGT
GTGGCTCCCAAGCAGAAGGAGTTTGTCATCACCAGCCTGAAGGAGCTGGGCTACGTGACCCTCATGTG
TGGGGATGGCACCAACGACGTGGGCGCCCTGAAGCATGCTGACGTGGGTGTGGCGCTCTTGGCCAATG
CCCCTGAGCGGGTTGTCGAGCGGCGACGGCGGCCCCGGGACAGCCCAACCCTGAGCAACAGTGGCATC
AGAGCCACCTCCAGGACAGCCAAGCAGCGGTCGGGGCTCCCTCCCTCCGAGGAGCAGCCAACCTCCCA
GAGGGACCGCCTGAGCCAGGTGCTGCGAGACCTCGAGGACGAGAGTACGCCCATTGTGAAACTGGGGG
ATGCCAGCATCGCAGCACCCTTCACCTCCAAGCTCTCATCCATCCAGTGCATCTGCCACGTGATCAAG
CAGGGCCGCTGCACGCTGGTGACCACGCTACAGATGTTCAAGATCCTGGCGCTCAATGCCCTCATCCT
GGCCTACAGCCAGAGCGTCCTCTACCTGGAGGGAGTCAAGTTCAGTGACTTCCAGGCCACCCTACAGG
GGCTGCTGCTGGCCGGCTGCTTCCTCTTCATCTCCCGTTCCAAGCCCCTCAAGACCCTCTCCCGAGAA

CGGCCCCTGCCCAACATCTTCAACCTGTACACCATCCTCACCGTCATGCTCCAGTTCTTTGTGCACTT
CCTGAGCCTTGTCTACCTGTACCGTGAGGCCCAGGCCCGGAGCCCCGAGAAGCAGGAGCAGTTCGTGG

ACTTGTACAAGGAGTTTGAGCCAAGCCTGGTCAACAGCACCGTCTACATCATGGCCATGGCCATGCAG
ATGGCCACCTTCGCCATCAATTACAAAGGCCCGCCCTTCATGGAGAGCCTGCCCGAGAACAAGCCCCT
GGTGTGGAGTCTGGCAGTTTCACTCCTGGCCATCATTGGCCTGCTCCTCGGCTCCTCGCCCGACTTCA
ACAGCCAGTTTGGCCTCGTGGACATCCCTGTGGAGTTCAAGCTGGTCATTGCCCAGGTCCTGCTCCTG
GACTTCTGCCTGGCGCTCCTGGCCGACCGCGTCCTGCAGTTCTTCCTGGGGACCCCGAAGCTGAAAGT
GCCTTCC**TGA**GATGGCAGTGCTGGTACCCACTGCCCACCCTGGCTGCCGCTGGGCGGGAACCCCAACA
GGGCCCCGGGAGGGAACCCTGCCCCCAACCCCCCACAGCAAGGCTGTACAGTCTCGCCCTTGGAAGAC
TGAGCTGGGACCCCCACAGCCATCCGCTGGCTTGGCCAGCAGAACCAGCCCCAAGCCAGCACCTTTGG
TAAATAAAGCAGCATCTGAGATTTTAAA

NOV66b, CG57562-01 Protein Sequence	SEQ ID NO: 988	1204 aa	MW at 133030.2kD

MAAAAAVGNAVPCGARPCGVRPDGQPKPGRSRRALLAAGPALIANGDELVAAVWPYRRLALLRRLTVL
PFAGLLYPAWLGAAAAGCWGWGSSWVQIPEAALLVLATICLAHALTVLSGHWSVHAHCALTCTPEYDP
SKATFVKVVPTPNNGSTELVALHRNEGEDGLEVLSFEFQKIKYSYDALEKKQFLPVAFPVGNAFSYYQ
SNRGFQEDSEIRAAEKKFGSNKAEMVVPDFSELFKERATAPFFVFQVFCVGLWCLDEYWYYSVFTLSM
LVAFEASLVQQQMRNMSEIRKMGNKPHMIQVYRSRKWRPIASDEIVPGDIVSIGRSPQENLVPCDVLL
LRGRCIVDEAMLTGESVPQMKEPIEDLSPDRVLDLQADSRLHVIFGGTKVVQHIPPQKATTGLKPVDS
GCVAYVLRTGFNTSQGKLLRTILFGVKRVTANNLETFIFILFLLVFAIAAAAYVWIEGTKDPSRNRYK
LFLECTLILTSVVPPELPIELSLAVNTSLIALAKLYMYCTEPFRIPFAGKVEVCCFDKTGTLTSDSLV
VRGVAGLRDGKEVTPVSSIPVETHRALASCHSLMQLDDGTLVGDPLEKAMLTAVDWTLTKDEKVFPRS
IKTQGLKIHQRFHFASALKRMSVLASYEKLGSTDLCYIAAVKGAPETLHSMFSQCPPDYHHIHTEISR
EGARVLALGYKELGHLTHQQAREVKREALECSLKFVGFIVVSCPLKADSKAVIREIQNASHRVVMITG
DNPLTACHVAQELHFIEKAHTLILQPPSEKGRQCEWRSIDGSIVLPLARGSPKALALEYALCLTGDGL
AHLQATDPQQLLRLIPHVQVFARVAPKQKEFVITSLKELGYVTLMCGDGTNDVGALKHADVGVALLAN
APERVVERRRRPRDSPTLSNSGIRATSRTAKQRSGLPPSEEQPTSQRDRLSQVLRDLEDESTPIVKLG
DASIAAPFTSKLSSIQCICHVIKQGRCTLVTTLQMFKILALNALILAYSQSVLYLEGVKFSDFQATLQ
GLLLAGCFLFISRSKPLKTLSRERPLPNIFNLYTILTVMLQFFVHFLSLVYLYREAQARSPEKQEQFV
DLYKEFEPSLVNSTVYIMAMAMQMATFAINYKGPPFMESLPENKPLVWSLAVSLLAIIGLLLGSSPDF
NSQFGLVDIPVEFKLVIAQVLLLDFCLALLADRVLQFFLGTPKLKVPS

NOV66c, SNP13380762 of CG57562-02, DNA Sequence	SEQ ID NO: 989	3517 bp
	ORF Start: ATG at 5	ORF Stop: TGA at 3506
	SNP Pos: 756	SNP Change: T to C

AAAG**ATG**GCGGCAGCGGCGGCGGTGGGCAACGCGGTGCCCTGCGGGGCCCGGCCTTGCGGGGTCCGGC
CTGACGGGCAGCCCAAGCCCGGGCCGCAGCCGCGCGCGCTCCTTGCCGCCGGGCCGGCGCTCATAGCG
AACGGTGACGAGCTGGTGGCTGCCGTGTGGCCGTACCGGCGGTTGGCGCTGTTGCGGCGCCTCACGGT
GCTGCCATTCGCCGGGCTGCTTTACCCGGCCTGGTTGGGTGCCGCAGCCGCTGGCTGCTGGGGCTGGG
GCAGCAGTTGGGTGCAGATCCCCGAAGCTGCGCTGCTCGTGCTTGCCACCATCTGCCTCGCGCACGCG
CTCACTGTCCTCTCGGGGCATTGGTCTGTGCACGCGCATTGCGCGCTCACCTGCACCCCGGAGTACGA
CCCCAGCAAAGCGACCTTTGTGAAGGTGGTGCCAACCCCCAACAATGGCTCCACGGAGCTCGTGGCCC
TGCACCGCAATGAGGGCGAAGACGGGCTTGAGGTGCTGTCCTTCGAATTCCAGAAGATCAAGTATTCC
TACGATGCCCTGGAGAAGAAGCAGTTTCTCCCCGTGGCCTTTCCTGTGGGAAACGCCTTCTCATACTA
TCAGAGCAACAGAGGCTTCCAGGAAGACTCAGAGATCCGAGCAGCTGAGAAGAAATTTGGGAGCAACA
AGGCCGAGATGGTGGTGCCTGACTTCTCGGAGCTTTTCAAGGAGAGAGCCACAGCCCCCTTCTTTGTA
TTTCAGGCGTTCTGTGTGGGGCTCTGGTGCCTGGATGAGTACTGGTACTACAGCGTCTTTACGCTATC
CATGCTGGTGGCGTTCGAGGCCTCGCTGGTGCAGCAGCAGATGCGGAACATGTCGGAGATCCGGAAGA
TGGGCAACAAGCCCCACATGATCCAGGTCTACCGAAGCCGCAAGTGGAGGCCCATTGCCAGTGATGAG
ATCGTACCAGGGGACATCGTCTCCATCGGCCGCTCCCCACAGGAGAACCTGGTGCCATGTGACGTGCT
TCTGCTGCGAGGCCGCTGCATCGTAGACGAGGCCATGCTCACGGGGGAGTCCGTGCCACAGATGAAGG
AGCCCATCGAAGACCTCAGCCCAGACCGGGTGCTGGACCTCCAGGCTGATTCCCGGCTGCACGTCATC
TTCGGGGGCACCAAGGTGGTGCAGCACATCCCCCCACAGAAAGCCACCACGGGCCTGAAGCCGGTTGA
CAGCGGGTGCGTGGCCTACGTCCTGCGGACCGGATTCAACACATCCCAGGGCAAGCTGCTGCGCACCA
TCCTCTTCGGGGTCAAGAGGGTGACTGCGAACAACCTGGAGACCTTCATCTTCATCCTCTTCCTCCTG
GTGTTTGCCATCGCTGCAGCTGCCTATGTATGGATTGAAGGTACCAAGGACCCCAGCCGGAACCGCTA
CAAGCTGTTTCTGGAGTGCACCCTGATCCTCACCTCGGTCGTGCCTCCTGAGCTGCCCATCGAGCTGT

CCCTGGCCGTCAACACCTCCCTCATCGCCCTGGCCAAGCTCTACATGTACTGCACAGAGCCCTTCCGG
ATCCCCTTTGCTGGCAAGGTCGAGGTGTGCTGCTTTGACAAGACGGGGACGTTGACCAGTGACAGCCT
GGTGGTGCGCGGTGTGGCCGGGCTGAGAGACGGGAAGGAGGTGACCCCAGTGTCCAGCATCCCTGTAG
AAACACACCGGGCCCTGGCCTCGTGCCACTCGCTCATGCAGCTGGACGACGGCACCCTCGTGGGTGAC
CCTCTAGAGAAGGCCATGCTGACGGCCGTGGACTGGACGCTGACCAAAGATGAGAAAGTATTCCCCCG
AAGTATTAAAACTCAGGGGCTGAAAATTCACCAGCGCTTTCATTTTGCCAGTGCCCTGAAGCGAATGT
CCGTGCTTGCCTCGTATGAGAAGCTGGGCTCCACCGACCTTTGTTACATCGCGGCCGTGAAGGGGGCC
CCCGAAACTCTGCACTCCATGGCCCGGGAGGTCAAGCGGGAGGCCCTGGAGTGCAGCCTCAAGTTCGT
CGGCTTCATTGTGGTCTCCTGCCCGCTCAAGGCTGACTCGAAGGCCGTGATCCGGGAGATCCAGAATG
CGTCCCACCGGGTGTTCATGATCACGGGAGACAACCCGCTCACTGCATGCCACGTGGCCCAGGAGCTG
CACTTCATTGAAAAGGCCCACACGCTGATCCTGCAGCCTCCCTCCGAGAAAGGCCGGCAGTGCGAGTG
GCGCTCCATTGACGGCAGCATCGTGCTGCCCCTGGCCCGGGGCTCCCCAAAGGCACTGGCCCTGGAGT
ACGCACTGTGCCTCACAGGCGACGGCTTGGCCCACCTGCAGGCCACCGACCCCCAGCAGCTGCTCCGC
CTCATCCCCCATGTGCAGGTGTTCGCCCGTGTGGCTCCCAAGCAGAAGGAGTTTGTCATCACCAGCCT
GAAGGAGCTGGGCTACGTGACCCTCATGTGTGGGGATGGCACCAACGACGTGGGCGCCCTGAAGCATG
CTGACGTGGGTGTGGCGCTCTTGGCCAATGCCCCTGAGCGGGTTGTCGAGCGGCGACGGCGGCCCCGG
GACAGCCCAACCCTGAGCAACAGTGGCATCAGAGCCACCTCCAGGACAGCCAAGCAGCGGTCGGGGCT
CCCTCCCTCCGAGGAGCAGCCAACCTCCCAGAGGGACCGCCTGAGCCAGGTGCTGCGAGACCTCGAGG
ACGAGAGTACGCCCATTGTGAAACTGGGGGATGCCAGCATCGCAGCACCCTTCACCTCCAAGCTCTCA
TCCATCCAGTGCATCTGCCACGTGATCAAGCAGGGCCGCTGCACGCTGGTGACCACGCTACAGATGTT
CAAGATCCTGGCGCTCAATGCCCTCATCCTGGCCTACAGCCAGAGCGTCCTCTACCTGGAGGGAGTCA
AGTTCAGTGACTTCCAGGCCACCCTACAGGGGCTGCTGCTGGCCGGCTGCTTCCTCTTCATCTCCCGT
TCCAAGCCCCTCAAGACCCTCTCCCGAGAACGGCCCCTGCCCAACATCTTCAACCTGTACACCATCCT
CACCGTCATGCTCCAGTTCTTTGTGCACTTCCTGAGCCTTGTCTACCTGTACCGTGAGGCCCAGGCCC
GGAGCCCCGAGAAGCAGGAGCAGTTCGTGGACTTGTACAAGGAGTTTGAGCCAAGCCTGGTCAACAGC
ACCGTCTACATCATGGCCATGGCCATGCAGATGGCCACCTTCGCCATCAATTACAAAGGCCCGCCCTT
CATGGAGAGCCTGCCCGAGAACAAGCCCCTGGTGTGGAGTCTGGCAGTTTCACTCCTGGCCATCATTG

GCCTGCTCCTCGGCTCCTCGCCCGACTTCAACAGCCAGTTTGGCCTCGTGGACATCCCTGTGGAGTTC
AAGCTGGTCATTGCCCAGGTCCTGCTCCTGGACTTCTGCCTGGCGCTCCTGGCCGACCGCGTCCTGCA
GTTCTTCCTGGGGACCCCGAAGCTGAAAGTGCCTTCC**TGA**GATGGCAGT

NOV66c, SNP13380762 of CG57562-02, Protein Sequence	SEQ ID NO: 990	1167 aa	MW at 128721.4kD
	SNP Pos: 251		SNP Change: Val to Ala

MAAAAAVGNAVPCGARPCGVRPDGQPKPGPQPRALLAAGPALIANGDELVAAVWPYRRLALLRRLTVL
PFAGLLYPAWLGAAAAGCWGWGSSWVQIPEAALLVLATICLAHALTVLSGHWSVHAHCALTCTPEYDP
SKATFVKVVPTPNNGSTELVALHRNEGEDGLEVLSFEFQKIKYSYDALEKKQFLPVAFPVGNAFSYYQ
SNRGFQEDSEIRAAEKKFGSNKAEMVVPDFSELFKERATAPFFVFQ**A**FCVGLWCLDEYWYYSVFTLSM
LVAFEASLVQQQMRNMSEIRKMGNKPHMIQVYRSRKWRPIASDEIVPGDIVSIGRSPQENLVPCDVLL
LRGRCIVDEAMLTGESVPQMKEPIEDLSPDRVLDLQADSRLHVIFGGTKVVQHIPPQKATTGLKPVDS
GCVAYVLRTGFNTSQGKLLRTILFGVKRVTANNLETFIFILFLLVFAIAAAAYVWIEGTKDPSRNRYK
LFLECTLILTSVVPPELPIELSLAVNTSLIALAKLYMYCTEPFRIPFAGKVEVCCFDKTGTLTSDSLV
VRGVAGLRDGKEVTPVSSIPVETHRALASCHSLMQLDDGTLVGDPLEKAMLTAVDWTLTKDEKVFPRS
IKTQGLKIHQRFHFASALKRMSVLASYEKLGSTDLCYIAAVKGAPETLHSMAREVKREALECSLKFVG
FIVVSCPLKADSKAVIREIQNASHRVFMITGDNPLTACHVAQELHFIEKAHTLILQPPSEKGRQCEWR
SIDGSIVLPLARGSPKALALEYALCLTGDGLAHLQATDPQQLLRLIPHVQVFARVAPKQKEFVITSLK
ELGYVTLMCGDGTNDVGALKHADVGVALLANAPERVVERRRRPRDSPTLSNSGIRATSRTAKQRSGLP
PSEEQPTSQRDRLSQVLRDLEDESTPIVKLGDASIAAPFTSKLSSIQCICHVIKQGRCTLVTTLQMFK
ILALNALILAYSQSVLYLEGVKFSDFQATLQGLLLAGCFLFISRSKPLKTLSRERPLPNIFNLYTILT
VMLQFFVHFLSLVYLYREAQARSPEKQEQFVDLYKEFEPSLVNSTVYIMAMAMQMATFAINYKGPPFM
ESLPENKPLVWSLAVSLLAIIGLLLGSSPDFNSQFGLVDIPVEFKLVIAQVLLLDFCLALLADRVLQF
FLGTPKLKVPS

NOV66d, SNP13380787 of CG57562-02, DNA Sequence	SEQ ID NO: 991	3517 bp
	ORF Start: ATG at 5	ORF Stop: TGA at 3506
	SNP Pos: 2732	SNP Change: C to A

AAAG**ATG**GCGGCAGCGGCGGCGGTGGGCAACGCGGTGCCCTGCGGGGCCCGGCCTTGCGGGGTCCGGC
CTGACGGGCAGCCCAAGCCCGGGCCGCAGCCGCGCGCGCTCCTTGCCGCCGGGCCGGCGCTCATAGCG
AACGGTGACGAGCTGGTGGCTGCCGTGTGGCCGTACCGGCGGTTGGCGCTGTTGCGGCGCCTCACGGT

GCTGCCATTCGCCGGGCTGCTTTACCCGGCCTGGTTGGGTGCCGCAGCCGCTGGCTGCTGGGGCTGGG
GCAGCAGTTGGGTGCAGATCCCCGAAGCTGCGCTGCTCGTGCTTGCCACCATCTGCCTCGCGCACGCG
CTCACTGTCCTCTCGGGGCATTGGTCTGTGCACGCGCATTGCGCGCTCACCTGCACCCCGGAGTACGA
CCCCAGCAAAGCGACCTTTGTGAAGGTGGTGCCAACCCCCAACAATGGCTCCACGGAGCTCGTGGCCC
TGCACCGCAATGAGGGCGAAGACGGGCTTGAGGTGCTGTCCTTCGAATTCCAGAAGATCAAGTATTCC
TACGATGCCCTGGAGAAGAAGCAGTTTCTCCCCGTGGCCTTTCCTGTGGGAAACGCCTTCTCATACTA
TCAGAGCAACAGAGGCTTCCAGGAAGACTCAGAGATCCGAGCAGCTGAGAAGAAATTTGGGAGCAACA
AGGCCGAGATGGTGGTGCCTGACTTCTCGGAGCTTTTCAAGGAGAGAGCCACAGCCCCCTTCTTTGTA
TTTCAGGTGTTCTGTGTGGGGCTCTGGTGCCTGGATGAGTACTGGTACTACAGCGTCTTTACGCTATC
CATGCTGGTGGCGTTCGAGGCCTCGCTGGTGCAGCAGCAGATGCGGAACATGTCGGAGATCCGGAAGA
TGGGCAACAAGCCCCACATGATCCAGGTCTACCGAAGCCGCAAGTGGAGGCCCATTGCCAGTGATGAG
ATCGTACCAGGGGACATCGTCTCCATCGGCCGCTCCCCACAGGAGAACCTGGTGCCATGTGACGTGCT
TCTGCTGCGAGGCCGCTGCATCGTAGACGAGGCCATGCTCACGGGGGAGTCCGTGCCACAGATGAAGG
AGCCCATCGAAGACCTCAGCCCAGACCGGGTGCTGGACCTCCAGGCTGATTCCCGGCTGCACGTCATC
TTCGGGGGCACCAAGGTGGTGCAGCACATCCCCCCACAGAAAGCCACCACGGGCCTGAAGCCGGTTGA
CAGCGGGTGCGTGGCCTACGTCCTGCGGACCGGATTCAACACATCCCAGGGCAAGCTGCTGCGCACCA
TCCTCTTCGGGGTCAAGAGGGTGACTGCGAACAACCTGGAGACCTTCATCTTCATCCTCTTCCTCCTG
GTGTTTGCCATCGCTGCAGCTGCCTATGTATGGATTGAAGGTACCAAGGACCCCAGCCGGAACCGCTA
CAAGCTGTTTCTGGAGTGCACCCTGATCCTCACCTCGGTCGTGCCTCCTGAGCTGCCCATCGAGCTGT
CCCTGGCCGTCAACACCTCCCTCATCGCCCTGGCCAAGCTCTACATGTACTGCACAGAGCCCTTCCGG
ATCCCCTTTGCTGGCAAGGTCGAGGTGTGCTGCTTTGACAAGACGGGGACGTTGACCAGTGACAGCCT
GGTGGTGCGCGGTGTGGCCGGGCTGAGAGACGGGAAGGAGGTGACCCCAGTGTCCAGCATCCCTGTAG
AAACACACCGGGCCCTGGCCTCGTGCCACTCGCTCATGCAGCTGGACGACGGCACCCTCGTGGGTGAC
CCTCTAGAGAAGGCCATGCTGACGGCCGTGGACTGGACGCTGACCAAAGATGAGAAAGTATTCCCCCG
AAGTATTAAAACTCAGGGGCTGAAAATTCACCAGCGCTTTCATTTTGCCAGTGCCCTGAAGCGAATGT
CCGTGCTTGCCTCGTATGAGAAGCTGGGCTCCACCGACCTTTGTTACATCGCGGCCGTGAAGGGGGCC
CCCGAAACTCTGCACTCCATGGCCCGGGAGGTCAAGCGGGAGGCCCTGGAGTGCAGCCTCAAGTTCGT
CGGCTTCATTGTGGTCTCCTGCCCGCTCAAGGCTGACTCGAAGGCCGTGATCCGGGAGATCCAGAATG
CGTCCCACCGGGTGTTCATGATCACGGGAGACAACCCGCTCACTGCATGCCACGTGGCCCAGGAGCTG
CACTTCATTGAAAAGGCCCACACGCTGATCCTGCAGCCTCCCTCCGAGAAAGGCCGGCAGTGCGAGTG
GCGCTCCATTGACGGCAGCATCGTGCTGCCCCTGGCCCGGGGCTCCCCAAAGGCACTGGCCCTGGAGT
ACGCACTGTGCCTCACAGGCGACGGCTTGGCCCACCTGCAGGCCACCGACCCCCAGCAGCTGCTCCGC
CTCATCCCCCATGTGCAGGTGTTCGCCCGTGTGGCTCCCAAGCAGAAGGAGTTTGTCATCACCAGCCT
GAAGGAGCTGGGCTACGTGACCCTCATGTGTGGGGATGGCACCAACGACGTGGGCGCCCTGAAGCATG
CTGACGTGGGTGTGGCGCTCTTGGCCAATGCCCCTGAGCGGGTTGTCGAGCGGCGACGGCGGCCCCGG
GACAGCCCAACCCTGAGCAACAGTGGCATCAGAGCCACCTCCAGGACAGCCAAGCAGCGGTCGGGGCT
CCCTCCCTCCGAGGAGCAGCCAACCTCCCAGAGGGACCGCCTGAGCCAGGTGCTGCGAGACCTCGAGG
ACGAGAGTACGACCATTGTGAAACTGGGGGATGCCAGCATCGCAGCACCCTTCACCTCCAAGCTCTCA
TCCATCCAGTGCATCTGCCACGTGATCAAGCAGGGCCGCTGCACGCTGGTGACCACGCTACAGATGTT
CAAGATCCTGGCGCTCAATGCCCTCATCCTGGCCTACAGCCAGAGCGTCCTCTACCTGGAGGGAGTCA
AGTTCAGTGACTTCCAGGCCACCCTACAGGGGCTGCTGCTGGCCGGCTGCTTCCTCTTCATCTCCCGT
TCCAAGCCCCTCAAGACCCTCTCCCGAGAACGGCCCCTGCCCAACATCTTCAACCTGTACACCATCCT
CACCGTCATGCTCCAGTTCTTTGTGCACTTCCTGAGCCTTGTCTACCTGTACCGTGAGGCCCAGGCCC
GGAGCCCCGAGAAGCAGGAGCAGTTCGTGGACTTGTACAAGGAGTTTGAGCCAAGCCTGGTCAACAGC
ACCGTCTACATCATGGCCATGGCCATGCAGATGGCCACCTTCGCCATCAATTACAAAGGCCCGCCCTT
CATGGAGAGCCTGCCCGAGAACAAGCCCCTGGTGTGGAGTCTGGCAGTTTCACTCCTGGCCATCATTG

GCCTGCTCCTCGGCTCCTCGCCCGACTTCAACAGCCAGTTTGGCCTCGTGGACATCCCTGTGGAGTTC
AAGCTGGTCATTGCCCAGGTCCTGCTCCTGGACTTCTGCCTGGCGCTCCTGGCCGACCGCGTCCTGCA
GTTCTTCCTGGGGACCCCGAAGCTGAAAGTGCCTTCCTGAGATGGCAGT

NOV66d, SNP13380787 of CG57562-02, Protein Sequence	SEQ ID NO: 992	1167 aa	MW at 128753.4kD
	SNP Pos: 910		SNP Change: Pro to Thr

MAAAAAVGNAVPCGARPCGVRPDGQPKPGPQPRALLAAGPALIANGDELVAAVWPYRRLALLRRLTVL
PFAGLLYPAWLGAAAAGCWGWGSSWVQIPEAALLVLATICLAHALTVLSGHWSVHAHCALTCTPEYDP
SKATFVKVVPTPNNGSTELVALHRNEGEDGLEVLSFEFQKIKYSYDALEKKQFLPVAFPVGNAFSYYQ
SNRGFQEDSEIRAAEKKFGSNKAEMVVPDFSELFKERATAPFFVFQVFCVGLWCLDEYWYYSVFTLSM
LVAFEASLVQQQMRNMSEIRKMGNKPHMIQVYRSRKWRPIASDEIVPGDIVSIGRSPQENLVPCDVLL
LRGRCIVDEAMLTGESVPQMKEPIEDLSPDRVLDLQADSRLHVIFGGTKVVQHIPPQKATTGLKPVDS
GCVAYVLRTGFNTSQGKLLRTILFGVKRVTANNLETFIFILFLLVFAIAAAAYVWIEGTKDPSRNRYK

LFLECTLILTSVVPPELPIELSLAVNTSLIALAKLYMYCTEPFRIPFAGKVEVCCFDKTGTLTSDSLV
VRGVAGLRDGKEVTPVSSIPVETHRALASCHSLMQLDDGTLVGDPLEKAMLTAVDWTLTKDEKVFPRS
IKTQGLKIHQRFHFASALKRMSVLASYEKLGSTDLCYIAAVKGAPETLHSMAREVKREALECSLKFVG
FIVVSCPLKADSKAVIREIQNASHRVFMITGDNPLTACHVAQELHFIEKAHTLILQPPSEKGRQCEWR
SIDGSIVLPLARGSPKALALEYALCLTGDGLAHLQATDPQQLLRLIPHVQVFARVAPKQKEFVITSLK
ELGYVTLMCGDGTNDVGALKHADVGVALLANAPERVVERRRRPRDSPTLSNSGIRATSRTAKQRSGLP
PSEEQPTSQRDRLSQVLRDLEDESTTIVKLGDASIAAPFTSKLSSIQCICHVIKQGRCTLVTTLQMFK
ILALNALILAYSQSVLYLEGVKFSDFQATLQGLLLAGCFLFISRSKPLKTLSRERPLPNIFNLYTILT
VMLQFFVHFLSLVYLYREAQARSPEKQEQFVDLYKEFEPSLVNSTVYIMAMAMQMATFAINYKGPPFM
ESLPENKPLVWSLAVSLLAIIGLLLGSSPDFNSQFGLVDIPVEFKLVIAQVLLLDFCLALLADRVLQF
FLGTPKLKVPS

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 66B.

Table 66B. Comparison of the NOV66 protein sequences.

```
NOV66a  MAAAAAVGNAVPCGARPCGVRPDGQPKPGPQPRALLAAGPALIANGDELVAAVWPYRRLA
NOV66b  MAAAAAVGNAVPCGARPCGVRPDGQPKPGRSRRALLAAGPALIANGDELVAAVWPYRRLA

NOV66a  LLRRLTVLPFAGLLYPAWLGAAAAGCWGWGSSWVQIPEAALLVLATICLAHALTVLSGHW
NOV66b  LLRRLTVLPFAGLLYPAWLGAAAAGCWGWGSSWVQIPEAALLVLATICLAHALTVLSGHW

NOV66a  SVHAHCALTCTPEYDPSKATFVKVVPTPNNGSTELVALHRNEGEDGLEVLSFEFQKIKYS
NOV66b  SVHAHCALTCTPEYDPSKATFVKVVPTPNNGSTELVALHRNEGEDGLEVLSFEFQKIKYS

NOV66a  YDALEKKQFLPVAFPVGNAFSYYQSNRGFQEDSEIRAAEKKFGSNKAEMVVPDFSELFKE
NOV66b  YDALEKKQFLPVAFPVGNAFSYYQSNRGFQEDSEIRAAEKKFGSNKAEMVVPDFSELFKE

NOV66a  RATAPFFVFQVFCVGLWCLDEYWYYSVFTLSMLVAFEASLVQQQMRNMSEIRKMGNKPHM
NOV66b  RATAPFFVFQVFCVGLWCLDEYWYYSVFTLSMLVAFEASLVQQQMRNMSEIRKMGNKPHM

NOV66a  IQVYRSRKWRPIASDEIVPGDIVSIGRSPQENLVPCDVLLLRGRCIVDEAMLTGESVPQM
NOV66b  IQVYRSRKWRPIASDEIVPGDIVSIGRSPQENLVPCDVLLLRGRCIVDEAMLTGESVPQM

NOV66a  KEPIEDLSPDRVLDLQADSRLHVIFGGTKVVQHIPPQKATTGLKPVDSGCVAYVLRTGFN
NOV66b  KEPIEDLSPDRVLDLQADSRLHVIFGGTKVVQHIPPQKATTGLKPVDSGCVAYVLRTGFN

NOV66a  TSQGKLLRTILFGVKRVTANNLETFIFILFLLVFAIAAAAYVWIEGTKDPSRNRYKLFLE
NOV66b  TSQGKLLRTILFGVKRVTANNLETFIFILFLLVFAIAAAAYVWIEGTKDPSRNRYKLFLE

NOV66a  CTLILTSVVPPELPIELSLAVNTSLIALAKLYMYCTEPFRIPFAGKVEVCCFDKTGTLTS
NOV66b  CTLILTSVVPPELPIELSLAVNTSLIALAKLYMYCTEPFRIPFAGKVEVCCFDKTGTLTS

NOV66a  DSLVVRGVAGLRDGKEVTPVSSIPVETHRALASCHSLMQLDDGTLVGDPLEKAMLTAVDW
NOV66b  DSLVVRGVAGLRDGKEVTPVSSIPVETHRALASCHSLMQLDDGTLVGDPLEKAMLTAVDW

NOV66a  TLTKDEKVFPRSIKTQGLKIHQRFHFASALKRMSVLASYEKLGSTDLCYIAAVKGAPETL
NOV66b  TLTKDEKVFPRSIKTQGLKIHQRFHFASALKRMSVLASYEKLGSTDLCYIAAVKGAPETL

NOV66a  HSM-----------------------------------AREVKREALECSLKFVGFIV
NOV66b  HSMFSQCPPDYHHIHTEISREGARVLALGYKELGHLTHQQAREVKREALECSLKFVGFIV

NOV66a  VSCPLKADSKAVIREIQNASHRVFMITGDNPLTACHVAQELHFIEKAHTLILQPPSEKGR
```

```
NOV66b  VSCPLKADSKAVIREIQNASHRVVMITGDNPLTACHVAQELHFIEKAHTLILQPPSEKGR

NOV66a  QCEWRSIDGSIVLPLARGSPKALALEYALCLTGDGLAHLQATDPQQLLRLIPHVQVFARV
NOV66b  QCEWRSIDGSIVLPLARGSPKALALEYALCLTGDGLAHLQATDPQQLLRLIPHVQVFARV

NOV66a  APKQKEFVITSLKELGYVTLMCGDGTNDVGALKHADVGVALLANAPERVVERRRRPRDSP
NOV66b  APKQKEFVITSLKELGYVTLMCGDGTNDVGALKHADVGVALLANAPERVVERRRRPRDSP

NOV66a  TLSNSGIRATSRTAKQRSGLPPSEEQPTSQRDRLSQVLRDLEDESTPIVKLGDASIAAPF
NOV66b  TLSNSGIRATSRTAKQRSGLPPSEEQPTSQRDRLSQVLRDLEDESTPIVKLGDASIAAPF

NOV66a  TSKLSSIQCICHVIKQGRCTLVTTLQMFKILALNALILAYSQSVLYLEGVKFSDFQATLQ
NOV66b  TSKLSSIQCICHVIKQGRCTLVTTLQMFKILALNALILAYSQSVLYLEGVKFSDFQATLQ

NOV66a  GLLLAGCFLFISRSKPLKTLSRERPLPNIFNLYTILTVMLQFFVHFLSLVYLYREAQARS
NOV66b  GLLLAGCFLFISRSKPLKTLSRERPLPNIFNLYTILTVMLQFFVHFLSLVYLYREAQARS

NOV66a  PEKQEQFVDLYKEFEPSLVNSTVYIMAMAMQMATFAINYKGPPFMESLPENKPLVWSLAV
NOV66b  PEKQEQFVDLYKEFEPSLVNSTVYIMAMAMQMATFAINYKGPPFMESLPENKPLVWSLAV

NOV66a  SLLAIIGLLLGSSPDFNSQFGLVDIPVEFKLVIAQVLLLDFCLALLADRVLQFFLGTPKL
NOV66b  SLLAIIGLLLGSSPDFNSQFGLVDIPVEFKLVIAQVLLLDFCLALLADRVLQFFLGTPKL

NOV66a  KVPS
NOV66b  KVPS

NOV66a  (SEQ ID NO:  986)
NOV66b  (SEQ ID NO:  988)
```

Further analysis of the NOV66a protein yielded the following properties shown in Table 66C.

Table 66C. Protein Sequence Properties NOV66a

SignalP analysis:	Cleavage site between residues 16 and 17

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 0;  pos.chg 0;  neg.chg 0
      H-region:  length 15;  peak value   7.85
      PSG score:   3.45

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -4.35
      possible cleavage site: between 46 and 47

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:  11
      INTEGRAL    Likelihood = -7.38   Transmembrane  100 - 116
      INTEGRAL    Likelihood = -2.23   Transmembrane  243 - 259
```

```
      INTEGRAL    Likelihood = -0.16   Transmembrane  265 - 281
      INTEGRAL    Likelihood =-11.94   Transmembrane  446 - 462
      INTEGRAL    Likelihood =  0.10   Transmembrane  493 - 509
      INTEGRAL    Likelihood = -0.64   Transmembrane  672 - 688
      INTEGRAL    Likelihood = -0.69   Transmembrane  944 - 960
      INTEGRAL    Likelihood = -1.38   Transmembrane  978 - 994
      INTEGRAL    Likelihood = -3.61   Transmembrane 1017 -1033
      INTEGRAL    Likelihood = -9.45   Transmembrane 1097 -1113
      INTEGRAL    Likelihood = -8.65   Transmembrane 1134 -1150
      PERIPHERAL  Likelihood =  1.01 (at 67)
      ALOM score: -11.94  (number of TMSs: 11)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 107
      Charge difference:  3.0   C( 2.0) - N(-1.0)
      C > N: C-terminal side will be inside

>>>Caution: Inconsistent mtop result with signal peptide
>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       2       Hyd Moment(75):  2.14
      Hyd Moment(95):  0.67    G content:       3
      D/E content:     1       S/T content:     0
      Score: -5.79

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 43  PRA|LL

NUCDISC: discrimination of nuclear localization signals
      pat4: RRRR (5) at  855
      pat4: RRRP (4) at  856
      pat4: RRPR (4) at  857
      pat7: none
      bipartite: RKMGNKPHMIQVYRSRK at  292
      content of basic residues:  10.4%
      NLS Score:  0.83

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: LKVP

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none
```

```
NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       77.8 %: endoplasmic reticulum
       11.1 %: mitochondrial
       11.1 %: vacuolar

>> prediction for CG57562-02 is end (k=9)
```

A search of the NOV66a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 66D.

Table 66D. Geneseq Results for NOV66a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV66a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU91186	Human HEAT-3 polypeptide - Homo sapiens, 1204 aa. [WO200216591-A2, 28-FEB-2002]	1..1167 1..1204	1165/1204 (96%) 1165/1204 (96%)	0.0
AAE16780	Human transporter and ion channel-17 (TRICH-17) protein - Homo sapiens, 1197 aa. [WO200192304-A2, 06-DEC-2001]	1..1167 1..1197	1159/1204 (96%) 1159/1204 (96%)	0.0
ABG30096	Novel human diagnostic protein #30087 - Homo sapiens, 914 aa. [WO200175067-A2, 11-OCT-2001]	328..1167 7..914	829/915 (90%) 830/915 (90%)	0.0
AAB42279	Human ORFX ORF2043 polypeptide sequence SEQ ID NO:4086 - Homo sapiens, 789 aa. [WO200058473-A2, 05-OCT-2000]	418..1167 1..789	734/789 (93%) 737/789 (93%)	0.0
AAM93412	Human polypeptide, SEQ ID NO: 3024 - Homo sapiens, 572 aa. [EP1130094-A2, 05-SEP-2001]	633..1167 1..572	530/572 (92%) 530/572 (92%)	0.0

In a BLAST search of public sequence databases, the NOV66a protein was found to have homology to the proteins shown in the BLASTP data in Table 66E.

Table 66E. Public BLASTP Results for NOV66a				
Protein Accession Number	**Protein/Organism/Length**	**NOV66a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9HD20	Probable cation-transporting ATPase 2 (EC 3.6.3.-) (CGI-152) - Homo sapiens (Human), 1204 aa.	1..1167 1..1204	1166/1204 (96%) 1166/1204 (96%)	0.0
CAD29018	Sequence 8 from Patent WO0216591 - Homo sapiens (Human), 1204 aa.	1..1167 1..1204	1165/1204 (96%) 1165/1204 (96%)	0.0
Q9EPE9	Probable cation-transporting ATPase 2 (EC 3.6.3.-) (CATP) - Mus musculus (Mouse), 1200 aa.	5..1167 2..1200	1103/1200 (91%) 1127/1200 (93%)	0.0
Q9VKJ6	BCDNA:GH06032 protein - Drosophila melanogaster (Fruit fly), 1225 aa.	98..1164 105..1219	594/1123 (52%) 770/1123 (67%)	0.0
Q8NC73	Hypothetical protein FLJ90439 - Homo sapiens (Human), 572 aa.	633..1167 1..572	530/572 (92%) 530/572 (92%)	0.0

PFam analysis indicates that the NOV66a protein contains the domains shown in the Table 66F.

Table 66F. Domain Analysis of NOV66a			
Pfam Domain	**NOV66a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
E1-E2_ATPase	266..511	62/284 (22%) 166/284 (58%)	7.4e-05
Hydrolase	527..848	35/327 (11%) 199/327 (61%)	2.8e-06

Example 67.

The NOV67 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 67A.

Table 67A. NOV67 Sequence Analysis		
NOV67a, CG57758-03 DNA Sequence	SEQ ID NO: 993	3147 bp
	ORF Start: ATG at 2	ORF Stop: TAG at 1706

GATGGCCTCGGCGCTGAGCTATGTCTCCAAGTTCAAGTCCTTCGTGATCTTGTTCGTCACCCCGCTCC
TGCTGCTGCCACTCGTCATTCTGATGCCCGCCAAGTTTGTCAGGTGTGCCTACGTCATCATCCTCATG
GCCATTTACTGGTGCACAGAAGTCATCCCTCTGGCTGTCACCTCTCTCATGCCTGTCTTGCTTTTCCC
ACTCTTCCAGATTCTGGACTCCAGGCAGGTGTGTGTCCAGTACATGAAGGACACCAACATGCTGTTCC
TGGGCGGCCTCATCGTGGCCGTGGCTGTGGAGCGCTGGAACCTGCACAAGAGGATCGCCCTGCGCACG
CTCCTCTGGGTGGGGGCCAAGCCTGCACGGCTGATGCTGGGCTTCATGGGCGTCACAGCCCTCCTGTC
CATGTGGATCAGTAACACGGCAACCACGGCCATGATGGTGCCCATCGTGGAGGCCATATTGCAGCAGA
TGGAAGCCACAAGCGCAGCCACCGAGGCCGGCCTGGAGCTGGTGGACAAGGGCAAGGCCAAGGAGCTG
CCAGGGAGTCAAGTGATTTTTGAAGGCCCCACTCTGGGGCAGCAGGAAGACCAAGAGCGGAAGAGGTT
GTGTAAGGCCATGACCCTGTGCATCTGCTACGCGGCCAGCATCGGGGGCACCGCCACCCTGACCGGGA
CGGGACCCAACGTGGTGCTCCTGGGCCAGATGAACGAGTTGTTTCCTGACAGCAAGGACCTCGTGAAC
TTTGCTTCCTGGTTTGCATTTGCCTTTCCCAACATGCTGGTGATGCTGCTGTTCGCCTGGCTGTGGCT
CCAGTTTGTTTACATGAGATTCAATTTTAAAAAGTCCTGGGGCTGCGGGCTAGAGAGCAAGAAAAACG
AGAAGGCTGCCCTCAAGGTGCTGCAGGAGGAGTACCGGAAGTTGGGGCCCTTGTCCTTCGCGGAGATC
AACGTGCTGATCTGCTTCTTCCTGCTGGTCATCCTGTGGTTCTCCCGAGACCCCGGCTTCATGCCCGG
CTGGCTGACTGTTGCCTGGGTGGAGGGTGAGACAAAGTATGTCTCCGATGCCACTGTGGCCATCTTTG
TGGCCACCCTGCTATTCATTGTGCCTTCACAGAAGCCCAAGTTTAACTTCCGCAGCCAGACTGAGGAA
GAAAGGAAAACTCCATTTTATCCCCCTCCCCTGCTGGATTGGAAGGTAACCCAGGAGAAAGTGCCCTG
GGGCATCGTGCTGCTACTAGGGGGCGGATTTGCTCTGGCTAAAGGATCCGAGGCCTCGGGGCTGTCCG
TGTGGATGGGGAAGCAGATGGAGCCCTTGCACGCAGTGCCCCCGGCAGCCATCACCTTGATCTTGTCC
TTGCTCGTTGCCGTGTTCACTGAGTGCACAAGCAACGTGGCCACCACCACCTTGTTCCTGCCCATCTT
TGCCTCCATGTCTCGCTCCATCGGCCTCAATCCGCTGTACATCATGCTGCCCTGTACCCTGAGTGCCT
CCTTTGCCTTCATGTTGCCTGTGGCCACCCCTCCAAATGCCATCGTGTTCACCTATGGGCACCTCAAG
GTTGCTGACATGGTGAAAACAGGAGTCATAATGAACATAATTGGAGTCTTCTGTGTGTTTTTGGCTGT
CAACACCTGGGGACGGGCCATATTTGACTTGGATCATTTCCCTGACTGGGCTAATGTGACACATATTG
AGACTTAGGAAGAGCCACAAGACCACACACACAGCCCTTACCCTCCTCAGGACTACCGAACCTTCTGG
CACACCTTGTACAGAGTTTTGGGGTTCACACCCCAAAATGACCCAACGATGTCCACACACCACCAAAA
CCCAGCCAATGGGCCACCTCTTCCTCCAAGCCCAGATGCAGAGATGGTCATGGGCAGCTGGAGGGTAG
GCTCAGAAATGAAGGGAACCCCTCAGTGGGCTGCTGGACCCATCTTTCCCAAGCCTTGCCATTATCTC
TGTGAGGGAGGCCAGGTAGCCGAGGGATCAGGATGCAGGCTGCTGTACCCGCTCTGCCTCAAGCATCC
CCCACACAGGGCTCTGGTTTTCACTCGCTTCGTCCTAGATAGTTTAAATGGGAATCGGATCCCCTGGT
TGAGAGCTAAGACAACCACCTACCAGTGCCCATGTCCCTTCCAGCTCACCTTGAGCAGCCTCAGATCA
TCTCTGTCACTCTGGAAGGGACACCCCAGCCAGGGACGGAATGCCTGGTCTTGAGCAACCTCCCACTG
CTGGAGTGCGAGTGGGAATCAGAGCCTCCTGAAGCCTCTGGGAACTCCTCCTGTGGCCACCACCAAAG
GATGAGGAATCTGAGTTGCCAACTTCAGGACGACACCTGGCTTGCCACCCACAGTGCACCACAGGCCA
ACCTACGCCCTTCATCACTTGGTTCTGTTTTAATCGACTGGCCCCCTGTCCCACCTCTCCAGTGAGCC
TCCTTCAACTCCTTGGTCCCCTGTTGTCTGGGTCAACATTTGCCGAGACGCCTTGGCTGGCACCCTCT
GGGGTCCCCCTTTTCTCCCAGGCAGGTCATCTTTTCTGGGAGATGCTTCCCCTGCCATCCCCAAATAG
CTAGGATCACACTCCAAGTATGGGCAGTGATGGCGCTCTGGGGGCCACAGTGGGCTATCTAGGCCCTC
CCTCACCTGAGGCCCAGAGTGGACACAGCTGTTAATTTCCACTGGCTATGCCACTTCAGAGTCTTTCA
TGCCAGCGTTTGAGCTCCTCTGGGTAAAATCTTCCCTTTGTTGACTGGCCTTCACAGCCATGGCTGGT
GACAACAGAGGATCGTTGAGATTGAGCAGCGCTTGGTGATCTCTCAGCAAACAACCCCTGCCCGTGGG
CCAATCTACTTGAAGTTACTCGGACAAAGACCCCAAAGTGGGGCAACAACTCCAGAGAGGCTGTGGGA
ATCTTCAGAAGCCCCCCTGTAAGAGACAGACATGAGAGACAAGCATCTTCTTTCCCCCGCAAGTCCAT
TTTATTTCCTTCTTGTGCTGCTCTGGAAGAGAGGCAGTAGCAAAGAGATGAGCTCCTGGATGGCATTT
TCCAGGGCAGGAGAAAGTATGAGAGCCTCAGGAAACCCCATCAAGGACCGAGTATGTGTCTGGTTCCT
TTGGTGGTTGGCTTCTGGC

NOV67a, CG57758-03 Protein Sequence	SEQ ID NO: 994	568 aa	MW at 63061.4kD

MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKFVRCAYVIILMAIYWCTEVIPLAVTSLMPVLLFP
LFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLWVGAKPARLMLGFMGVTALLS
MWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVDKGKAKELPGSQVIFEGPTLGQQEDQERKRL
CKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLVNFASWFAFAFPNMLVMLLFAWLWL
QFVYMRFNFKKSWGCGLESKKNEKAALKVLQEEYRKLGPLSFAEINVLICFFLLVILWFSRDPGFMPG

WLTVAWVEGETKYVSDATVAIFVATLLFIVPSQKPKFNFRSQTEEERKTPFYPPPLLDWKVTQEKVPW
GIVLLLGGGFALAKGSEASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIF
ASMSRSIGLNPLYIMLPCTLSASFAFMLPVATPPNAIVFTYGHLKVADMVKTGVIMNIIGVFCVFLAV
NTWGRAIFDLDHFPDWANVTHIET

NOV67b, 308537854 DNA Sequence	SEQ ID NO: 995	1729 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCAGATCTCCCACCATGGCCTCGGCGCTGAGCTATGTCTCCAAGTTCAAGTCCTTCGTGATCTTGT
TCGTCACCCCGCTCCTGCTGCTGCCACTCGTCATTCTGATGCCCGCCAAGTTTGTCAGGTGTGCCTAC
GTCATCATCCTCATGGCCATTTACTGGTGCACAGAAGTCATCCCTCTGGCTGTCACCTCTCTCATGCC
TGTCTTGCTTTTCCCACTCTTCCAGATTCTGGACTCCAGGCAGGTGTGTGTCCAGTACATGAAGGACA
CCAACATGCTGTTCCTGGGCGGCCTCATCGTGGCCGTGGCTGTGGAGCGCTGGAACCTGCACAAGAGG
ATCGCCCTGCGCACGCTCCTCTGGGTGGGGGCCAAGCCTGCACGGCTGATGCTGGGCTTCATGGGCGT
CACAGCCCTCCTGTCCATGTGGATCAGTAACACGGCAACCACGGCCATGATGGTGCCCATCGTGGAGG
CCATATTGCAGCAGATGGAAGCCACAAGCGCAGCCACCGAGGCCGGCCTGGAGCTGGTGGACAAGGGC
AAGGCCAAGGAGCTGCCAGGGAGTCAAGTGATTTTTGAAGGCCCCACTCTGGGGCAGCAGGAAGACCA
AGAGCGGAAGAGGTTGTGTAAGGCCATGACCCTGTGCATCTGCTACGCGGCCAGCATCGGGGGCACCG
CCACCCTGACCGGGACGGGACCCAACGTGGTGCTCCTGGGCCAGATGAACGAGTTGTTTCCTGACAGC
AAGGACCTCGTGAACTTTGCTTCCTGGTTTGCATTTGCCTTTCCCAACATGCTGGTGATGCTGCTGTT
CGCCTGGCTGTGGCTCCAGTTTGTTTACATGAGATTCAATTTTAAAAAGTCCTGGGGCTGCGGGCTAG
AGAGCAAGAAAAACGAGAAGGCTGCCCTCAAGGTGCTGCAGGAGGAGTACCGGAAGTTGGGGCCCTTG
TCCTTCGCGGAGATCAACGTGCTGATCTGCTTCTTCCTGCTGGTCATCCTGTGGTTCTCCCGAGACCC
CGGCTTCATGCCCGGCTGGCTGACTGTTGCCTGGGTGGAGGGTGAGACAAAGTATGTCTCCGATGCCA
CTGTGGCCATCTTTGTGGCCACCCTGCTATTCATTGTGCCTTCACAGAAGCCCAAGTTTAACTTCCGC
AGCCAGACTGAGGAAGAAAGGAAAACTCCATTTTATCCCCCTCCCCTGCTGGATTGGAAGGTAACCCA
GGAGAAAGTGCCCTGGGGCATCGTGCTGCTACTAGGGGGCGGATTTGCTCTGGCTAAAGGATCCGAGG
CCTCGGGGCTGTCCGTGTGGATGGGGAAGCAGATGGAGCCCTTGCACGCAGTGCCCCCGGCAGCCATC
ACCTTGATCTTGTCCTTGCTCGTTGCCGTGTTCACTGAGTGCACAAGCAACGTGGCCACCACCACCTT
GTTCCTGCCCATCTTTGCCTCCATGTCTCGCTCCATCGGCCTCAATCCGCTGTACATCATGCTGCCCT
GTACCCTGAGTGCCTCCTTTGCCTTCATGTTGCCTGTGGCCACCCCTCCAAATGCCATCGTGTTCACC
TATGGGCACCTCAAGGTTGCTGACATGGTGAAAACAGGAGTCATAATGAACATAATTGGAGTCTTCTG
TGTGTTTTTGGCTGTCAACACCTGGGGACGGGCCATATTTGACTTGGATCATTTCCCTGACTGGGCTA
ATGTGACACATATTGAGACTCTCGAGGGC

NOV67b, 308537854 Protein Sequence	SEQ ID NO: 996	576 aa	MW at 63903.4kD

TRSPTMASALSYVSKFKSFVILFVTPLLLLPLVILMPAKFVRCAYVIILMAIYWCTEVIPLAVTSLMP
VLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLWVGAKPARLMLGFMGV
TALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVDKGKAKELPGSQVIFEGPTLGQQEDQ
ERKRLCKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLVNFASWFAFAFPNMLVMLLF
AWLWLQFVYMRFNFKKSWGCGLESKKNEKAALKVLQEEYRKLGPLSFAEINVLICFFLLVILWFSRDP
GFMPGWLTVAWVEGETKYVSDATVAIFVATLLFIVPSQKPKFNFRSQTEEERKTPFYPPPLLDWKVTQ
EKVPWGIVLLLGGGFALAKGSEASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTL
FLPIFASMSRSIGLNPLYIMLPCTLSASFAFMLPVATPPNAIVFTYGHLKVADMVKTGVIMNIIGVFC
VFLAVNTWGRAIFDLDHFPDWANVTHIETLEG

NOV67c, CG57758-01 DNA Sequence	SEQ ID NO: 997	1790 bp
	ORF Start: ATG at 16	ORF Stop: TAG at 1720

TCTCCCTCCCGCGCGATGGCCTCGGCGCTGAGCTATGTCTCCAAGTTCAAGTCCTTCGTGATCTTGTT
CGTCACCCCGCTCCTGCTGCTGCCACTCGTCATTCTGATGCCCGCCAAGGTCAGTTGTGCCTACGTCA
TCATCCTCATGGCCATTTACTGGTGCACAGAAGTCATCCCTCTGGCTGTCACCTCTCTCATGCCTGTC
TTGCTTTTCCCACTCTTCCAGATTCTGGACTCCAGGCAGGTGTGTGTCCAGTACATGAAGGACACCAA
CATGCTGTTCCTGGGCGGCCTCATCGTGGCCGTGGCTGTGGAGCGCTGGAACCTGCACAAGAGGATCG
CCCTGCGCACGCTCCTCTGGGTGGGGGCCAAGCCTGCACGGCTGATGCTGGGCTTCATGGGCGTCACA
GCCCTCCTGTCCATGTGGATCAGTAACACGGCAACCACGGCCATGATGGTGCCCATCGTGGAGGCCAT
ATTGCAGCAGATGGAAGCCACAAGCGCAGCCACCGAGGCCGGCCTGGAGCTGGTGGACAAGGGCAAGG
CCAAGGAGCTGCCAGGGAGTCAAGTGATTTTTGAAGGCCCCACTCTGGGGCAGCAGGAAGACCAAGAG
CGGAAGAGGTTGTGTAAGGCCATGACCCTGTGCATCTGCTACGCGGCCAGCATCGGGGGCACCGCCAC

CCTGACCGGGACGGGACCCAACGTGGTGCTCCTGGGCCAGATGAACGAGTTGTTTCCTGACAGCAAGG
ACCTCGTGAACTTTGCTTCCTGGTTTGCATTTGCCTTTCCCAACATGCTGGTGATGCTGCTGTTCGCC
TGGCTGTGGCTCCAGTTTGTTTACATGTTCTCCAGTTTTAAAAAGTCCTGGGGCTGCGGGCTAGAGAG
CAAGAAAAACGAGAAGGCTGCCCTCAAGGTGCTGCAGGAGGAGTACCGGAAGCTGGGGCCCTTGTCCT
TCGCGGAGATCAACGTGCTGATCTGCTTCTTCCTGCTGGTCATCCTGTGGTTCTCCCGAGACCCCGGC
TTCATGCCCGGCTGGCTGACTGTTGCCTGGGTGGAGGGTGAGACAAAGTATGTCTCCGATGCCACTGT
GGCCATCTTTGTGGCCACCCTGCTATTCATTGTGCCTTCACAGAAGCCCAAGTTTAACTTCCGCAGCC
AGACTGAGGAAGGTAAGTCTCCTGTTCTGATCGCCCCCCCTCCCCTGCTGGATTGGAAGGTAACCCAG
GAGAAAGTGCCCTGGGGCATCGTGCTGCTACTAGGGGGCGGATTTGCTCTGGCTAAAGGATCCGAGGC
CTCGGGGCTGTCCGTGTGGATGGGGAAGCAGATGGAGCCCTTGCACGCAGTGCCCCCGGCAGCCATCA
CCTTGATCTTGTCCTTGCTCGTTGCCGTGTTCACTGAGTGCACAAGCAACGTGGCCACCACCACCTTG
TTCCTGCCCATCTTTGCCTCCATGTCTCGCTCCATCGGCCTCAATCCGCTGTACATCATGCTGCCCTG
TACCCTGAGTGCCTCCTTTGCCTTCATGTTGCCTGTGGCCACCCCTCCAAATGCCATCGTGTTCACCT
ATGGGCACCTCAAGGTTGCTGACATGGTGAAAACAGGAGTCATAATGAACATAATTGGAGTCTTCTGT
GTGTTTTTGGCTGTCAACACCTGGGGACGGGCCATATTTGACTTGGATCATTTCCCTGACTGGGCTAA
TGTGACACATATTGAGACTTAGGAAGAGCCACAAGACCACACACACAGCCCTTACCCTCCTCAGGACT
ACCGAACCTTCTGGCACACCTT

NOV67c, CG57758-01 Protein Sequence	SEQ ID NO: 998	568 aa	MW at 62592.9kD

MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKVSCAYVIILMAIYWCTEVIPLAVTSLMPVLLFPL
FQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLWVGAKPARLMLGFMGVTALLSM
WISNTATTAMMVPIVEAILQQMEATSAATEAGLELVDKGKAKELPGSQVIFEGPTLGQQEDQERKRLC
KAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLVNFASWFAFAFPNMLVMLLFAWLWLQ
FVYMFSSFKKSWGCGLESKKNEKAALKVLQEEYRKLGPLSFAEINVLICFFLLVILWFSRDPGFMPGW
LTVAWVEGETKYVSDATVAIFVATLLFIVPSQKPKFNFRSQTEEGKSPVLIAPPPLLDWKVTQEKVPW
GIVLLLGGGFALAKGSEASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIF
ASMSRSIGLNPLYIMLPCTLSASFAFMLPVATPPNAIVFTYGHLKVADMVKTGVIMNIIGVFCVFLAV
NTWGRAIFDLDHFPDWANVTHIET

NOV67d, CG57758-02 DNA Sequence	SEQ ID NO: 999	1899 bp
	ORF Start: ATG at 31	ORF Stop: TAG at 1879

CGTCTCGCCCGCCAGTCTCCCTCCCGCGCGATGGCCTCGGCGCTGAGCTATGTCTCCAAGTTCAAGTC
CTTCGTGATCTTGTTCGTCACCCCGCTCCTGCTGCTGCCACTCGTCATTCTGATGCCCGCCAAGGTCA
GTTGCTGTGCCTACGTCATCATCCTCATGGCCATTTACTGGTGCACAGAAGTCATCCCTCTGGCTGTC
ACCTCTCTCATGCCTGTCTTGCTTTTCCCACTCTTCCAGATTCTGGACTCCAGGCAGGTGTGTGTCCA
GTACATGAAGGACACCAACATGCTGTTCCTGGGCGGCCTCATCGTGGCCGTGGCTGTGGAGCGCTGGA
ACCTGCACAAGAGGATCGCCCTGCGCACGCTCCTCTGGGTGGGGGCCAAGCCTGCACGGCTGATGCTG
GGCTTCATGGGCGTCACAGCCCTCCTGTCCATGTGGATCAGTAACACGGCAACCACGGCCATGATGGT
GCCCATCGTGGAGGCCATATTGCAGCAGATGGAAGCCACAAGCGCAGCCACCGAGGCCGGCCTGGAGG
GACAAGGTACCACAATAAACAACCTGAATGCACTGGAGGATGATACAGTGAAAGCAGTACTAGGAGGA
AAGTGTGTAGCTATAATAAGCACTTACGTCAAAAAAGTAGAAAAACTTCAAATAAACAATCTAATGAC
ACCTCTTAAAAAACTAGAAAAGCAAGAGCAACAGGACCTAGGGCCTGGCATCAGGCCTCAGGACTCTG
CCCAGTGCCAGGAAGACCAAGAGCGGAAGAGGTTGTGTAAGGCCATGACCCTGTGCATCTGCTACGCG
GCCAGCATCGGGGGCACCGCCACCCTGACCGGGACGGGACCCAACGTGGTGCTCCTGGGCCAGATGAA
CGAGTTGTTTCCTGACAGCAAGGACCTCGTGAACTTTGCTTCCTGGTTTGCATTTGCCTTTCCCAACA
TGCTGGTGATGCTGCTGTTCGCCTGGCTGTGGCTCCAGTTTGTTTACATGTTCTCCAGTTTTAAAAAG
TCCTGGGGCTGCGGGCTAGAGAGCAAGAAAAACGAGAAGGCTGCCCTCAAGGTGCTGCAGGAGGAGTA
CCGGAAGCTGGGGCCCTTGTCCTTCGCGGAGATCAACGTGCTGATCTGCTTCTTCCTGCTGGTCATCC
TGTGGTTCTCCCGAGACCCCGGCTTCATGCCCGGCTGGCTGACTGTTGCCTGGGTGGAGGGTGAGACA
AAGTCAGTCTCCGATGCCACTGTGGCCATCTTTGTGGCCACCCTGCTATTCATTGTGCCTTCACAGAA
GCCCAAGTTTAACTTCCGCAGCCAGACTGAGGAAGGTAAGTCTCCTGTTCTGATCGCCCCCCCTCCCC
TGCTGGATTGGAAGGTAACCCAGGAGAAAGTGCCCTGGGGCATCGTGCTGCTACTAGGGGGCGGATTT
GCTCTGGCTAAAGGATCCGAGGCCTCGGGGCTGTCCGTGTGGATGGGGAAGCAGATGGAGCCCTTGCA
CGCAGTGCCCCCGGCAGCCATCACCTTGATCTTGTCCTTGCTCGTTGCCGTGTTCACTGAGTGCACAA
GCAACGTGGCCACCACCACCTTGTTCCTGCCCATCTTTGCCTCCATGTCTCGCTCCATCGGCCTCAAT
CCGCTGTACATCATGCTGCCCTGTACCCTGAGTGCCTCCTTTGCCTTCATGTTGCCTGTGGCCACCCC
TCCAAATGCCATCGTGTTCACCTATGGGCACCTCAAGGTTGCTGACATGGTAAAAACAGGAGTCATAA
TGAACATAATTGGAGTCTTCTGTGTGTTTTTGGCTGTCAACACCTGGGGACGGGCCATATTTGACTTG

GATCATTTCCCTGACTGGGCTAATGTGACACATATTGAGACTTAGGAAGAGCCACAAGACCAC

NOV67d, CG57758-02 Protein Sequence	SEQ ID NO: 1000	616 aa	MW at 67816.9kD

MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKVSCCAYVIILMAIYWCTEVIPLAVTSLMPVLLFP
LFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLWVGAKPARLMLGFMGVTALLS
MWISNTATTAMMVPIVEAILQQMEATSAATEAGLEGQGTTINNLNALEDDTVKAVLGGKCVAIISTYV
KKVEKLQINNLMTPLKKLEKQEQQDLGPGIRPQDSAQCQEDQERKRLCKAMTLCICYAASIGGTATLT
GTGPNVVLLGQMNELFPDSKDLVNFASWFAFAFPNMLVMLLFAWLWLQFVYMFSSFKKSWGCGLESKK
NEKAALKVLQEEYRKLGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVEGETKSVSDATVAI
FVATLLFIVPSQKPKFNFRSQTEEGKSPVLIAPPPLLDWKVTQEKVPWGIVLLLGGGFALAKGSEASG
LSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIFASMSRSIGLNPLYIMLPCTL
SASFAFMLPVATPPNAIVFTYGHLKVADMVKTGVIMNIIGVFCVFLAVNTWGRAIFDLDHFPDWANVT
HIET

NOV67e, CG57758-04 DNA Sequence	SEQ ID NO: 1001	1606 bp
	ORF Start: ATG at 2	ORF Stop: TAG at 1568

GATGGCCTCGGCGCTGAGCTATGTCTCCAAGTTCAAGTCCTTCGTGATCTTGTTCGTCACCCCGCTCC
TGCTGCTGCCACTCGTCATTCTGATGCCCGCCAAGTTTGTCAGGTGTGCCTACGTCATCATCCTCATG
GCCATTTACTGGTGCACAGAAGTCATCCCTCTGGCTGTCACCTCTCTCATGCCTGTCTTGCTTTTCCC
ACTCTTCCAGATTCTGGACTCCAGGCAGGTGTGTGTCCAGTACATGAAGGACACCAACATGCTGTTCC
TGGGCGGCCTCATCGTGGCCGTGGCTGTGGAGCGCTGGAACCTGCACAAGAGGATCGCCCTGCGCACG
CTCCTCTGGGTGGGGGCCAAGCCTGCACGGCTGATGCTGGGCTTCATGGGCGTCACAGCCCTCCTGTC
CATGTGGATCAGTAACACGGCAACCACGGCCATGATGGTGCCCATCGTGGAGGCCATATTGCAGCAGA
TGGAAGCCACAAGCGCAGCCACCGAGGCCGGCCTGGAGCTGGTGGACAAGGGCAAGGCCAAGGAGCTG
CCAGGGAGTCAAGTGATTTTTGAAGGCCCCACTCTGGGGCAGCAGGAAGACCAAGAGCGGAAGAGGTT
GTGTAAGGCCATGACCCTGTGCATCTGCTACGCGGCCAGCATCGGGGGCACCGCCACCCTGACCGGGA
CGGGACCCAACGTGGTGCTCCTGGGCCAGATGAACGAGTTGTTTCCTGACAGCAAGGACCTCGTGAAC
TTTGCTTCCTGGTTTGCATTTGCCTTTCCCAACATGCTGGTGATGCTGCTGTTCGCCTGGCTGTGGCT
CCAGTTTGTTTACATGAGATTCAATTTTAAAAAGTCCTGGGGCTGCGGGCTAGAGAGCAAGAAAAACG
AGAAGGCTGCCCTCAAGGTGCTGCAGGAGGAGTACCGGAAGTTGGGGCCCTTGTCCTTCGCGGAGATC
AACGTGCTGATCTGCTTCTTCCTGCTGGTCATCCTGTGGTTCTCCCGAGACCCCGGCTTCATGCCCGG
CTGGCTGACTGTTGCCTGGGTGGAGGGTGAGACAAAGTATGTCTCCGATGCCACTGTGGCCATCTTTG
TGGCCACCCTGCTATTCATTGTGCCTTCACAGAAGCCCAAGTTTAACTTCCGCAGCCAGACTGAGGAA
GAAAGGAAAACTCCATTTTATCCCCCTCCCCTGCTGGATTGGAAGGTAACCCAGGAGAAAGTGCCCTG
GGGCATCGTGCTGCTACTAGGGGGCGGATTTGCTCTGGCTAAAGGATCCGAGGCCTCGGGGCTGTCCG
TGTGGATGGGGAAGCAGATGGAGCCCTTGCACGCAGTGCCCCCGGCAGCCATCACCTTGATCTTGTCC
TTGCTCGTTGCCGTGTTCACTGAGTGCACAAGCAACGTGGCCACCACCACCTTGTTCCTGCCCATCTT
TGCCTCCATGGTGAAAACAGGAGTCATAATGAACATAATTGGAGTCTTCTGTGTGTTTTTGGCTGTCA
ACACCTGGGGACGGGCCATATTTGACTTGGATCATTTCCCTGACTGGGCTAATGTGACACATATTGAG
ACTTAGGAAGAGCCACAAGACCACACACATAGCCCTTACCCT

NOV67e, CG57758-04 Protein Sequence	SEQ ID NO: 1002	522 aa	MW at 58109.6kD

MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKFVRCAYVIILMAIYWCTEVIPLAVTSLMPVLLFP
LFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLWVGAKPARLMLGFMGVTALLS
MWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVDKGKAKELPGSQVIFEGPTLGQQEDQERKRL
CKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLVNFASWFAFAFPNMLVMLLFAWLWL
QFVYMRFNFKKSWGCGLESKKNEKAALKVLQEEYRKLGPLSFAEINVLICFFLLVILWFSRDPGFMPG
WLTVAWVEGETKYVSDATVAIFVATLLFIVPSQKPKFNFRSQTEEERKTPFYPPPLLDWKVTQEKVPW
GIVLLLGGGFALAKGSEASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIF
ASMVKTGVIMNIIGVFCVFLAVNTWGRAIFDLDHFPDWANVTHIET

NOV67f, CG57758-05 DNA Sequence	SEQ ID NO: 1003	1781 bp
	ORF Start: ATG at 2	ORF Stop: TGA at 1550

GATGGCCTCGGCGCTGAGCTATGTCTCCAAGTTCAAGTCCTTCGTGATCTTGTTCGTCACCCCGCTCC
TGCTGCTGCCACTCGTCATTCTGATGCCCGCCAAGTTTGTCAGGTGTGCCTACGTCATCATCCTCATG
GCCATTTACTGGTGCACAGAAGTCATCCCTCTGGCTGTCACCTCTCTCATGCCTGTCTTGCTTTTCCC

ACTCTTCCAGATTCTGGACTCCAGGCAGGTGTGTGTCCAGTACATGAAGGACACCAACATGCTGTTCC
TGGGCGGCCTCATCGTGGCCGTGGCTGTGGAGCGCTGGAACCTGCACAAGAGGATCGCCCTGCGCACG
CTCCTCTGGGTGGGGGCCAAGCCTGCACGGCTGATGCTGGGCTTCATGGGCGTCACAGCCCTCCTGTC
CATGTGGATCAGTAACACGGCAACCACGGCCATGATGGTGCCCATCGTGGAGGCCATATTGCAGCAGA
TGGAAGCCACAAGCGCAGCCACCGAGGCCGGCCTGGAGCTGGTGGACAAGGGCAAGGCCAAGGAGCTG
CCAGGGAGTCAAGTGATTTTTGAAGGCCCCACTCTGGGGCAGCAGGAAGACCAAGAGCGGAAGAGGTT
GTGTAAGGCCATGACCCTGTGCATCTGCTACGCGGCCAGCATCGGGGGCACCGCCACCCTGACCGGGA
CGGGACCCAACGTGGTGCTCCTGGGCCAGATGAACGAGTTGTTTCCTGACAGCAAGGACCTCGTGAAC
TTTGCTTCCTGGTTTGCATTTGCCTTTCCCAACATGCTGGTGATGCTGCTGTTCGCCTGGCTGTGGCT
CCAGTTTGTTTACATGAGATTCAATTTTAAAAAGTCCTGGGGCTGCGGGCTAGAGAGCAAGAAAAACG
AGAAGGCTGCCCTCAAGGTGCTGCAGGAGGAGTACCGGAAGTTGGGGCCCTTGTCCTTCGCGGAGATC
AACGTGCTGATCTGCTTCTTCCTGCTGGTCATCCTGTGGTTCTCCCGAGACCCCGGCTTCATGCCCGG
CTGGCTGACTGTTGCCTGGGTGGAGGGTGAGACAAAGTATGTCTCCGATGCCACTGTGGCCATCTTTG
TGGCCACCCTGCTATTCATTGTGCCTTCACAGAAGCCCAAGTTTAACTTCCGCAGCCAGACTGAGGAA
GAAAGGAAAACTCCATTTTATCCCCCTCCCCTGCTGGATTGGAAGGTAACCCAGGAGAAAGTGCCCTG
GGGCATCGTGCTGCTACTAGGGGGCGGATTTGCTCTGGCTAAAGGATCCGAGGCCTCGGGGCTGTCCG
TGTGGATGGGGAAGCAGATGGAGCCCTTGCACGCAGTGCCCCCGGCAGCCATCACCTTGATCTTGTCC
TTGCTCGTTGCCGTGTTCACTGAGTGCACAAGCAACGTGGCCACCACCACCTTGTTCCTGCCCATCTT
TGCCTCCATGAATCACGTCCCCAAGAGCTTCTGTGTTCTGTACGGTGATGTTGCAGTGCTGTCTTTCC
GCAGTCTCGCTCCATCGGCCTCAATCCGCTGTACATCATGCTGCCCTGTACCCTGAGTGCCTCCTTTG
CCTTCATGTTGCCTGTGGCCACCCCTCCAAATGCCATCGTGTTCACCTATGGGCACCTCAAGGTTGCT
GACATGGTGAAAACAGGAGTCATAATGAACATAATTGGAGTCTTCTGTGTGTTTTTGGCTGTCAACAC
CTGGGGACGGGCCATATTTGACTTGGATCATTTCCCTGACTGGGCTAATGTGACACATATTGAGACTT
AGGAAGAGCCACA

NOV67f, CG57758-05 Protein Sequence	SEQ ID NO: 1004	516 aa	MW at 57173.5kD

MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKFVRCAYVIILMAIYWCTEVIPLAVTSLMPVLLFP
LFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLWVGAKPARLMLGFMGVTALLS
MWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVDKGKAKELPGSQVIFEGPTLGQQEDQERKRL
CKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLVNFASWFAFAFPNMLVMLLFAWLWL
QFVYMRFNFKKSWGCGLESKKNEKAALKVLQEEYRKLGPLSFAEINVLICFFLLVILWFSRDPGFMPG
WLTVAWVEGETKYVSDATVAIFVATLLFIVPSQKPKFNFRSQTEEERKTPFYPPPLLDWKVTQEKVPW
GIVLLLGGGFALAKGSEASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIF
ASMNHVPKSFCVLYGDVAVLSFRSLAPSASIRCTSCCPVP

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 67B.

Table 67B. Comparison of the NOV67 protein sequences.

```
NOV67a  -----MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKFVRCAYVIILMAIYWCTEVIP
NOV67b  TRSPTMASALSYVSKFKSFVILFVTPLLLLPLVILMPAKFVRCAYVIILMAIYWCTEVIP
NOV67c  -----MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKVS-CAYVIILMAIYWCTEVIP
NOV67d  -----MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKVSCCAYVIILMAIYWCTEVIP
NOV67e  -----MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKFVRCAYVIILMAIYWCTEVIP
NOV67f  -----MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKFVRCAYVIILMAIYWCTEVIP

NOV67a  LAVTSLMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLL
NOV67b  LAVTSLMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLL
NOV67c  LAVTSLMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLL
NOV67d  LAVTSLMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLL
NOV67e  LAVTSLMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLL
NOV67f  LAVTSLMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLL

NOV67a  WVGAKPARLMLGFMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVDK
```

```
NOV67b  WVGAKPARLMLGFMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVDK
NOV67c  WVGAKPARLMLGFMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVDK
NOV67d  WVGAKPARLMLGFMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLEGQGT
NOV67e  WVGAKPARLMLGFMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVDK
NOV67f  WVGAKPARLMLGFMGVTALLSMWISNTATTAMMVPIVEAILQQMEATSAATEAGLELVDK

NOV67a  GK----------------------------------------AKELPGSQVIFEGPTLG
NOV67b  GK----------------------------------------AKELPGSQVIFEGPTLG
NOV67c  GK----------------------------------------AKELPGSQVIFEGPTLG
NOV67d  TINNLNALEDDTVKAVLGGKCVAIISTYVKKVEKLQINNLMTPLKKLEKQEQQDLGPGIR
NOV67e  GK----------------------------------------AKELPGSQVIFEGPTLG
NOV67f  GK----------------------------------------AKELPGSQVIFEGPTLG

NOV67a  -Q-----QEDQERKRLCKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLV
NOV67b  -Q-----QEDQERKRLCKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLV
NOV67c  -Q-----QEDQERKRLCKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLV
NOV67d  PQDSAQCQEDQERKRLCKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLV
NOV67e  -Q-----QEDQERKRLCKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLV
NOV67f  -Q-----QEDQERKRLCKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLV

NOV67a  NFASWFAFAFPNMLVMLLFAWLWLQFVYMRFNFKKSWGCGLESKKNEKAALKVLQEEYRK
NOV67b  NFASWFAFAFPNMLVMLLFAWLWLQFVYMRFNFKKSWGCGLESKKNEKAALKVLQEEYRK
NOV67c  NFASWFAFAFPNMLVMLLFAWLWLQFVYMFSSFKKSWGCGLESKKNEKAALKVLQEEYRK
NOV67d  NFASWFAFAFPNMLVMLLFAWLWLQFVYMFSSFKKSWGCGLESKKNEKAALKVLQEEYRK
NOV67e  NFASWFAFAFPNMLVMLLFAWLWLQFVYMRFNFKKSWGCGLESKKNEKAALKVLQEEYRK
NOV67f  NFASWFAFAFPNMLVMLLFAWLWLQFVYMRFNFKKSWGCGLESKKNEKAALKVLQEEYRK

NOV67a  LGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVEGETKYVSDATVAIFVATLLF
NOV67b  LGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVEGETKYVSDATVAIFVATLLF
NOV67c  LGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVEGETKYVSDATVAIFVATLLF
NOV67d  LGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVEGETKSVSDATVAIFVATLLF
NOV67e  LGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVEGETKYVSDATVAIFVATLLF
NOV67f  LGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVEGETKYVSDATVAIFVATLLF

NOV67a  IVPSQKPKFNFRSQTEEERK-TPFYPPPLLDWKVTQEKVPWGIVLLLGGGFALAKGSEAS
NOV67b  IVPSQKPKFNFRSQTEEERK-TPFYPPPLLDWKVTQEKVPWGIVLLLGGGFALAKGSEAS
NOV67c  IVPSQKPKFNFRSQTEEGKSPVLIAPPPLLDWKVTQEKVPWGIVLLLGGGFALAKGSEAS
NOV67d  IVPSQKPKFNFRSQTEEGKSPVLIAPPPLLDWKVTQEKVPWGIVLLLGGGFALAKGSEAS
NOV67e  IVPSQKPKFNFRSQTEEERK-TPFYPPPLLDWKVTQEKVPWGIVLLLGGGFALAKGSEAS
NOV67f  IVPSQKPKFNFRSQTEEERK-TPFYPPPLLDWKVTQEKVPWGIVLLLGGGFALAKGSEAS

NOV67a  GLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIFASMSR---SIG
NOV67b  GLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIFASMSR---SIG
NOV67c  GLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIFASMSR---SIG
NOV67d  GLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIFASMSR---SIG
NOV67e  GLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIFASMVKTGVIMN
NOV67f  GLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIFASMN------H

NOV67a  LNPLYIMLPCTLSASFAFMLPVATPPNAIVFTYGHLKVADMVKTGVIMNIIGVFCVFLAV
NOV67b  LNPLYIMLPCTLSASFAFMLPVATPPNAIVFTYGHLKVADMVKTGVIMNIIGVFCVFLAV
NOV67c  LNPLYIMLPCTLSASFAFMLPVATPPNAIVFTYGHLKVADMVKTGVIMNIIGVFCVFLAV
NOV67d  LNPLYIMLPCTLSASFAFMLPVATPPNAIVFTYGHLKVADMVKTGVIMNIIGVFCVFLAV
NOV67e  IIGVFCVFLAVNTWGRAIFDLDHFPDWANVTHIET-----------------------
NOV67f  VPKSFCVLYGD----VAVLSFRSLAPSASIRCTSCCPVP--------------------
```

```
NOV67a  NTWGRAIFDLDHFPDWANVTHIET---
NOV67b  NTWGRAIFDLDHFPDWANVTHIETLEG
NOV67c  NTWGRAIFDLDHFPDWANVTHIET---
NOV67d  NTWGRAIFDLDHFPDWANVTHIET---
NOV67e  --------------------------
NOV67f  --------------------------

NOV67a  (SEQ ID NO:  994)
NOV67b  (SEQ ID NO:  996)
NOV67c  (SEQ ID NO:  998)
NOV67d  (SEQ ID NO:  1000)
NOV67e  (SEQ ID NO:  1002)
NOV67f  (SEQ ID NO:  1004)
```

Further analysis of the NOV67a protein yielded the following properties shown in Table 67C.

Table 67C. Protein Sequence Properties NOV67a

SignalP analysis:	Cleavage site between residues 39 and 40

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 10;  pos.chg 1;  neg.chg 0
      H-region:  length 1;  peak value   5.97
      PSG score:   1.57

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -2.36
      possible cleavage site: between 30 and 31

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:  11
      INTEGRAL    Likelihood =-12.21   Transmembrane   14 -  30
      INTEGRAL    Likelihood = -1.81   Transmembrane   35 -  51
      INTEGRAL    Likelihood = -4.99   Transmembrane   53 -  69
      INTEGRAL    Likelihood = -1.75   Transmembrane  124 - 140
      INTEGRAL    Likelihood = -3.72   Transmembrane  261 - 277
      INTEGRAL    Likelihood = -8.01   Transmembrane  312 - 328
      INTEGRAL    Likelihood = -6.26   Transmembrane  354 - 370
      INTEGRAL    Likelihood =  0.21   Transmembrane  405 - 421
      INTEGRAL    Likelihood = -7.11   Transmembrane  443 - 459
      INTEGRAL    Likelihood = -1.70   Transmembrane  490 - 506
      INTEGRAL    Likelihood = -7.43   Transmembrane  528 - 544
      PERIPHERAL  Likelihood =  0.74 (at 89)
      ALOM score: -12.21  (number of TMSs: 11)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 21
```

```
      Charge difference: -1.0   C( 2.0) - N( 3.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75):  1.18
      Hyd Moment(95):  3.95    G content:       0
      D/E content:     1       S/T content:     6
      Score: -3.21

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 47  VRC|AY

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.7%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
```

```
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       22.2 %: mitochondrial
       11.1 %: nuclear

>> prediction for CG57758-03 is end (k=9)
```

A search of the NOV67a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 67D.

5

Table 67D. Geneseq Results for NOV67a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV67a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU79946	Human transporter protein sequence - Homo sapiens, 568 aa. [US2002019028-A1, 14-FEB-2002]	1..568 1..568	566/568 (99%) 566/568 (99%)	0.0
AAE21181	Human TRICH-25 protein - Homo sapiens, 539 aa. [WO200212340-A2, 14-FEB-2002]	1..568 1..539	531/576 (92%) 531/576 (92%)	0.0
ABB82951	Human SLC13A related protein (GenBank Identifier No. GI#4506979) - Homo sapiens, 592 aa. [WO200298468-A1, 12-DEC-2002]	1..561 1..572	317/579 (54%) 423/579 (72%)	0.0
ABB82950	Human SLC13A related protein (GenBank Identifier No. GI#2499523) - Homo sapiens, 592 aa. [WO200298468-A1, 12-DEC-2002]	1..561 1..572	317/579 (54%) 423/579 (72%)	0.0
ABB82952	Human SLC13A related protein (GenBank Identifier No. GI#13653602) - Homo sapiens, 602 aa. [WO200298468-A1, 12-DEC-2002]	1..563 4..582	284/591 (48%) 387/591 (65%)	e-156

In a BLAST search of public sequence databases, the NOV67a protein was found to have homology to the proteins shown in the BLASTP data in Table 67E.

Table 67E. Public BLASTP Results for NOV67a				
Protein Accession Number	**Protein/Organism/Length**	**NOV67a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAN86530	Na+-coupled citrate transporter protein - Homo sapiens (Human), 568 aa.	1..568 1..568	568/568 (100%) 568/568 (100%)	0.0
Q8CJ44	Sodium-coupled citrate transporter - Rattus norvegicus (Rat), 572 aa.	1..568 1..572	442/572 (77%) 502/572 (87%)	0.0
AAH44437	Similar to solute carrier family 13, member 2 - Brachydanio rerio (Zebrafish) (Danio rerio), 613 aa.	5..564 10..605	331/598 (55%) 437/598 (72%)	0.0
AAO27449	Sodium dicarboxylate co-transporter - Didelphis marsupialis virginiana (North American opossum), 605 aa.	11..563 11..587	318/582 (54%) 422/582 (71%)	0.0
Q13183	Solute carrier family 13, member 2 (Renal sodium/dicarboxylate cotransporter) (Na(+)/dicarboxylate cotransporter 1) (NaDC-1) - Homo sapiens (Human), 592 aa.	1..561 1..572	317/579 (54%) 423/579 (72%)	0.0

PFam analysis indicates that the NOV67a protein contains the domains shown in the Table 67F.

Table 67F. Domain Analysis of NOV67a			
Pfam Domain	**NOV67a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
DcuC	30..548	88/593 (15%) 324/593 (55%)	0.73
Na_sulph_symp	6..554	159/603 (26%) 423/603 (70%)	2e-143

Example 68.

The NOV68 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 68A.

Table 68A. NOV68 Sequence Analysis

NOV68a, CG58504-01 DNA Sequence	SEQ ID NO: 1005	5115 bp
	ORF Start: ATG at 327	ORF Stop: TAA at 5106

GAATTCCGGGAGCGGGCGGGCTGCGAGGCCGCGGGGCATGCGGGAGGCGGAGGGGTGGGACCGGGTGG
CTGCGCCCATTCCACACCCGCCGAAAGCGGACACTGTCAGCTGAATCACTCCCCTTTTAGGAGGAGGG
AGGGGGAAAAGGTGTCTAGCTAATTTCTGCTTAAAAAAGCACAGGAGATCGCGGGTCAGCTTTGCAGT
CGCTGCCTTCTCGCGCCTGACCATGCACCCCTGCATCTTCCTGCTGGGCACAGGCGAGCGCTTTATTT
CTGGAGCTGAGGGCTAAAACTTTTTTCACTTTTCTTCTCCTCAACATCTGAATCATGCCATGTGCCCA
GAGGAGCTGGCTTGCAAACCTTTCCGTGGTGGCTCAGCTCCTTAACTTTGGGGCGCTTTGCTATGGGA
GACAGCCTCAGCCAGGCCCGGTTCGCTTCCCGGACAGGAGGCAAGAGCATTTTATCAAGGGCCTGCCA
GAATACCACGTGGTGGGTCCAGTCCGAGTAGATGCCAGTGGGCATTTTTTGTCATATGGCTTGCACTA
TCCCATCACGAGCAGCAGGAGGAAGAGAGATTTGGATGGCTCAGAGGACTGGGTGTACTACAGAATTT
CTCACGAGGAGAAGGACCTGTTTTTTAACTTGACGGTCAATCAAGGATTTCTTTCCAATAGCTACATC
ATGGAGAAGAGATATGGGAACCTCTCCCATGTTAAGATGATGGCTTCCTCTGCCCCCCTCTGCCATCT
CAGTGGCACGGTTCTACAGCAGGGCACCAGAGTTGGGACGGCAGCCCTCAGTGCCTGCCATGGACTGA
CTGGATTTTTCCAACTACCACATGGAGACTTTTTCATTGAACCCGTGAAGAAGCATCCACTGGTTGAG
GGAGGGTACCACCCGCACATCGTTTACAGGAGGCAGAAAGTTCCAGAAACCAAGGAGCCAACCTGTGG
ATTAAAGGACAGTGTTAACATCTCCCAGAAGCAAGAGCTATGGCGGGAGAAGTGGGAGAGGCACAACT
TGCCAAGCAGAAGCCTCTCTCGGCGTTCCATCAGCAAGGAGAGATGGGTGGAGACACTGGTGGTGGCC
GACACAAAGATGATTGAATACCATGGGAGTGAGAATGTGGAGTCCTACATCCTCACCATCATGAACAT
GGTCACTGGGTTGTTCCATAACCCAAGCATTGGCAATGCAATTCACATTGTTGTGGTTCGGCTCATTC
TACTCGAAGAAGAAGAGCAAGGACTGAAAATAGTTCACCATGCAGAAAAGACACTGTCTAGCTTCTGC
AAGTGGCAGAAGAGTATCAATCCCAAGAGTGACCTCAATCCTGTTCATCACGACGTGGCTGTCCTTCT
CACCAGAAAGGACATCTGTGCTGGTTTCAATCGCCCCTGCGAGACCCTGGGCCTGTCTCACCTTTCAG
GAATGTGTCAGCCTCACCGCAGTTGTAACATCAATGAAGATTCGGGACTCCCTCTGGCTTTCACAATT
GCCCATGAGCTAGGACACAGCTTCGGCATCCAGCATGATGGGAAAGAAAATGACTGTGAGCCTGTGGG
CAGACATCCGTACATCATGTCCCGCCAGCTCCAGTACGATCCCACTCCGCTGACATGGTCCAAGTGCA
GCGAGGAGTACATCACCCGCTTCTTGGACCGAGGCTGGGGGTTCTGTCTTGATGACATACCTAAAAAG
AAAGGCTTGAAGTCCAAGGTCATTGCCCCCGGAGTGATCTATGATGTTCACCACCAGTGCCAGCTACA
ATATGGACCCAATGCTACCTTCTGCCAGGAAGTAGAAAACGTCTGCCAGACACTGTGGTGCTCCGTGA
AGGGCTTTTGTCGCTCTAAGCTGGACGCTGCTGCAGATGGAACTCAATGTGGTGAGAAGAAGTGGTGT
ATGGCAGGCAAGTGCATCACAGTGGGGAAGAAACCAGAGAGCATTCCTGGAGGCTGGGGCCGCTGGTC
ACCCTGGTCCCACTGTTCCAGGACCTGTGGGGCTGGAGTCCAGAGCGCAGAGAGGCTCTGCAACAACC
CCGAGCCAAAGTTTGGAGGGAAATATTGCACTGGAGAAAGAAAACGCTATCGCTTGTGCAACGTCCAC
CCCTGTCGCTCAGAGGCACCAACATTTCGGCAGATGCAGTGCAGTGAATTTGACACTGTTCCCTACAA
GAATGAACTCTACCACTGGTTTCCCATTTTTAACCCAGCACATCCTTGTGAGCTCTACTGCCGACCCA
TAGATGGCCAGTTTTCTGAGAAAATGCTGGATGCTGTCATTGATGGTACCCCTTGCTTTGAAGGCGGC
AACAGCAGAAATGTCTGTATTAATGGCATATGTAAGATGGTTGGCTGTGACTATGAGATCGATTCCAA
TGCCACCGAGGATCGCTGCGGTGTGTGCCTGGGAGATGGCTCTTCCTGCCAGACTGTGAGAAAGATGT
TTAAGCAGAAGGAAGGATCTGGTTATGTTGACATTGGGCTCATTCCAAAAGGAGCAAGGGACATAAGA
GTGATGGAAATTGAGGGAGCTGGAAACTTCCTGGCCATCAGGAGTGAAGATCCTGAAAAATATTACCT
GAATGGAGGGTTTATTATCCAGTGGAACGGGAACTATAAGCTGGCAGGGACTGTCTTTCAGTATGACA
GGAAAGGAGACCTGGAAAAGCTGATGGCCACAGGTCCCACCAATGAGTCTGTGTGGATCCAGCTTCTA
TTCCAGGTGACTAACCCTGGCATCAAGTATGAGTACACAATCCAGAAAGATGGCCTTGACAATGATGT
TGAGCAGATGTACTTCTGGCAGTACGGCCACTGGACAGAGTGCAGTGTGACCTGCGGGACAGGTATCC
GCCGCCAAACTGCCCATTGCATAAAGAAGGGCCGCGGGATGGTGAAAGCTACATTCTGTGACCCAGAA
ACACAGCCCAATGGGAGACAGAAGAAGTGCCATGAAAAGGCTTGTCCACCCAGGTGGTGGGCAGGGGA
GTGGGAAGCATGCTCGGCGACATGCGGGCCCCACGGGGAGAAGAAGCGAACCGTGCTGTGCATCCAGA
CCATGGTCTCTGACGAGCAGGCTCTCCCGCCCACAGACTGCCAGCACCTGCTGAAGCCCAAGACCCTC
CTTTCCTGCAACAGAGACATCCTGTGCCCCTCGGACTGGACAGTGGGCAACTGGAGTGAGTGTTCTGT
TTCCTGTGGTGGTGGAGTGCGGATTCGCAGTGTCACATGTGCCAAGAACCATGATGAACCTTGCGATG
TGACAAGGAAACCCAACAGCCGAGCTCTGTGTGGCCTCCAGCAATGCCCTTCTAGCCGGAGAGTTCTG

AAACCAAACAAAGGCACTATTTCCAATGGAAAAAACCCACCAACACTAAAGCCCGTCCCTCCACCTAC
ATCCAGGCCCAGAATGCTGACCACACCCACAGGGCCTGAGTCTATGAGCACAAGCACTCCAGCAATCA

GCAGCCCTAGTCCTACCACAGCCTCCAAAGAAGGAGACCTGGGTGGGAAACAGTGGCAAGATAGCTCA
ACCCAACCTGAGCTGAGCTCTCGCTATCTCATTTCCACTGGAAGCACTTCCCAGCCCATCCTCACTTC
CCAATCCTTGAGCATTCAGCCAAGTGAGGAAAATGTTTCCAGTTCAGATACTGGTCCTACCTCGGAGG
GAGGCCTTGTAGCTACAACAACAAGTGGTTCTGGCTTGTCATCTTCCCGCAACCCTATCACTTGGCCT
GTGACTCCATTTTACAATACCTTGACCAAAGGTCCAGAAATGGAGATTCACAGTGGCTCAGGGGAAGA
AAGAGAACAGCCTGAGGACAAAGATGAAAGCAATCCTGTAATATGGACCAAGATCAGAGTACCTGGAA
ATGACGCTCCAGTGGAAAGTACAGAAATGCCACTTGCACCTCCACTAACACCAGATCTCAGCAGGGAG
TCCTGGTGGCCACCCTTCAGCACAGTAATGGAAGGACTGCTCCCCAGCCAAAGGCCCACTACTTCCGA
AACTGGGACACCCAGAGTTGAGGGGATGGTTACTGAAAAGCCAGCCAACACTCTGCTCCCTCTGGGAG
GAGACCACCAGCCAGAACCCTCAGGAAAGACGGCAAACCGTAACCACCTGAAACTTCCAAACAACATG
AACCAAACAAAAAGTTCTGAACCAGTCCTGACTGAGGAGGATGCAACAAGTCTGATTACTGAGGGCTT
TTTGCTAAATGCCTCCAATTACAAGCAGCTCACAAACGGCCACGGCTCTGCACACTGGATCGTCGGAA
ACTGGAGCGAGTGCTCCACCACATGTGGCCTGGGGGCCTACTGGAAAAGGGTGGAGTGCACCACCCAG
ATGGATTCTGACTGTGCGGCCATCCAGAGACCTGACCCTGCAAAAAGATGCCACCTCCGTCCCTGTGC
TGGCTGGAAAGTGGGAAACTGGAGCAAGTGCTCCAGAAACTGCAGTGGGGGCTTCAAGATACGCGAGA
TTCAGTGCGTGGACAGCCGGGACCACCGGAACCTGAGGCCATTTCACTGCCAGTTCCTGGCCGGCATT
CCTCCCCCATTGAGCATGAGCTGTAACCCGGAGCCCTGTGAGGCGTGGCAGGTGGAGCCTTGGAGCCA
GTGCTCCAGGTCCTGTGGAGGTGGAGTTCAGGAGAGAGGAGTGTTCTGTCCAGGAGGCCTCTGTGATT
GGACAAAAAGACCCACATCCACCATGTCTTGCAATGAGCACCTGTGCTGTCACTGGGCCACTGGGAAC
TGGGACCTGTGTTCCACTTCCTGTGGAGGTGGCTTTCAGAAGAGGATTGTCCAATGTGTGCCCTCAGA
GGGCAATAAAACTGAAGACCAAGACCAATGTCTATGTGATCACAAACCCAGACCTCCAGAATTCAAAA
AATGCAACCAGCAGGCCTGCAAGAAAAGTGCCGATTTACTTTGCACTAAGGACAAACTGTCAGCCAGT
TTCTGCCAGACACTGAAAGCCATGAAGAAATGTTCTGTGCCCACCGTGAGGGCTGAGTGCTGCTTCTC
GTGTCCCCAGACACACATCACACACCCAAAGGCAAAGAAGGCAACGGTTGCTCCAAAAGTCAAAAG
AACTC**TAA**GCCCAAA

NOV68a, CG58504-01 Protein Sequence	SEQ ID NO: 1006	1593 aa	MW at 177543.9kD

MPCAQRSWLANLSVVAQLLNFGALCYGRQPQPGPVRFPDRRQEHFIKGLPEYHVVGPVRVDASGHFLS
YGLHYPITSSRRKRDLDGSEDWVYYRISHEEKDLFFNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSA
PLCHLSGTVLQQGTRVGTAALSACHGLTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRRQKVPETK
EPTCGLKDSVNISQKQELWREKWERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGSENVESYIL
TIMNMVTGLFHNPSIGNAIHIVVVRLILLEEEEQGLKIVHHAEKTLSSFCKWQKSINPKSDLNPVHHD
VAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSCNINEDSGLPLAFTIAHELGHSFGIQHDGKEND
CEPVGRHPYIMSRQLQYDPTPLTWSKCSEEYITRFLDRGWGFCLDDIPKKKGLKSKVIAPGVIYDVHH
QCQLQYGPNATFCQEVENVCQTLWCSVKGFCRSKLDAAADGTQCGEKKWCMAGKCITVGKKPESIPGG
WGRWSPWSHCSRTCGAGVQSAERLCNNPEPKFGGKYCTGERKRYRLCNVHPCRSEAPTFRQMQCSEFD
TVPYKNELYHWFPIFNPAHPCELYCRPIDGQFSEKMLDAVIDGTPCFEGGNSRNVCINGICKMVGCDY
EIDSNATEDRCGVCLGDGSSCQTVRKMFKQKEGSGYVDIGLIPKGARDIRVMEIEGAGNFLAIRSEDP
EKYYLNGGFIIQWNGNYKLAGTVFQYDRKGDLEKLMATGPTNESVWIQLLFQVTNPGIKYEYTIQKDG
LDNDVEQMYFWQYGHWTECSVTCGTGIRRQTAHCIKKGRGMVKATFCDPETQPNGRQKKCHEKACPPR
WWAGEWEACSATCGPHGEKKRTVLCIQTMVSDEQALPPTDCQHLLKPKTLLSCNRDILCPSDWTVGNW
SECSVSCGGGVRIRSVTCAKNHDEPCDVTRKPNSRALCGLQQCPSSRRVLKPNKGTISNGKNPPTLKP
VPPPTSRPRMLTTPTGPESMSTSTPAISSPSPTTASKEGDLGGKQWQDSSTQPELSSRYLISTGSTSQ
PILTSQSLSIQPSEENVSSSDTGPTSEGGLVATTTSGSGLSSSRNPITWPVTPFYNTLTKGPEMEIHS
GSGEEREQPEDKDESNPVIWTKIRVPGNDAPVESTEMPLAPPLTPDLSRESWWPPFSTVMEGLLPSQR
PTTSETGTPRVEGMVTEKPANTLLPLGGDHQPEPSGKTANRNHLKLPNNMNQTKSSEPVLTEEDATSL
ITEGFLLNASNYKQLTNGHGSAHWIVGNWSECSTTCGLGAYWKRVECTTQMDSDCAAIQRPDPAKRCH
LRPCAGWKVGNWSKCSRNCSGGFKIREIQCVDSRDHRNLRPFHCQFLAGIPPPLSMSCNPEPCEAWQV
EPWSQCSRSCGGGVQERGVFCPGGLCDWTKRPTSTMSCNEHLCCHWATGNWDLCSTSCGGGFQKRIVQ
CVPSEGNKTEDQDQCLCDHKPRPPEFKKCNQQACKKSADLLCTKDKLSASFCQTLKAMKKCSVPTVRA
ECCFSCPQTHITHTQRQRRQRLLQKSKEL

NOV68b, 169648376 DNA Sequence	SEQ ID NO: 1007	1068 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTAACTTGACGGTCAATCAAGGATTTCTTTCCAATAGCTACATCATGGAGAAGAGATATGGGAA
CCTCTCCCATGTTAAGATGATGGCTTCCTCTGCCCCCCTCTGCCATCTCAGTGGCACGGTTCTACAGC
AGGGCACCAGAGTTGGGATGGCAGCCCTCAGTGCCTGCCATGGACTGACTGGATTTTTCCAACTACCA
CATGGAGACTTTTTCATTGAACCCGTGAAGAAGCATCCACTGGTTGAGGGAGGGTACCACCCGCACAT
CGTTTACAGGAGGCAGAAAGTTCCAGAAACCAAGGAGCCAACCTGTGGATTAAAGGACAGTGTTAACA
TCTCCCAGAAGCAAGAGCTATGGCGGGAGAAGTGGGAGAGGCACAACTTGCCAAGCAGAAGCCTCTCT
CGGCGTTCCATCAGCAAGGAGAGATGGGTGGAGACACTGGTGGTGGCCGACACAAAGATGATTGAATA

CCATGGGAGTGAGAATGTGGAGTCCTGCATCCTCACCATCATGAACATGGTCACTGGGTTGTTCCATA
ACCCAAGCATTGGCAATGCAATTCACATTGTTGTGGTTCGGCTCATTCTACTCGAAGAAGAAGAGCAA
GGACTGAAAATAGTTCACCATGCAGAAAAGACACTGTCTAGCTTCTGCAAGTGGCAGAAGAGTATCAA
TCCCAAGAGTGACCTCAATCCTGTTCATCACGACGTGGCTGTCCTTCTCACCAGAAAGGACATCTGTG
CTGGTTTCAATCGCCCCTGCGAGACCCTGGGCCTGTCTCACCTTTCAGGAATGTGTCAGCCTCACCGC
AGTTGTAACATCAATGAAGATTCGGGACTCCCTCTGGCTTTCACAATTGCCCATGAGCTAGGACACAG
CTTCGGCATCCAGCATGATGGGAAAGAAAATGACTGTGAGCCTGTGGGCAGACATCCGTACATCATGT
CCCGCCAGCTCCAGTACGATCCCACTCCGCTGACATGGTCCAAGTGCAGCGAGGAGTACATCACCCGC
TTCTTGGACCGAGGCTGGGGGTTCTGTCTTGATGACATACCTCTCGAG

NOV68b, 169648376 Protein Sequence	SEQ ID NO: 1008	356 aa	MW at 40336.9kD

KLNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGMAALSACHGLTGFFQLP
HGDFFIEPVKKHPLVEGGYHPHIVYRRQKVPETKEPTCGLKDSVNISQKQELWREKWERHNLPSRSLS
RRSISKERWVETLVVADTKMIEYHGSENVESCILTIMNMVTGLFHNPSIGNAIHIVVVRLILLEEEEQ
GLKIVHHAEKTLSSFCKWQKSINPKSDLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHR
SCNINEDSGLPLAFTIAHELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPTPLTWSKCSEEYITR
FLDRGWGFCLDDIPLE

NOV68c, 169648388 DNA Sequence	SEQ ID NO: 1009	1068 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTAACTTGACGGTCAATCAAGGATTTCTTTCCAATAGCTACATCATGGAGAAGAGATATGGGAA
CCTCTCCCATGTTAAGATGATGGCTTCCTCTGCCCCCCTCTGCCATCTCAGTGGCACGGTTCTACAGC
AGGGCACCAGAGTTGGGACGGCAGCCCTCAGTGCCTGCCATGGACTGACTGGATTTTTCCAACTACCA
CATGGAGACTTTTTCATTGAACCCGTGAAGAAGCATCCACTGGTTGAGGGAGGGTACCACCCGCACAT
CGTTTACAGGAGGCAGAAAGTTCCAGAAACCAAGGAGCCAACCTGTGGATTAAAGGACAGTGTTAACA
TCTCCCAGAAGCAAGAGCTATGGCGGGAGAAGTGGGAGAGGCACAACTTGCCAAGCAGAAGCCTCTCT
CGGCGTTCCATCAGCAAGGAGAGATGGGTGGAGACACTGGTGGTGGCCGACACAAAGATGATTGAATA
CCATGGGAGTGAGAATGTGGAGTCCTACATCCTCACCATCATGAACATGATCACTGGGTTGTTCCATA
ACCCAAGCATTGGCAATGCAATTCACATTGTTGTGGTTCGGCTCATTCTACTCGAAGAAGAAGAGCAA
GGACTGAAAATAGTTCACCATGCAGAAAAGACACTGTCTAGCTTCTGCAAGTGGCAGAAGAGTATCAA
TCCCAAGAGTGACCTCAATCCTGTTCATCACGACGTGGCTGTCCTTCTCACCAGAAAGGACATCTGTG
CTGGTTTCAATCGCCCCTGCGAGACCCTGGGCCTGTCTCACCTTTCAGGAATGTGTCAGCCTCACCGC
AGTTGTAACATCAATGAAGATTCGGGACTCCCTCTGGCTTTCACAATTGCCCATGAGCTAGGACACAG
CTTCGGCATCCAGCATGATGGGAAAGAAAATGACTGTGAGCCTGTGGGCAGACATCCGTACATCATGT
CCCGCCAGCTCCAGTACGATCCCACTCCGCTGACATGGTCCAAGTGCAGCGAGGAGTACATCACCCGC
TTCTTGGACCGAGGCTGGGGGTTCTGTCTTGATGACATACCTCTCGAG

NOV68c, 169648388 Protein Sequence	SEQ ID NO: 1010	356 aa	MW at 40380.8kD

KLNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHGLTGFFQLP
HGDFFIEPVKKHPLVEGGYHPHIVYRRQKVPETKEPTCGLKDSVNISQKQELWREKWERHNLPSRSLS
RRSISKERWVETLVVADTKMIEYHGSENVESYILTIMNMITGLFHNPSIGNAIHIVVVRLILLEEEEQ
GLKIVHHAEKTLSSFCKWQKSINPKSDLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHR
SCNINEDSGLPLAFTIAHELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPTPLTWSKCSEEYITR
FLDRGWGFCLDDIPLE

NOV68d, 169648365 DNA Sequence	SEQ ID NO: 1011	1068 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTAACTTGACGGTCAATCAAGGATTTCTTTCCAATAGCTACATCATGGAGAAGAGATATGGGAA
CCTCTCCCATGTTAAGATGATGGCTTCCTCTGCCCCCCTCTGCCATCTCAGTGGCACGGTTCTACAGC
AGGGCACCAGAGTTGGGACGGCAGCCCTCAGTGCCTGCCATGGACTGACTGGATTTTTCCAACTACCA
CATGGAGACTTTTTCATTGAACCCGTGAAGAAGCATCCACTGGTTGAGGGAGGGTACCACCCGCACAT
CGTTTACAGGAGGCAGAAAGTTCCAGAAACCAAGGAGCCAACCTGTGGATTAAAGGACAGTGTTAACA
TCTCCCAGAAGCAAGAGCTATGGCGGGAGAAGTGGGAGAGGCACAACTTGCCAAGCAGAAGCCTCTCT
CGGCGTTCCATCAGCAAGGAGAGATGGGTGGAGACACTGGTGGTGGCCGACACAAAGATGATTGAATA
CCATGGGAGTGAGAATGTGGAGTCCTACATCCTCACCATCATGAACATGGTCACTGGGTTGTTCCATA
ACCCAAGCATTGGCAATGCAATTCACATTGTTGTGGTTCGGCTCATTCTACTCGAAGAAGAAGAGCAA
GGACTGAAAATAGTTCACCATGCAGAAAAGACACTGTCTAGCTTCTGCAAGTGGCAGAAGAGTATCAA
TCCCAAGAGTGACCTCAATCCTGTTCATCACGACGTGGCTGTCCTTCTCACCAGAAAGGACATCTGTG
CTGGTTTCAATCGCCCCTGCGAGACCCTGGGCCTGTCTCACCTTTCAGGAATGTGTCAGCCTCACCGC
AGTTGTAACATCAATGAAGATTCGGGACTCCCTCTGGCTTTCACAATTGCCCATGAGCTAGGACACAG

CTTCGGCATCCAGCATGATGGGAAAGAAAATGACTGTGAGCCTGTGGGCAGACATCCGTACATCATGT
CCCGCCAGCTCCAGTACGATCCCACTCCGCTGACATGGTCCAAGTGCAGCGAGGAGTACATCACCCGC
TTCTTGGACCGAGGCTGGGGGTTCTGTCTTGATGACATACCTCTCGAG

NOV68d, 169648365 Protein Sequence	SEQ ID NO: 1012	356 aa	MW at 40366.8kD

KLNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHGLTGFFQLP
HGDFFIEPVKKHPLVEGGYHPHIVYRRQKVPETKEPTCGLKDSVNISQKQELWREKWERHNLPSRSLS
RRSISKERWVETLVVADTKMIEYHGSENVESYILTIMNMVTGLFHNPSIGNAIHIVVVRLILLEEEEQ
GLKIVHHAEKTLSSFCKWQKSINPKSDLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHR
SCNINEDSGLPLAFTIAHELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPTPLTWSKCSEEYITR
FLDRGWGFCLDDIPLE

NOV68e, 284068250 DNA Sequence	SEQ ID NO: 1013	4777 bp
	ORF Start: at 1	ORF Stop: TGA at 4729

AGATCTTATGGGAGACAGCCTCAGCCAGGCCCGGTTCGCTTCCCGGACAGGAGGCAAGAGCATTTTAT
CAAGGGCCTGCCAGAATACCACGTGGTGGGTCCAGTCCGAGTAGATGCCAGTGGGCATTTTTTGTCAT
ATGGCTTGCACTATCCCATCACGAGCAGCAGGAGGAAGAGAGATTTGGATGGCTCAGAGGACTGGGTG
TACTACAGAATTTCTCACGAGGAGAAGGACCTGTTTTTTAACTTGACGGTCAATCAAGGATTTCTTTC
CAATAGCTACATCATGGAGAAGAGATATGGGAACCTCTCCCATGTTAAGATGATGGCTTCTTCTGCCC
CCCTCTGCCATCTCAATGGCACGGTTCTACAGCAGGGCACCAGAGTTGGGACGGCAGCCCTCAGTGCC
TGCCATGGACTGACTGGATTTTTCCAACTACCACATGGAGACTTTTTCATTGAACCCGTGAAGAAGCA
TCCACTGGTTGAGGGAGGGTACCACCCGCACATCGTTTACAGGAGGCAGAAAGTTCCAGAAACCAAGG
AGCCAACCTGTGGATTAAAGGACAGTGTTAACATCTCCCAGAAGCAAGAGCTATGGCGGGAGAAGTGG
GAGAGGCACAACTTGCCAAGCAGAAGCCTCTCTCGGCGTTCCATCAGCAAGGAGAGATGGGTGGAGAC
ACTGGTGGTGGCCGACACAAAGATGATTGAATACCATGGGAGTGAGAATGTGGAGTCCTACATCCTCA
CCATCATGAACATGGTCACTGGGTTGTTCCATAACCCAAGCATTGGCAATGCAATTCACATTGTTGTG
GTTCGGCTCATTCTACTCGAAGAAGAAGAGCAAGGACTGAAAATAGTTCACCATGCAGAAAAGACACT
GTCTAGCTTCTGCAAGTGGCAGAAGAGTATCAATCCCAAGAGTGACCTCAATCCTGTTCATCACGACG
TGGCTGTCCTTCTCACCAGAAAGGACATCTGTGCTGGTTTCAATCGCCCCTGCGAGACCCTGGGCCTG
TCTCACCTTTCAGGAATGTGTCAGCCTCACCGCAGTTGTAACATCAATGAAGATTCGGGACTCCCTCT
GGCTTTCACAATTGCCCATGAGCTAGGACACAGCTTCGGCATCCAGCATGATGGGAAAGAAAATGACT
GTGAGCCTGTGGGCAGACATCCGTACATCATGTCCCGCCAGCTCCAGTACGATCCCACTCCGCTGACA
TGGTCCAAGTGCAGCGAGGAGTACATCACCCGCTTCTTGGACCGAGGCTGGGGGTTCTGTCTTGATGA
CATACCTAAAAAGAAAGGCTTGAAGTCCAAGGTCATTGCCCCCGGAGTGATCTATGATGTTCACCACC
AGTGCCAGCTACAATATGGACCCAATGCTACCCTCTGCCAGGAAGTAGAAAACGTCTGCCAGACACTG
TGGTGCTCCGTGAAGGGCTTTTGTCGCTCTAAGCTGGACGCTGCTGCAGATGGAACTCAATGTGGTGA
GAAGAAGTGGTGTATGGCAGGCAAGTGCATCACAGTGGGGAAGAAACCAGAGAGCATTCCTGGAGGCT
GGGGCCGCTGGTCACCCTGGTCCCACTGTTCCAGGACCTGTGGGGCTGGAGTCCAGAGCGCAGAGAGG
CTCTGCAACAACCCCGAGCCAAAGTTTGGAGGGAAATATTGCACTGGAGAAAGAAAACGCTATCGCTT
GTGCAACGTCCACCCCTGTCGCTCAGAGGCACCAACATTTCGGCAGATGCAGTGCAGTGAATTTGACA
CTGTTCCCTACAAGAATGAACTCTACCACTGGTTTCCCATTTTTAACCCAGCACATCCTTGTGAGCTC
TACTGCCGACCCATAGATGGCCAGTTTTCTGAGAAAATGCTGGATGCTGTCATTGATGGTACCCCTTG
CTTTGAAGGCGGCAACAGCAGAAATGTCTGTATTAATGGCATATGTAAGATGGTTGGCTGTGACTATG
AGATCGATTCCAATGCCACCGAGGATCGCTGCGGTGTGTGCCTGGGAGATGGCTCTTCCTGCCAGACT
GTGAGAAAGATGTTTAAGCAGAAGGAAGGATCTGGTTATGTTGACATTGGGCTCATTCCAAAAGGAGC
AAGGGACATAAGAGTGATGGAAATTGAGGGAGCTGGAAACTTCCTGGCCATCAGGAGTGAAGATCCTG
AAAAATATTACCTGAATGGAGGGTTTATTATCCAGTGGAACGGGAACTATAAGCTGGCAGGGACTGTC
TTTCAGTATGACAGGAAAGGAGACCTGGAAAAGCTGATGGCCACAGGTCCCACCAATGAGTCTGTGTG
GATCCAGCTTCTATTCCAGGTGACTAACCCTGGCATCAAGTATGAGTACACAATCCAGAAAGATGGCC
TTGACAATGATGTTGAGCAGCAGATGTACTTCTGGCAGTACGGCCACTGGACAGAGTGCAGTGTGACC
TGCGGGACAGGTATCCGCCGCCAAACTGCCCATTGCATAAAGAAGGGCCGCGGGATGGTGAAAGCTAC
ATTCTGTGACCCAGAAACACAGCCCAATGGGAGACAGAAGAAGTGCCATGAAAAGGCTTGTCCACCCA
GGTGGTGGGCAGGGGAGTGGGAAGCATGCTCGGCGACATGCGGGCCCCACGGGGAGAAGAAGCGAACC
GTGCTGTGCATCCAGACCATGGTCTCTGACGAGCAGGCTCTCCCGCCCACAGACTGCCAGCACCTGCT
GAAGCCCAAGACCCTCCTTTCCTGCAACAGAGACATCCTGTGCCCCTCGGACTGGACAGTGGGCAACT
GGAGTGAGTGTTCTGTTTCCTGTGGTGGTGGAGTGCGGATTCGCAGTGTCACATGTGCCAAGAACCAT
GATGAACCTTGCGATGTGACAAGGAAACCCAACAGCCGAGCTCTGTGTGGCCTCCAGCAATGCCCTTC
TAGCCGGAGAGTTCTGAAACCAAACAAAGGCACTATTTCCAATGGAAAAAACCCACCAACACTAAAGC
CCGTCCCTCCACCTACATCCAGGCCCAGAATGCTGACCACACCCACAGGGCCTGAGTCTATGAGCACA
AGCACTCCAGCAATCAGCAGCCCTAGTCCTACCACAGCCTCCAAAGAAGGAGACCTGGGTGGGAAACA
GTGGCAAGATAGCTCAACCCAACCTGAGCTGAGCTCTCGCTATCTCATTTCCACTGGAAGCACTTCCC

AGCCCATCCTCACTTCCCAATCCTTGAGCATTCAGCCAAGTGAGGAAAATGTTTCCAGTTCAGATACT
GGTCCTACCTCGGAGGGAGGCCTTGTAGCTACAACAACAAGTGGTTCTGGCTTGTCATCTTCCCGCAA

CCCTATCACTTGGCCTGTGACTCCATTTTACAATACCTTGACCAAAGGTCCAGAAATGGAGATTCACA
GTGGCTCAGGGGAAGAAAGAGAACAGCCTGAGGACAAAGATGAAAGCAATCCTGTAATATGGACCAAG
ATCAGAGTACCTGGAAATGACGCTCCAGTGGAAAGTACAGAAATGCCACTTGCACCTCCACTAACACC
AGATCTCAGCAGGGAGTCCTGGTGGCCACCCTTCAGCACAGTAATGGAAGGACTGCTCCCCAGCCAAA
GGCCCACTACTTCCGAAACTGGGACACCCAGAGTTGAGGGGATGGTTACTGAAAAGCCAGCCAACACT
CTGCTCCCTCTGGGAGGAGACCACCAGCCAGAACCCTCAGGAAAGACGGCAAACCGTAACCACCTGAA
ACTTCCAAACAACATGAACCAAACAAAAAGTTCTGAACCAGTCCTGACTGAGGAGGATGCAACAAGTC
TGATTACTGAGGGCTTTTTGCTAAATGCCTCCAATTACAAGCAGCTCACAAACGGCCACGGCTCTGCA
CACTGGATCGTCGGAAACTGGAGCGAGTGCTCCACCACATGTGGCCTGGGGGCCTACTGGAGAAGGGT
GGAGTGCAGCACCCAGATGGATTCTGACTGTGCGGCCATCCAGAGACCTGACCCTGCAAAAAGATGCC
ACCTCCGTCCCTGTGCTGGCTGGAAAGTGGGAAACTGGAGCAAGTGCTCCAGAAACTGCAGTGGGGGC
TTCAAGATACGCGAGATTCAGTGCGTGGACAGCCGGGACCACCGGAACCTGAGGCCATTTCACTGCCA
GTTCCTGGCCGGCATTCCTCCCCCATTGAGCATGAGCTGTAACCCGGAGCCCTGTGAGGCGTGGCAGG
TGGAGCCTTGGAGCCAGTGCTCCAGGTCCTGTGGAGGTGGAGTTCAGGAGAGAGGAGTGTTCTGTCCA
GGAGGCCTCTGTGATTGGACAAAAAGACCCACATCCACCATGTCTTGCAATGAGCACCTGTGCTGTCA
CTGGGCCACTGGGAACTGGGACCTGTGTTCCACTTCCTGTGGAGGTGGCTTTCAGAAGAGGACTGTCC
AATGTGTGCCCTCAGAGGGCAATAAAACTGAAGACCAAGACCAATGTCTATGTGATCACAAACCCAGA
CCTCCAGAATTCAAAAAATGCAACCAGCAGGCCTGCAAGAAAAGTGCCGATTTACTTTGCACTAAGGA
CAAACTGTCAGCCAGTTTCTGCCAGACACTGAAAGCCATGAAGAAATGTTCTGTGCCCACCGTGAGGG
CTGAGTGCTGCTTCTCGTGTCCCCAGACACACATCACACACACCCAAAGGCAAAGAAGGCAACGGTTG
CTCCAAAAGTCAAAAGAACTCCTCGAGGAGTTCTTTTGACTTTTGGAGCAACCGGGCCTGGCTGAGGC
TGTCTCCCATAAGATCT

NOV68e, 284068250 Protein Sequence	SEQ ID NO: 1014	1576 aa	MW at 175887.8kD

RSYGRQPQPGPVRFPDRRQEHFIKGLPEYHVVGPVRVDASGHFLSYGLHYPITSSRRKRDLDGSEDWV
YYRISHEEKDLFFNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLNGTVLQQGTRVGTAALSA
CHGLTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRRQKVPETKEPTCGLKDSVNISQKQELWREKW
ERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGSENVESYILTIMNMVTGLFHNPSIGNAIHIVV
VRLILLEEEEQGLKIVHHAEKTLSSFCKWQKSINPKSDLNPVHHDVAVLLTRKDICAGFNRPCETLGL
SHLSGMCQPHRSCNINEDSGLPLAFTIAHELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPTPLT
WSKCSEEYITRFLDRGWGFCLDDIPKKKGLKSKVIAPGVIYDVHHQCQLQYGPNATLCQEVENVCQTL
WCSVKGFCRSKLDAAADGTQCGEKKWCMAGKCITVGKKPESIPGGWGRWSPWSHCSRTCGAGVQSAER
LCNNPEPKFGGKYCTGERKRYRLCNVHPCRSEAPTFRQMQCSEFDTVPYKNELYHWFPIFNPAHPCEL
YCRPIDGQFSEKMLDAVIDGTPCFEGGNSRNVCINGICKMVGCDYEIDSNATEDRCGVCLGDGSSCQT
VRKMFKQKEGSGYVDIGLIPKGARDIRVMEIEGAGNFLAIRSEDPEKYYLNGGFIIQWNGNYKLAGTV
FQYDRKGDLEKLMATGPTNESVWIQLLFQVTNPGIKYEYTIQKDGLDNDVEQQMYFWQYGHWTECSVT
CGTGIRRQTAHCIKKGRGMVKATFCDPETQPNGRQKKCHEKACPPRWWAGEWEACSATCGPHGEKKRT
VLCIQTMVSDEQALPPTDCQHLLKPKTLLSCNRDILCPSDWTVGNWSECSVSCGGGVRIRSVTCAKNH
DEPCDVTRKPNSRALCGLQQCPSSRRVLKPNKGTISNGKNPPTLKPVPPPTSRPRMLTTPTGPESMST
STPAISSPSPTTASKEGDLGGKQWQDSSTQPELSSRYLISTGSTSQPILTSQSLSIQPSEENVSSSDT
GPTSEGGLVATTTSGSGLSSSRNPITWPVTPFYNTLTKGPEMEIHSGSGEEREQPEDKDESNPVIWTK
IRVPGNDAPVESTEMPLAPPLTPDLSRESWWPPFSTVMEGLLPSQRPTTSETGTPRVEGMVTEKPANT
LLPLGGDHQPEPSGKTANRNHLKLPNNMNQTKSSEPVLTEEDATSLITEGFLLNASNYKQLTNGHGSA
HWIVGNWSECSTTCGLGAYWRRVECSTQMDSDCAAIQRPDPAKRCHLRPCAGWKVGNWSKCSRNCSGG
FKIREIQCVDSRDHRNLRPFHCQFLAGIPPPLSMSCNPEPCEAWQVEPWSQCSRSCGGGVQERGVFCP
GGLCDWTKRPTSTMSCNEHLCCHWATGNWDLCSTSCGGGFQKRTVQCVPSEGNKTEDQDQCLCDHKPR
PPEFKKCNQQACKKSADLLCTKDKLSASFCQTLKAMKKCSVPTVRAECCFSCPQTHITHTQRQRRQRL
LQKSKELLEEFF

NOV68f, 305867866 DNA Sequence	SEQ ID NO: 1015	4777 bp
	ORF Start: at 1	ORF Stop: TGA at 4729

AGATCTTATGGGAGACAGCCTCAGCCAGGCCCGGTTCGCTTCCCGGACAGGAGGCAAGAGCATTTTAT
CAAGGGCCTGCCAGAATACCACGTGGTGGGTCCAGTCCGAGTAGATGCCAGTGGGCATTTTTTGTCAT
ATGGCTTGCACTATCCCATCACGAGCAGCAGGAGGAAGAGAGATTTGGATGGCTCAGAGGACTGGGTG
TACTACAGAATTTCTCACGAGGAGAAGGACCTGTTTTTTAACTTGACGGTCAATCAAGGATTTCTTTC
CAATAGCTACATCATGGAGAAGAGATATGGGAACCTCTCCCATGTTAAGATGATGGCTTCCTCTGCCC
CCCTCTGCCATCTCAGTGGCACGGTTCTACAGCAGGGCACCAGAGTTGGGACGGCAGCCCTCAGTGCC
TGCCATGGACTGACTGGATTTTTCCAACTACCACATGGAGACTTTTTCATTGAACCCGTGAAGAAGCA
TCCACTGGTTGAGGGAGGGTACCACCCGCACATCGTTTACAGGAGGCAGAAAGTTCCAGAAACCAAGG

AGCCAACCTGTGGATTAAAGGACAGTGTTAACATCTCCCAGAAGCAAGAGCTATGGCGGGAGAAGTGG
GAGAGGCACAACTTGCCAAGCAGAAGCCTCTCTCGGCGTTCCATCAGCAAGGAGAGATGGGTGGAGAC
ACTGGTGGTGGCCGACACAAAGATGATTGAATACCATGGGAGTGAGAATGTGGAGTCCTACATCCTCA
CCATCATGAACATGGTCACTGGGTTGTTCCATAACCCAAGCATTGGCAATGCAATTCACATTGTTGTG
GTTCGGCTCATTCTACTCGAAGAAGAAGAGCAAGGACTGAAAATAGTTCACCATGCAGAAAAGACACT
GTCTAGCTTCTGCAAGTGGCAGAAGAGTATCAATCCCAAGAGTGACCTCAATCCTGTTCATCACGACG
TGGCTGTCCTTCTCACTAGAAAGGACATCTGTGCTGGTTTCAATCGCCCCTGCGAGACCCTGGGCCTG
TCTCACCTTTCAGGAATGTGTCAGCCTCACCGCAGTTGTAACATCAATGAAGATTCGGGACTCCCTCT
GGCTTTCACAATTGCCCATGAGCTAGGACACAGCTTCGGCATCCAGCATGATGGGAAAGAAAATGACT
GTGAGCCTGTGGGCAGACATCCGTACATCATGTCCCGCCAGCTCCAGTACGATCCCACTCCGCTGACA
TGGTCCAAGTGCAGCGAGGAGTACATCACCCGCTTCTTGGACCGAGGCTGGGGGTTCTGTCTTGATGA
CATACCTAAAAAGAAAGGCTTGAAGTCCAAGGTCATTGCCCCCGGAGTGATCTATGATGTTCACCACC
AGTGCCAGCTACAATATGGACCCAATGCTACCTTCTGCCAGGAAGTAGAAAACGTCTGCCAGACACTG
TGGTGCTCCGTGAAGGGCTTTTGTCGCTCTAAGCTGGACGCTGCTGCAGATGGAACTCAATGTGGTGA
GAAGAAGTGGTGTATGGCAGGCAAGTGCATCACAGTGGGGAAGAAACCAGAGAGCATTCCTGGAGGCT
GGGGCCGCTGGTCACCCTGGTCCCACTGTTCCAGGACCTGTGGGGCTGGAGTCCAGAGCGCAGAGAGG
CTCTGCAACAACCCCGAGCCAAAGTTTGGAGGGAAATATTGCACTGGAGAAAGAAAACGCTATCGCTT
GTGCAACGTCCACCCCTGTCGCTCAGAGGCACCAACATTTCGGCAGATGCAGTGCAGTGAATTTGACA
CTGTTCCCTACAAGAATGAACTCTACCACTGGTTTCCCATTTTTAACCCAGCACATCCTTGTGAGCTC
TACTGCCGACCCATAGATGGCCAGTTTTCTGAGAAAATGCTGGATGCTGTCATTGATGGTACCCCTTG
CTTTGAAGGCGGCAACAGCAGAAATGTCTGTATTAATGGCATATGTAAGATGGTTGGCTGTGACTATG
AGATCGATTCCAATGCCACCGAGGATCGCTGCGGTGTGTGCCTGGGAGATGGCTCTTCCTGCCAGACT
GTGAGAAAGATGTTTAAGCAGAAGGAAGGATCTGGTTATGTTGACATTGGGCTCATTCCAAAAGGAGC
AAGGGACATAAGAGTGATGGAAATTGAGGGAGCTGGAAACTTCCTGGCCATCAGGAGTGAAGATCCTG
AAAAATATTACCTGAATGGAGGGTTTATTATCCAGTGGAACGGGAACTATAAGCTGGCAGGGACTGTC
TTTCAGTATGACAGGAAAGGAGACCTGGAAAAGCTGATGGCCACAGGTCCCACCAATGAGTCTGTGTG
GATCCAGCTTCTATTCCAGGTGACTAACCCTGGCATCAAGTATGAGTACACAATCCAGAAAGATGGCC
TTGACAATGATGTTGAGCAGCAGATGTACTTCTGGCAGTACGGCCACTGGACAGAGTGCAGTGTGACC
TGCGGGACAGGTATCCGCCGCCAAACTGCCCATTGCATAAAGAAGGGCCGCGGGATGGTGAAAGCTAC
ATTCTGTGACCCAGAAACACAGCCCAATGGGAGACAGAAGAAGTGCCATGAAAAGGCTTGTCCACCCA
GGTGGTGGGCAGGGGAGTGGGAAGCATGCTCGGCGACATGCGGGCCCCACGGGGAGAAGAAGCGAACC
GTGCTGTGCATCCAGACCATGGTCTCTGACGAGCAGGCTCTCCCGCCCACAGACTGCCAGCACCTGCT
GAAGCCCAAGACCCTCCTTTCCTGCAACAGAGACATCCTGTGCCCCTCGGACTGGACAGTGGGCAACT
GGAGTGAGTGTTCTGTTTCCTGTGGTGGTGGAGTGCGGATTCGCAGTGTCACATGTGCCAAGAACCAT
GATGAACCTTGCGATGTGACAAGGAAACCCAACAGCCGAGCTCTGTGTGGCCTCCAGCAATGCCCTTC
TAGCCGGAGAGTTCTGAAACCAAACAAAGGCACTATTTCCAATGGAAAAAACCCACCAACACTAAAGC
CCGTCCCTCCACCTACATCCAGGCCCAGAATGCTGACCACACCCACAGGGCCTGAGTCTATGAGCACA
AGCACTCCAGCAATCAGCAGCCCTAGTCCTACCACAGCCTCCAAAGAAGGAGACCTGGGTGGGAAACA
GTGGCAAGATAGCTCAACCCAACCTGAGCTGAGCTCTCGCTATCTCATTTCCACTGGAAGCACTTCCC
AGCCCATCCTCACTTCCCAATCCTTGAGCATTCAGCCAAGTGAGGAAAATGTTTCCAGTTCAGATACT
GGTCCTACCTCGGAGGGAGGCCTTGTAGCTACAACAACAAGTGGTTCTGGCTTGTCATCTTCCCGCAA

CCCTATCACTTGGCCTGTGACTCCATTTTACAATACCTTGACCAAAGGTCCAGAAATGGAGATTCACA
GTGGCTCAGGGGAAGAAAGAGAACAGCCTGAGGACAAAGATGAAAGCAATCCTGTAATATGGACCAAG
ATCAGAGTACCTGGAAATGACGCTCCAGTGGAAAGTACAGAAATGCCACTTGCACCTCCACTAACACC
AGATCTCAGCAGGGAGTCCTGGTGGCCACCCTTCAGCACAGTAATGGAAGGACTGCTCCCCAGCCAAA
GGCCCACTACTTCCGAAACTGGGACACCCAGAGTTGAGGGGATGGTTACTGAAAAGCCAGCCAACACT
CTGCTCCCTCTGGGAGGAGACCACCAGCCAGAACCCTCAGGAAAGACGGCAAACCGTAACCACCTGAA
ACTTCCAAACAACATGAACCAAACAAAAAGTTCTGAACCAGTCCTGACTGAGGAGGATGCAACAAGTC
TGATTACTGAGGGCTTTTTGCTAAATGCCTCCAATTACAAGCAGCTCACAAACGGCCACGGCTCTGCA
CACTGGATCGTCGGAAACTGGAGCGAGTGCTCCACCACATGTGGCCTGGGGGCCTACTGGAGAAGGGT
GGAGTGCAGCACCCAGATGGATTCTGACTGTGCGGCCATCCAGAGACCTGACCCTGCAAAAAGATGCC
ACCTCCGTCCCTGTGCTGGCTGGAAAGTGGGAAACTGGAGCAAGTGCTCCAGAAACTGCAGTGGGGGC
TTCAAGATACGCGAGATTCAGTGCGTGGACAGCCGGGACCACCGGAACCTGAGGCCATTTCACTGCCA
GTTCCTGGCCGGCATTCCTCCCCCATTGAGCATGAGCTGTAACCCGGAGCCCTGTGAGGCGTGGCAGG
TGGAGCCTTGGAGCCAGTGCTCCAGGTCCTGTGGAGGTGGAGTTCAGGAGAGAGGAGTGTTCTGTCCA
GGAGGCCTCTGTGATTGGACAAAAAGACCCACATCCACCATGTCTTGCAATGAGCACCTGTGCTGTCA
CTGGGCCACTGGGAACTGGGACCTGTGTTCCACTTCCTGTGGAGGTGGCTTTCAGAAGAGGACTGTCC
AATGTGTGCCCTCAGAGGGCAATAAAACTGAAGACCAAGACCAATGTCTATGTGATCACAAACCCAGA
CCTCCAGAATTCAAAAAATGCAACCAGCAGGCCTGCAAGAAAAGTGCCGATTTACTTTGCACTAAGGA
CAAACTGTCAGCCAGTTTCTGCCAGACACTGAAAGCCATGAAGAAATGTTCTGTGCCCACCGTGAGGG
CTGAGTGCTGCTTCTCGTGTCCCCAGACACACATCACACACACCCAAAGGCAAAGAAGGCAACGGTTG
CTCCAAAAGTCAAAAGGACTCCTCGAGGAGTTCTTTTGACTTTTGGAGCAACCGGGCCTGGCTGAGGC

TGTCTCCCATAAGATCT

NOV68f, 305867866 Protein Sequence	SEQ ID NO: 1016	1576 aa	MW at 175822.7kD

RSYGRQPQPGPVRFPDRRQEHFIKGLPEYHVVGPVRVDASGHFLSYGLHYPITSSRRKRDLDGSEDWV
YYRISHEEKDLFFNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSA
CHGLTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRRQKVPETKEPTCGLKDSVNISQKQELWREKW
ERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGSENVESYILTIMNMVTGLFHNPSIGNAIHIVV
VRLILLEEEEQGLKIVHHAEKTLSSFCKWQKSINPKSDLNPVHHDVAVLLTRKDICAGFNRPCETLGL
SHLSGMCQPHRSCNINEDSGLPLAFTIAHELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPTPLT
WSKCSEEYITRFLDRGWGFCLDDIPKKKGLKSKVIAPGVIYDVHHQCQLQYGPNATFCQEVENVCQTL
WCSVKGFCRSKLDAAADGTQCGEKKWCMAGKCITVGKKPESIPGGWGRWSPWSHCSRTCGAGVQSAER
LCNNPEPKFGGKYCTGERKRYRLCNVHPCRSEAPTFRQMQCSEFDTVPYKNELYHWFPIFNPAHPCEL
YCRPIDGQFSEKMLDAVIDGTPCFEGGNSRNVCINGICKMVGCDYEIDSNATEDRCGVCLGDGSSCQT
VRKMFKQKEGSGYVDIGLIPKGARDIRVMEIEGAGNFLAIRSEDPEKYYLNGGFIIQWNGNYKLAGTV
FQYDRKGDLEKLMATGPTNESVWIQLLFQVTNPGIKYEYTIQKDGLDNDVEQQMYFWQYGHWTECSVT
CGTGIRRQTAHCIKKGRGMVKATFCDPETQPNGRQKKCHEKACPPRWWAGEWEACSATCGPHGEKKRT
VLCIQTMVSDEQALPPTDCQHLLKPKTLLSCNRDILCPSDWTVGNWSECSVSCGGGVRIRSVTCAKNH
DEPCDVTRKPNSRALCGLQQCPSSRRVLKPNKGTISNGKNPPTLKPVPPPTSRPRMLTTPTGPESMST
STPAISSPSPTTASKEGDLGGKQWQDSSTQPELSSRYLISTGSTSQPILTSQSLSIQPSEENVSSSDT
GPTSEGGLVATTTSGSGLSSSRNPITWPVTPFYNTLTKGPEMEIHSGSGEEREQPEDKDESNPVIWTK
IRVPGNDAPVESTEMPLAPPLTPDLSRESWWPPFSTVMEGLLPSQRPTTSETGTPRVEGMVTEKPANT
LLPLGGDHQPEPSGKTANRNHLKLPNNMNQTKSSEPVLTEEDATSLITEGFLLNASNYKQLTNGHGSA
HWIVGNWSECSTTCGLGAYWRRVECSTQMDSDCAAIQRPDPAKRCHLRPCAGWKVGNWSKCSRNCSGG
FKIREIQCVDSRDHRNLRPFHCQFLAGIPPPLSMSCNPEPCEAWQVEPWSQCSRSCGGGVQERGVFCP
GGLCDWTKRPTSTMSCNEHLCCHWATGNWDLCSTSCGGGFQKRTVQCVPSEGNKTEDQDQCLCDHKPR
PPEFKKCNQQACKKSADLLCTKDKLSASFCQTLKAMKKCSVPTVRAECCFSCPQTHITHTQRQRRQRL
LQKSKGLLEEFF

NOV68g, 318176397 DNA Sequence	SEQ ID NO: 1017	3174 bp
	ORF Start: at 1	ORF Stop: end of sequence

CTGGAATGCGCCCTTAGATCTGGCCGCTGGTCACCCTGGTCCCACTGTTCCAGGACCTGTGGGGCTGG
AGTCCAGAGCGCAGAGAGGCTCTGCAACAACCCCGAGCCAAAGTTTGGAGGGAAATATTGCACTGGAG
AAAGAAAACGCTATCGCTTGTGCAACGTCCACCCCTGTCGCTCAGAGGCACCAACATTTCGGCAGATG
CAGTGCAGTGAATTTGACACTGTTCCCTACAAGAATGAACTCTACCACTGGTTTCCCATTTTTAACCC
AGCACATCCTTGTGAGCTCTACTGCCGACCCATAGATGGCCAGTTTTCTGAGAAAATGCTGGATGCTG
TCATTGATGGTACCCCTTGCTTTGAAGGCGGCAACAGCAGAAATGTCTGTATTAATGGCATATGTAAG
ATGGTTGGCTGTGACTATGAGATCGATTCCAATGCCACCGAGGATCGCTGCGGTGTGTGCCTGGGAGA
TGGCTCTTCCTGCCAGACTGCGAGAAAGATGTTTAAGCAGAAGGAAGGATCTGGTTATGTTGACATTG
GGCTCATTCCAAAAGGAGCAAGGGACATAAGAGTGATGGAAATTGAGGGAGCTGGAAACTTCCTGGCC
ATCAGGAGTGAAGATCCTGAAAAATATTACCTGAATGGAGGGTTTATTATCCAGTGGAACGGGAACTA
TAAGCTGGCAGGGACTGTCTTTCAGTATGACAGGAAAGGAGACCTGGAAAAGCTGATGGCCACAGGTC
CCACCAATGAGTCTGTGTGGATCCAGCTTCTATTCCAGGTGACTAACCCTGGCATCAAGTATGAGTAC
ACAATCCAGAAAGATGGCCTTGACAATGATGTTGAGCAGCAGATGTACTTCTGGCAGTACGGCCACTG
GACAGAGTGCAGTGTGACCTGCGGGACAGGTATCCGCCGCCAAACTGCCCATTGCATAAAGAAGGGCC
GCGGGATGGTGAAAGCTACATTCTGTGACCCAGAAACACAGCCCAATGGGAGACAGAAGAAGTGCCAT
GAAAAGGCTTGTCCACCCAGGTGGTGGGCAGGGGAGTGGGAAGCATGCTCGGCGACATGCGGGCCCCA
CGGGGAGAAGAAGCGAACCGTGCTGTGCATCCAGACCATGGTCTCTGACGAGCAGGCTCTCCCGCCCA
CAGACTGCCAGCACCTGCTGAAGCCCAAGACCCTCCTTTCCTGCAACAGAGACATCCTGTGCCCCTCG
GACTGGACAGTGGGCAACTGGAGTGAGTGTTCTGTTTCCTGTGGTGGTGGAGTGCGGATTCGCAGTGT
CACATGTGCCAAGAACCATGATGAACCTTGCGATGTGACAAGGAAACCCAACAGCCGAGCTCTGTGTG
GCCTCCAGCAATGCCCTTCTAGCCGGAGAGTTCTGAAACCAAACAAAGGCACTATTTCCAATGGAAAA
AACCCACCAACACTAAAGCCCGTCCCTCCACCTACATCCAGGCCCAGAATGCTGACCACACCCACAGG
GCCTGAGTCTATGAGCACAAGCACTCCGGCAATCAGCAGCCCTAGTCCTACCACAGCCTCCAAAGAAG
GAGACCTGGGTGGGAAACAGTGGCAAGATAGCTCAACCCAACCTGAGCTGAGCTCTCGCTATCTCATT
TCCACTGGAAGCACTTCCCAGCCCATCCTCACTTCCCAATCCTTGAGCATTCAGCCAAGTGAGGAAAA
TGTTTCCAGTTCAGATACTGGTCCTACCTCGGAGGGAGGCCTTGTAGCTACAACAACAAGTGGTTCTG
GCTTGTCATCTTCCCGCAACCCTATCACTTGGCCTGTGACTCCATTTTACAATACCTTGACCAAAGGT
CCAGAAATGGAGATTCACAGTGGCTCAGGGGAAGAAAGAGAACAGCCTGAGGACAAAGATGAAAGCAA
TCCTGTAATATGGACCAAGATCAGAGTACCTGGAAATGACGCTCCAGTGGAAAGTACAGAAATGCCAC
TTGCACCTCCACTAACACCAGATCTCAGCAGGGAGTCCTGGTGGCCACCCTTCAGCACAGTAATGGAA
GGACTGCTCCCCAGCCAAAGGCCCACTACTTCCGAAACTGGGACACCCAGAGTTGAGGGGATGGTTAC

TGAAAAGCCAGCCAACACTCTGCTCCCTCTGGGAGGAGACCACCAGCCAGAACCCTCAGGAAAGACGG
CAAACCGTAACCACCTGAAACTTCCAAACAACATGAACCAAACAAAAAGTTCTGAACCAGTCCTGACT
GAGGAGGATGCAACAAGTCTGATTACTGAGGGCTTTTTGCTAAATGCCTCCAATTACAAGCAGCTCAC
AAACGGCTACGGCTCTGCACACTGGATCGTCGGAAACTGGAGCGAGTGCTCCACCACATGTGGCCTGG
GGGCCTACTGGAGAAGGGTGGAGTGCAGCACCCAGATGGATTCTGACTGTGCGGCCATCCAGAGACCT
GACCCTGCAAAAAGATGCCACCTCCGTCCCTGTGCTGGCTGGAAAGTGGGAAACTGGAGCAAGTGCTC
CAGAAACTGCAGTGGGGGCTTCAAGATACGCGAGATTCAGTGCGTGGACAGCCGGGACCACCGGAACC
TGAGGCCATTTCACTGCCAGTTCCTGGCCGGCATTCCTCCCCCATTGAGCATGAGCTGTAACCCGGAG
CCCTGTGAGGCGTGGCAGGTGGAGCCTTGGAGCCAGTGCTCCAGGTCCTGTGGAGGTGGAGTTCAGGA
GAGAGGAGTGTTCTGTCCAGGAGGCCTCTGTGATTGGACAAAAAGACCCACATCCACCATGTCTTGCA
ATGAGCACCTGTGCTGTCACTGGGCCACTGGGAACTGGGACCTGTGTTCCACTTCCTGTGGAGGCGGC
TTTCAGAAGAGGACTGTCCAATGTGTGCCCTCAGAGGGCAATAAAACTGAAGACCAAGACCAATGTCT
ATGTGATCACAAACCCAGACCTCCAGAATTCAAAAAATGCAACCAGCAGGCCTGCAAGAAAAGTGCCG
ATTTACTTTGCACTAAGGACAAACTGTCAGCCAGTTTCTGCCAGACACTGAAAGCCATGAAGAAATGT
TCTGTGCCCACCGTGAGGGCTGAGTGCTGCTTCTCGTGTCCCCAGACACACATCACACACACCCAAAG
GCAAAGAAGGCAACGGTTGCTCCAAAAGTCAAAAGAACTCCTCGAG

NOV68g, 318176397 Protein Sequence	SEQ ID NO: 1018	1058 aa	MW at 117062.0kD

LECALRSGRWSPWSHCSRTCGAGVQSAERLCNNPEPKFGGKYCTGERKRYRLCNVHPCRSEAPTFRQM
QCSEFDTVPYKNELYHWFPIFNPAHPCELYCRPIDGQFSEKMLDAVIDGTPCFEGGNSRNVCINGICK
MVGCDYEIDSNATEDRCGVCLGDGSSCQTARKMFKQKEGSGYVDIGLIPKGARDIRVMEIEGAGNFLA
IRSEDPEKYYLNGGFIIQWNGNYKLAGTVFQYDRKGDLEKLMATGPTNESVWIQLLFQVTNPGIKYEY
TIQKDGLDNDVEQQMYFWQYGHWTECSVTCGTGIRRQTAHCIKKGRGMVKATFCDPETQPNGRQKKCH
EKACPPRWWAGEWEACSATCGPHGEKKRTVLCIQTMVSDEQALPPTDCQHLLKPKTLLSCNRDILCPS
DWTVGNWSECSVSCGGGVRIRSVTCAKNHDEPCDVTRKPNSRALCGLQQCPSSRRVLKPNKGTISNGK
NPPTLKPVPPPTSRPRMLTTPTGPESMSTSTPAISSPSPTTASKEGDLGGKQWQDSSTQPELSSRYLI
STGSTSQPILTSQSLSIQPSEENVSSSDTGPTSEGGLVATTTSGSGLSSSRNPITWPVTPPFYNTLTKG
PEMEIHSGSGEEREQPEDKDESNPVIWTKIRVPGNDAPVESTEMPLAPPLTPDLSRESWWPPFSTVME
GLLPSQRPTTSETGTPRVEGMVTEKPANTLLPLGGDHQPEPSGKTANRNHLKLPNNMNQTKSSEPVLT
EEDATSLITEGFLLNASNYKQLTNGYGSAHWIVGNWSECSTTCGLGAYWRRVECSTQMDSDCAAIQRP
DPAKRCHLRPCAGWKVGNWSKCSRNCSGGFKIREIQCVDSRDHRNLRPFHCQFLAGIPPPLSMSCNPE
PCEAWQVEPWSQCSRSCGGGVQERGVFCPGGLCDWTKRPTSTMSCNEHLCCHWATGNWDLCSTSCGGG
FQKRTVQCVPSEGNKTEDQDQCLCDHKPRPPEFKKCNQQACKKSADLLCTKDKLSASFCQTLKAMKKC
SVPTVRAECCFSCPQTHITHTQRQRRQRLLQKSKELLE

NOV68h, CG58504-02 DNA Sequence	SEQ ID NO: 1019	1068 bp
	ORF Start: at 7	ORF Stop: at 1063

AAGCTTAACTTGACGGTCAATCAAGGATTTCTTTCCAATAGCTACATCATGGAGAAGAGATATGGGAA
CCTCTCCCATGTTAAGATGATGGCTTCCTCTGCCCCCCTCTGCCATCTCAGTGGCACGGTTCTACAGC
AGGGCACCAGAGTTGGGATGGCAGCCCTCAGTGCCTGCCATGGACTGACTGGATTTTTCCAACTACCA
CATGGAGACTTTTTCATTGAACCCGTGAAGAAGCATCCACTGGTTGAGGGAGGGTACCACCCGCACAT
CGTTTACAGGAGGCAGAAAGTTCCAGAAACCAAGGAGCCAACCTGTGGATTAAAGGACAGTGTTAACA
TCTCCCAGAAGCAAGAGCTATGGCGGGAGAAGTGGGAGAGGCACAACTTGCCAAGCAGAAGCCTCTCT
CGGCGTTCCATCAGCAAGGAGAGATGGGTGGAGACACTGGTGGTGGCCGACACAAAGATGATTGAATA
CCATGGGAGTGAGAATGTGGAGTCCTGCATCCTCACCATCATGAACATGGTCACTGGGTTGTTCCATA
ACCCAAGCATTGGCAATGCAATTCACATTGTTGTGGTTCGGCTCATTCTACTCGAAGAAGAAGAGCAA
GGACTGAAAATAGTTCACCATGCAGAAAAGACACTGTCTAGCTTCTGCAAGTGGCAGAAGAGTATCAA
TCCCAAGAGTGACCTCAATCCTGTTCATCACGACGTGGCTGTCCTTCTCACCAGAAAGGACATCTGTG
CTGGTTTCAATCGCCCCTGCGAGACCCTGGGCCTGTCTCACCTTTCAGGAATGTGTCAGCCTCACCGC
AGTTGTAACATCAATGAAGATTCGGGACTCCCTCTGGCTTTCACAATTGCCCATGAGCTAGGACACAG
CTTCGGCATCCAGCATGATGGGAAAGAAAATGACTGTGAGCCTGTGGGCAGACATCCGTACATCATGT
CCCGCCAGCTCCAGTACGATCCCACTCCGCTGACATGGTCCAAGTGCAGCGAGGAGTACATCACCCGC
TTCTTGGACCGAGGCTGGGGGTTCTGTCTTGATGACATACCTCTCGAG

NOV68h, CG58504-02 Protein Sequence	SEQ ID NO: 1020	352 aa	MW at 39853.3kD

NLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGMAALSACHGLTGFFQLPHG
DFFIEPVKKHPLVEGGYHPHIVYRRQKVPETKEPTCGLKDSVNISQKQELWREKWERHNLPSRSLSRR
SISKERWVETLVVADTKMIEYHGSENVESCILTIMNMVTGLFHNPSIGNAIHIVVVRLILLEEEEQGL
KIVHHAEKTLSSFCKWQKSINPKSDLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSC
NINEDSGLPLAFTIAHELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPTPLTWSKCSEEYITRFL

DRGWGFCLDDIP

NOV68i, CG58504-03 DNA Sequence	SEQ ID NO: 1021	1068 bp
	ORF Start: at 7	ORF Stop: at 1063

AAGCTTAACTTGACGGTCAATCAAGGATTTCTTTCCAATAGCTACATCATGGAGAAGAGATATGGGAA
CCTCTCCCATGTTAAGATGATGGCTTCCTCTGCCCCCCTCTGCCATCTCAGTGGCACGGTTCTACAGC
AGGGCACCAGAGTTGGGACGGCAGCCCTCAGTGCCTGCCATGGACTGACTGGATTTTTCCAACTACCA
CATGGAGACTTTTTCATTGAACCCGTGAAGAAGCATCCACTGGTTGAGGGAGGGTACCACCCGCACAT
CGTTTACAGGAGGCAGAAAGTTCCAGAAACCAAGGAGCCAACCTGTGGATTAAAGGACAGTGTTAACA
TCTCCCAGAAGCAAGAGCTATGGCGGGAGAAGTGGGAGAGGCACAACTTGCCAAGCAGAAGCCTCTCT
CGGCGTTCCATCAGCAAGGAGAGATGGGTGGAGACACTGGTGGTGGCCGACACAAAGATGATTGAATA
CCATGGGAGTGAGAATGTGGAGTCCTACATCCTCACCATCATGAACATGATCACTGGGTTGTTCCATA
ACCCAAGCATTGGCAATGCAATTCACATTGTTGTGGTTCGGCTCATTCTACTCGAAGAAGAAGAGCAA
GGACTGAAAATAGTTCACCATGCAGAAAAGACACTGTCTAGCTTCTGCAAGTGGCAGAAGAGTATCAA
TCCCAAGAGTGACCTCAATCCTGTTCATCACGACGTGGCTGTCCTTCTCACCAGAAAGGACATCTGTG
CTGGTTTCAATCGCCCCTGCGAGACCCTGGGCCTGTCTCACCTTTCAGGAATGTGTCAGCCTCACCGC
AGTTGTAACATCAATGAAGATTCGGGACTCCCTCTGGCTTTCACAATTGCCCATGAGCTAGGACACAG
CTTCGGCATCCAGCATGATGGGAAAGAAAATGACTGTGAGCCTGTGGGCAGACATCCGTACATCATGT
CCCGCCAGCTCCAGTACGATCCCACTCCGCTGACATGGTCCAAGTGCAGCGAGGAGTACATCACCCGC
TTCTTGGACCGAGGCTGGGGGTTCTGTCTTGATGACATACCTCTCGAG

NOV68i, CG58504-03 Protein Sequence	SEQ ID NO: 1022	352 aa	MW at 39897.2kD

NLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHGLTGFFQLPHG
DFFIEPVKKHPLVEGGYHPHIVYRRQKVPETKEPTCGLKDSVNISQKQELWREKWERHNLPSRSLSRR
SISKERWVETLVVADTKMIEYHGSENVESYILTIMNMITGLFHNPSIGNAIHIVVVRLILLEEEEQGL
KIVHHAEKTLSSFCKWQKSINPKSDLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSC
NINEDSGLPLAFTIAHELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPTPLTWSKCSEEYITRFL
DRGWGFCLDDIP

NOV68j, CG58504-04 DNA Sequence	SEQ ID NO: 1023	252 bp
	ORF Start: at 7	ORF Stop: at 247

AAGCTTCACCAGTGCCAGCTACAATATGGACCCAATGCTACCTTCTGCCAGGAAGTAGAAAACGTCTG
CCAGACACTGTGGTGCTCCGTGAAGGGCTTTTGTCGCTCTAAGCTGGACGCTGCTGCAGATGGAACTC
AATGTGGTGAGAAGAAGTGGTGTATGGCAGGCAAGTGCATCACAGTGGGGAAGAAACCAGAGAGCATT
CCTGGAGGCTGCGGCCGCTGGTCACCCTGGTCCCACTGTTCCCTCGAG

NOV68j, CG58504-04 Protein Sequence	SEQ ID NO: 1024	80 aa	MW at 8757.0kD

HQCQLQYGPNATFCQEVENVCQTLWCSVKGFCRSKLDAAADGTQCGEKKWCMAGKCITVGKKPESIPG
GCGRWSPWSHCS

NOV68k, CG58504-05 DNA Sequence	SEQ ID NO: 1025	4777 bp
	ORF Start: at 7	ORF Stop: TGA at 4729

AGATCTTATGGGAGACAGCCTCAGCCAGGCCCGGTTCGCTTCCCGGACAGGAGGCAAGAGCATTTTAT
CAAGGGCCTGCCAGAATACCACGTGGTGGGTCCAGTCCGAGTAGATGCCAGTGGGCATTTTTTGTCAT
ATGGCTTGCACTATCCCATCACGAGCAGCAGGAGGAAGAGAGATTTGGATGGCTCAGAGGACTGGGTG
TACTACAGAATTTCTCACGAGGAGAAGGACCTGTTTTTTAACTTGACGGTCAATCAAGGATTTCTTTC
CAATAGCTACATCATGGAGAAGAGATATGGGAACCTCTCCCATGTTAAGATGATGGCTTCCTCTGCCC
CCCTCTGCCATCTCAGTGGCACGGTTCTACAGCAGGGCACCAGAGTTGGGACGGCAGCCCTCAGTGCC
TGCCATGGACTGACTGGATTTTTCCAACTACCACATGGAGACTTTTTCATTGAACCCGTGAAGAAGCA
TCCACTGGTTGAGGGAGGGTACCACCCGCACATCGTTTACAGGAGGCAGAAAGTTCCAGAAACCAAGG
AGCCAACCTGTGGATTAAAGGACAGTGTTAACATCTCCCAGAAGCAAGAGCTATGGCGGGAGAAGTGG
GAGAGGCACAACTTGCCAAGCAGAAGCCTCTCTCGGCGTTCCATCAGCAAGGAGAGATGGGTGGAGAC
ACTGGTGGTGGCCGACACAAAGATGATTGAATACCATGGGAGTGAGAATGTGGAGTCCTACATCCTCA
CCATCATGAACATGGTCACTGGGTTGTTCCATAACCCAAGCATTGGCAATGCAATTCACATTGTTGTG
GTTCGGCTCATTCTACTCGAAGAAGAAGAGCAAGGACTGAAAATAGTTCACCATGCAGAAAAGACACT
GTCTAGCTTCTGCAAGTGGCAGAAGAGTATCAATCCCAAGAGTGACCTCAATCCTGTTCATCACGACG
TGGCTGTCCTTCTCACTAGAAAGGACATCTGTGCTGGTTTCAATCGCCCCTGCGAGACCCTGGGCCTG
TCTCACCTTTCAGGAATGTGTCAGCCTCACCGCAGTTGTAACATCAATGAAGATTCGGGACTCCCTCT
GGCTTTCACAATTGCCCATGAGCTAGGACACAGCTTCGGCATCCAGCATGATGGGAAAGAAAATGACT

GTGAGCCTGTGGGCAGACATCCGTACATCATGTCCCGCCAGCTCCAGTACGATCCCACTCCGCTGACA
TGGTCCAAGTGCAGCGAGGAGTACATCACCCGCTTCTTGGACCGAGGCTGGGGGTTCTGTCTTGATGA
CATACCTAAAAAGAAAGGCTTGAAGTCCAAGGTCATTGCCCCCGGAGTGATCTATGATGTTCACCACC
AGTGCCAGCTACAATATGGACCCAATGCTACCTTCTGCCAGGAAGTAGAAAACGTCTGCCAGACACTG
TGGTGCTCCGTGAAGGGCTTTTGTCGCTCTAAGCTGGACGCTGCTGCAGATGGAACTCAATGTGGTGA
GAAGAAGTGGTGTATGGCAGGCAAGTGCATCACAGTGGGGAAGAAACCAGAGAGCATTCCTGGAGGCT
GGGGCCGCTGGTCACCCTGGTCCCACTGTTCCAGGACCTGTGGGGCTGGAGTCCAGAGCGCAGAGAGG
CTCTGCAACAACCCCGAGCCAAAGTTTGGAGGGAAATATTGCACTGGAGAAAGAAAACGCTATCGCTT
GTGCAACGTCCACCCCTGTCGCTCAGAGGCACCAACATTTCGGCAGATGCAGTGCAGTGAATTTGACA
CTGTTCCCTACAAGAATGAACTCTACCACTGGTTTCCCATTTTTAACCCAGCACATCCTTGTGAGCTC
TACTGCCGACCCATAGATGGCCAGTTTTCTGAGAAAATGCTGGATGCTGTCATTGATGGTACCCCTTG
CTTTGAAGGCGGCAACAGCAGAAATGTCTGTATTAATGGCATATGTAAGATGGTTGGCTGTGACTATG
AGATCGATTCCAATGCCACCGAGGATCGCTGCGGTGTGTGCCTGGGAGATGGCTCTTCCTGCCAGACT
GTGAGAAAGATGTTTAAGCAGAAGGAAGGATCTGGTTATGTTGACATTGGGCTCATTCCAAAAGGAGC
AAGGGACATAAGAGTGATGGAAATTGAGGGAGCTGGAAACTTCCTGGCCATCAGGAGTGAAGATCCTG
AAAAATATTACCTGAATGGAGGGTTTATTATCCAGTGGAACGGGAACTATAAGCTGGCAGGGACTGTC
TTTCAGTATGACAGGAAAGGAGACCTGGAAAAGCTGATGGCCACAGGTCCCACCAATGAGTCTGTGTG
GATCCAGCTTCTATTCCAGGTGACTAACCCTGGCATCAAGTATGAGTACACAATCCAGAAAGATGGCC
TTGACAATGATGTTGAGCAGCAGATGTACTTCTGGCAGTACGGCCACTGGACAGAGTGCAGTGTGACC
TGCGGGACAGGTATCCGCCGCCAAACTGCCCATTGCATAAAGAAGGGCCGCGGGATGGTGAAAGCTAC
ATTCTGTGACCCAGAAACACAGCCCAATGGGAGACAGAAGAAGTGCCATGAAAAGGCTTGTCCACCCA
GGTGGTGGGCAGGGGAGTGGGAAGCATGCTCGGCGACATGCGGGCCCCACGGGGAGAAGAAGCGAACC
GTGCTGTGCATCCAGACCATGGTCTCTGACGAGCAGGCTCTCCCGCCCACAGACTGCCAGCACCTGCT
GAAGCCCAAGACCCTCCTTTCCTGCAACAGAGACATCCTGTGCCCCTCGGACTGGACAGTGGGCAACT
GGAGTGAGTGTTCTGTTTCCTGTGGTGGTGGAGTGCGGATTCGCAGTGTCACATGTGCCAAGAACCAT
GATGAACCTTGCGATGTGACAAGGAAACCCAACAGCCGAGCTCTGTGTGGCCTCCAGCAATGCCCTTC
TAGCCGGAGAGTTCTGAAACCAAACAAAGGCACTATTTCCAATGGAAAAAACCCACCAACACTAAAGC
CCGTCCCTCCACCTACATCCAGGCCCAGAATGCTGACCACACCCACAGGGCCTGAGTCTATGAGCACA
AGCACTCCAGCAATCAGCAGCCCTAGTCCTACCACAGCCTCCAAAGAAGGAGACCTGGGTGGGAAACA
GTGGCAAGATAGCTCAACCCAACCTGAGCTGAGCTCTCGCTATCTCATTTCCACTGGAAGCACTTCCC
AGCCCATCCTCACTTCCCAATCCTTGAGCATTCAGCCAAGTGAGGAAAATGTTTCCAGTTCAGATACT
GGTCCTACCTCGGAGGGAGGCCTTGTAGCTACAACAACAAGTGGTTCTGGCTTGTCATCTTCCCGCAA

CCCTATCACTTGGCCTGTGACTCCATTTTACAATACCTTGACCAAAGGTCCAGAAATGGAGATTCACA
GTGGCTCAGGGGAAGAAAGAGAACAGCCTGAGGACAAAGATGAAAGCAATCCTGTAATATGGACCAAG
ATCAGAGTACCTGGAAATGACGCTCCAGTGGAAAGTACAGAAATGCCACTTGCACCTCCACTAACACC
AGATCTCAGCAGGGAGTCCTGGTGGCCACCCTTCAGCACAGTAATGGAAGGACTGCTCCCCAGCCAAA
GGCCCACTACTTCCGAAACTGGGACACCCAGAGTTGAGGGGATGGTTACTGAAAAGCCAGCCAACACT
CTGCTCCCTCTGGGAGGAGACCACCAGCCAGAACCCTCAGGAAAGACGGCAAACCGTAACCACCTGAA
ACTTCCAAACAACATGAACCAAACAAAAAGTTCTGAACCAGTCCTGACTGAGGAGGATGCAACAAGTC
TGATTACTGAGGGCTTTTTGCTAAATGCCTCCAATTACAAGCAGCTCACAAACGGCCACGGCTCTGCA
CACTGGATCGTCGGAAACTGGAGCGAGTGCTCCACCACATGTGGCCTGGGGGCCTACTGGAGAAGGGT
GGAGTGCAGCACCCAGATGGATTCTGACTGTGCGGCCATCCAGAGACCTGACCCTGCAAAAAGATGCC
ACCTCCGTCCCTGTGCTGGCTGGAAAGTGGGAAACTGGAGCAAGTGCTCCAGAAACTGCAGTGGGGGC
TTCAAGATACGCGAGATTCAGTGCGTGGACAGCCGGGACCACCGGAACCTGAGGCCATTTCACTGCCA
GTTCCTGGCCGGCATTCCTCCCCCATTGAGCATGAGCTGTAACCCGGAGCCCTGTGAGGCGTGGCAGG
TGGAGCCTTGGAGCCAGTGCTCCAGGTCCTGTGGAGGTGGAGTTCAGGAGAGAGGAGTGTTCTGTCCA
GGAGGCCTCTGTGATTGGACAAAAAGACCCACATCCACCATGTCTTGCAATGAGCACCTGTGCTGTCA
CTGGGCCACTGGGAACTGGGACCTGTGTTCCACTTCCTGTGGAGGTGGCTTTCAGAAGAGGACTGTCC
AATGTGTGCCCTCAGAGGGCAATAAAACTGAAGACCAAGACCAATGTCTATGTGATCACAAACCCAGA
CCTCCAGAATTCAAAAAATGCAACCAGCAGGCCTGCAAGAAAAGTGCCGATTTACTTTGCACTAAGGA
CAAACTGTCAGCCAGTTTCTGCCAGACACTGAAAGCCATGAAGAAATGTTCTGTGCCCACCGTGAGGG
CTGAGTGCTGCTTCTCGTGTCCCCAGACACACATCACACACACCCAAAGGCAAAGAAGGCAACGGTTG
CTCCAAAAGTCAAAAGGACTCCTCGAGGAGTTCTTTT**TGA**CTTTTGGAGCAACCGGGCCTGGCTGAGGC
TGTCTCCCATAAGATCT

NOV68k, CG58504-05 Protein Sequence	SEQ ID NO: 1026	1574 aa	MW at 175579.4kD

YGRQPQPGPVRFPDRRQEHFIKGLPEYHVVGPVRVDASGHFLSYGLHYPITSSRRKRDLDGSEDWVYY
RISHEEKDLFFNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACH
GLTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRRQKVPETKEPTCGLKDSVNISQKQELWREKWER
HNLPSRSLSRRSISKERWVETLVVADTKMIEYHGSENVESYILTIMNMVTGLFHNPSIGNAIHIVVVR
LILLEEEEQGLKIVHHAEKTLSSFCKWQKSINPKSDLNPVHHDVAVLLTRKDICAGFNRPCETLGLSH

LSGMCQPHRSCNINEDSGLPLAFTIAHELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPTPLTWS
KCSEEYITRFLDRGWGFCLDDIPKKKGLKSKVIAPGVIYDVHHQCQLQYGPNATFCQEVENVCQTLWC
SVKGFCRSKLDAAADGTQCGEKKWCMAGKCITVGKKPESIPGGWGRWSPWSHCSRTCGAGVQSAERLC
NNPEPKFGGKYCTGERKRYRLCNVHPCRSEAPTFRQMQCSEFDTVPYKNELYHWFPIFNPAHPCELYC
RPIDGQFSEKMLDAVIDGTPCFEGGNSRNVCINGICKMVGCDYEIDSNATEDRCGVCLGDGSSCQTVR
KMFKQKEGSGYVDIGLIPKGARDIRVMEIEGAGNFLAIRSEDPEKYYLNGGFIIQWNGNYKLAGTVFQ
YDRKGDLEKLMATGPTNESVWIQLLFQVTNPGIKYEYTIQKDGLDNDVEQQMYFWQYGHWTECSVTCG
TGIRRQTAHCIKKGRGMVKATFCDPETQPNGRQKKCHEKACPPRWWAGEWEACSATCGPHGEKKRTVL
CIQTMVSDEQALPPTDCQHLLKPKTLLSCNRDILCPSDWTVGNWSECSVSCGGGVRIRSVTCAKNHDE
PCDVTRKPNSRALCGLQQCPSSRRVLKPNKGTISNGKNPPTLKPVPPPTSRPRMLTTPTGPESMSTST
PAISSPSPTTASKEGDLGGKQWQDSSTQPELSSRYLISTGSTSQPILTSQSLSIQPSEENVSSSDTGP
TSEGGLVATTTSGSGLSSSRNPITWPVTPFYNTLTKGPEMEIHSGSGEEREQPEDKDESNPVIWTKIR
VPGNDAPVESTEMPLAPPLTPDLSRESWWPPFSTVMEGLLPSQRPTTSETGTPRVEGMVTEKPANTLL
PLGGDHQPEPSGKTANRNHLKLPNNMNQTKSSEPVLTEEDATSLITEGFLLNASNYKQLTNGHGSAHW
IVGNWSECSTTCGLGAYWRRVECSTQMDSDCAAIQRPDPAKRCHLRPCAGWKVGNWSKCSRNCSGGFK
IREIQCVDSRDHRNLRPFHCQFLAGIPPPLSMSCNPEPCEAWQVEPWSQCSRSCGGGVQERGVFCPGG
LCDWTKRPTSTMSCNEHLCCHWATGNWDLCSTSCGGGFQKRTVQCVPSEGNKTEDQDQCLCDHKPRPP
EFKKCNQQACKKSADLLCTKDKLSASFCQTLKAMKKCSVPTVRAECCFSCPQTHITHTQRQRRQRLLQ
KSKGLLEEFF

NOV68l, CG58504-06 DNA Sequence	SEQ ID NO: 1027	1068 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTAACTTGACGGTCAATCAAGGATTTCTTTCCAATAGCTACATCATGGAGAAGAGATATGGGAA
CCTCTCCCATGTTAAGATGATGGCTTCCTCTGCCCCCCTCTGCCATCTCAGTGGCACGGTTCTACAGC
AGGGCACCAGAGTTGGGACGGCAGCCCTCAGTGCCTGCCATGGACTGACTGGATTTTTCCAACTACCA
CATGGAGACTTTTTCATTGAACCCGTGAAGAAGCATCCACTGGTTGAGGGAGGGTACCACCCGCACAT
CGTTTACAGGAGGCAGAAAGTTCCAGAAACCAAGGAGCCAACCTGTGGATTAAAGGACAGTGTTAACA
TCTCCCAGAAGCAAGAGCTATGGCGGGAGAAGTGGGAGAGGCACAACTTGCCAAGCAGAAGCCTCTCT
CGGCGTTCCATCAGCAAGGAGAGATGGGTGGAGACACTGGTGGTGGCCGACACAAAGATGATTGAATA
CCATGGGAGTGAGAATGTGGAGTCCTACATCCTCACCATCATGAACATGGTCACTGGGTTGTTCCATA
ACCCAAGCATTGGCAATGCAATTCACATTGTTGTGGTTCGGCTCATTCTACTCGAAGAAGAAGAGCAA
GGACTGAAAATAGTTCACCATGCAGAAAAGACACTGTCTAGCTTCTGCAAGTGGCAGAAGAGTATCAA
TCCCAAGAGTGACCTCAATCCTGTTCATCACGACGTGGCTGTCCTTCTCACCAGAAAGGACATCTGTG
CTGGTTTCAATCGCCCCTGCGAGACCCTGGGCCTGTCTCACCTTTCAGGAATGTGTCAGCCTCACCGC
AGTTGTAACATCAATGAAGATTCGGGACTCCCTCTGGCTTTCACAATTGCCCATGAGCTAGGACACAG
CTTCGGCATCCAGCATGATGGGAAAGAAAATGACTGTGAGCCTGTGGGCAGACATCCGTACATCATGT
CCCGCCAGCTCCAGTACGATCCCACTCCGCTGACATGGTCCAAGTGCAGCGAGGAGTACATCACCCGC
TTCTTGGACCGAGGCTGGGGGTTCTGTCTTGATGACATACCTCTCGAG

NOV68l, CG58504-06 Protein Sequence	SEQ ID NO: 1028	356 aa	MW at 40366.8kD

KLNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHGLTGFFQLP
HGDFFIEPVKKHPLVEGGYHPHIVYRRQKVPETKEPTCGLKDSVNISQKQELWREKWERHNLPSRSLS
RRSISKERWVETLVVADTKMIEYHGSENVESYILTIMNMVTGLFHNPSIGNAIHIVVVRLILLEEEEQ
GLKIVHHAEKTLSSFCKWQKSINPKSDLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHR
SCNINEDSGLPLAFTIAHELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPTPLTWSKCSEEYITR
FLDRGWGFCLDDIPLE

NOV68m, CG58504-07 DNA Sequence	SEQ ID NO: 1029	252 bp
	ORF Start: at 1	ORF Stop: end of sequence

AAGCTTCACCAGTGCCAGCTACAATATGGACCCAATGCTACCTTCTGCCAGGAAGTAGAAAACGTCTG
CCAGACACTGTGGTGCTCCGTGAAGGGCTTTTGTCGCTCTAAGCTGGACGCTGCTGCAGATGGAACTC
AATGTGGTGAGAAGAAGTGGTGTATGGCAGGCAAGTGCATCACAGTGGGGAAGAAACCAGAGAGCATT
CCTGGAGGCTGGGGCCGCTGGTCACCCTGGTCCCACTGTTCCCTCGAG

NOV68m, CG58504-07 Protein Sequence	SEQ ID NO: 1030	84 aa	MW at 9323.7kD

KLHQCQLQYGPNATFCQEVENVCQTLWCSVKGFCRSKLDAAADGTQCGEKKWCMAGKCITVGKKPESI
PGGWGRWSPWSHCSLE

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 68B.

Table 68B. Comparison of the NOV68 protein sequences.

```
NOV68a  MPCAQRSWLANLSVVAQLLNFGALCYGRQPQPGPVRFPDRRQEHFIKGLPEYHVVGPVRV
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  ------------------------------------------------------------
NOV68f  ------------------------------------------------------------
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  ------------------------------------------------------------
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  DASGHFLSYGLHYPITSSRRKRDLDGSEDWVYYRISHEEKDLFFNLTVNQGFLSNSYIME
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  ------------------------------------------------------------
NOV68f  ------------------------------------------------------------
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  ------------------------------------------------------------
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  KRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHGLTGFFQLPHGDFFIEPVK
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  ------------------------------------------------------------
NOV68f  ------------------------------------------------------------
NOV68g  ------------------------------LECALRSGRWSPWSHCSRTCGAGVQSAE
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  ------------------------------------------------------------
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  KHPLVEGGYHPHIVYRRQKVPETKEPTCGLKDSVNISQKQELWREKWERHNLPSRSLSRR
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  ------------------------------------------------------------
NOV68f  ------------------------------------------------------------
NOV68g  RLCNNPEPKFGGKYCTGERKRYRLCNVHPCRSEAPTFRQMQCSEFDTVPYKNELYHWFPI
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
```

```
NOV68k  ------------------------------------------------------------
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  SISKERWVETLVVADTKMIEYHGSENVESYILTIMNMVTGLFHNPSIGNAIHIVVVRLIL
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  ------------------------------------------------------------
NOV68f  ------------------------------------------------------------
NOV68g  FNPAHPCELYCRPIDGQFSEKMLDAVIDGTPCFEGGNSRNVCINGICKMVGCDYEIDSNA
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  ------------------------------------------------------------
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  LEEEEQGLKIVHHAEKTLSSFCKWQKSINPKSDLNPVHHDVAVLLTRKDICAGFNRPCET
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  ------------------------------------------RSYGRQPQPGPVRFPDRRQ
NOV68f  ------------------------------------------RSYGRQPQPGPVRFPDRRQ
NOV68g  TEDRCGVCLGDGSSCQTARKMFKQKEGSGYVDIGLIPKGARDIRVMEIEGAGNFLAIRSE
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  ------------------------------------------YGRQPQPGPVRFPDRRQ
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  LGLSHLSGMCQPHRSCNINEDSGLPLAFTIAHELGHSFGIQHDGKENDCEPVGRHPYIMS
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  EHFIKGLPEYHVVGPVRVDASGHFLSYGLHYPITSSRRKRDLDGSEDWVYYRISHEEKDL
NOV68f  EHFIKGLPEYHVVGPVRVDASGHFLSYGLHYPITSSRRKRDLDGSEDWVYYRISHEEKDL
NOV68g  DPEKYYLNGGFIIQWNGNYKLAGTVFQYDRKGDLEKLMATGPTNESVWIQLLFQVTNPGI
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  EHFIKGLPEYHVVGPVRVDASGHFLSYGLHYPITSSRRKRDLDGSEDWVYYRISHEEKDL
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  RQLQYDPTPLTWSKCSEEYITRFLDRGWGFCLDDIPKKKGLKSKVIAPGVIYDVHHQCQL
NOV68b  KLNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGMAALSACHG
NOV68c  KLNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHG
NOV68d  KLNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHG
NOV68e  FFNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLNGTVLQQGTRVGTAALSACHG
NOV68f  FFNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHG
NOV68g  KYEYTIQKDGLDNDVEQQMYFWQYGHWTECSVTCGTGIRRQTAHCIKKGRGMVKATFCDP
NOV68h  --NLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGMAALSACHG
NOV68i  --NLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHG
NOV68j  ------------------------------------------------------HQCQL
NOV68k  FFNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHG
```

```
NOV68l  KLNLTVNQGFLSNSYIMEKRYGNLSHVKMMASSAPLCHLSGTVLQQGTRVGTAALSACHG
NOV68m  ----------------------------------------------------KLHQCQL

NOV68a  QYGPNATFCQEVENVCQTLWCSVKGFCRSKLDAAA-----------------------
NOV68b  LTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRR-----------------------
NOV68c  LTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRR-----------------------
NOV68d  LTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRR-----------------------
NOV68e  LTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRR-----------------------
NOV68f  LTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRR-----------------------
NOV68g  ETQPNGRQKKCHEKACPPRWWAGEWEACSATCGPHGEKKRTVLCIQTMVSDEQALPPTDC
NOV68h  LTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRR-----------------------
NOV68i  LTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRR-----------------------
NOV68j  QYGPNATFCQEVENVCQTLWCSVKGFCRSKLDAAA-----------------------
NOV68k  LTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRR-----------------------
NOV68l  LTGFFQLPHGDFFIEPVKKHPLVEGGYHPHIVYRR-----------------------
NOV68m  QYGPNATFCQEVENVCQTLWCSVKGFCRSKLDAAA-----------------------

NOV68a  ----------------------------------------------------DGTQCGEKK-
NOV68b  ----------------------------------------------------QKVPETKEP-
NOV68c  ----------------------------------------------------QKVPETKEP-
NOV68d  ----------------------------------------------------QKVPETKEP-
NOV68e  ----------------------------------------------------QKVPETKEP-
NOV68f  ----------------------------------------------------QKVPETKEP-
NOV68g  QHLLKPKTLLSCNRDILCPSDWTVGNWSECSVSCGGGVRIRSVTCAKNHDEPCDVTRKPN
NOV68h  ----------------------------------------------------QKVPETKEP-
NOV68i  ----------------------------------------------------QKVPETKEP-
NOV68j  ----------------------------------------------------DGTQCGEKK-
NOV68k  ----------------------------------------------------QKVPETKEP-
NOV68l  ----------------------------------------------------QKVPETKEP-
NOV68m  ----------------------------------------------------DGTQCGEKK-

NOV68a  ---WCMAGKC------------------------------------------------------
NOV68b  ---TCGLKDS------------------------------------------------------
NOV68c  ---TCGLKDS------------------------------------------------------
NOV68d  ---TCGLKDS------------------------------------------------------
NOV68e  ---TCGLKDS------------------------------------------------------
NOV68f  ---TCGLKDS------------------------------------------------------
NOV68g  SRALCGLQQCPSSRRVLKPNKGTISNGKNPPTLKPVPPPTSRPRMLTTPTGPESMSTSTP
NOV68h  ---TCGLKDS------------------------------------------------------
NOV68i  ---TCGLKDS------------------------------------------------------
NOV68j  ---WCMAGKC------------------------------------------------------
NOV68k  ---TCGLKDS------------------------------------------------------
NOV68l  ---TCGLKDS------------------------------------------------------
NOV68m  ---WCMAGKC------------------------------------------------------

NOV68a  -------ITVGKKPESIPGGWGRWSPWSHCSRTCGAGVQSAERLCNNPEPKFGGKYCT--
NOV68b  -------VNISQKQELWREKWERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGS--
NOV68c  -------VNISQKQELWREKWERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGS--
NOV68d  -------VNISQKQELWREKWERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGS--
NOV68e  -------VNISQKQELWREKWERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGS--
NOV68f  -------VNISQKQELWREKWERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGS--
NOV68g  AISSPSPTTASKEGDLGGKQWQDSSTQPELSSRYLISTGSTSQPILTSQSLSIQPSEENV
NOV68h  -------VNISQKQELWREKWERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGS--
NOV68i  -------VNISQKQELWREKWERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGS--
NOV68j  -------ITVGKKPESIPGGCGRWSPWSHCS--------------------------------
NOV68k  -------VNISQKQELWREKWERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGS--
NOV68l  -------VNISQKQELWREKWERHNLPSRSLSRRSISKERWVETLVVADTKMIEYHGS--
```

```
NOV68m  -------ITVGKKPESIPGGWGRWSPWSHCSLE--------------------------

NOV68a  -------------------GERKRYRLCNVHPCRSEAPTFRQMQCSEFDTVPYKNELYH
NOV68b  -------------------ENVESCILTIMNMVTGLFHNPSIGNAIHIVVVRLILLEEE
NOV68c  -------------------ENVESYILTIMNMITGLFHNPSIGNAIHIVVVRLILLEEE
NOV68d  -------------------ENVESYILTIMNMVTGLFHNPSIGNAIHIVVVRLILLEEE
NOV68e  -------------------ENVESYILTIMNMVTGLFHNPSIGNAIHIVVVRLILLEEE
NOV68f  -------------------ENVESYILTIMNMVTGLFHNPSIGNAIHIVVVRLILLEEE
NOV68g  SSSDTGPTSEGGLVATTTSGSGLSSSRNPITWPVTPFYNTLTKGPEMEIHSGSGEEREQP
NOV68h  -------------------ENVESCILTIMNMVTGLFHNPSIGNAIHIVVVRLILLEEE
NOV68i  -------------------ENVESYILTIMNMITGLFHNPSIGNAIHIVVVRLILLEEE
NOV68j  ------------------------------------------------------------
NOV68k  -------------------ENVESYILTIMNMVTGLFHNPSIGNAIHIVVVRLILLEEE
NOV68l  -------------------ENVESYILTIMNMVTGLFHNPSIGNAIHIVVVRLILLEEE
NOV68m  ------------------------------------------------------------

NOV68a  WFP---------------------------------------------------------
NOV68b  EQG---------------------------------------------------------
NOV68c  EQG---------------------------------------------------------
NOV68d  EQG---------------------------------------------------------
NOV68e  EQG---------------------------------------------------------
NOV68f  EQG---------------------------------------------------------
NOV68g  EDKDESNPVIWTKIRVPGNDAPVESTEMPLAPPLTPDLSRESWWPPFSTVMEGLLPSQRP
NOV68h  EQG---------------------------------------------------------
NOV68i  EQG---------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  EQG---------------------------------------------------------
NOV68l  EQG---------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  ------------------------------------------IFNPAHPCELYCR-----
NOV68b  ------------------------------------------LKIVHHAEKTLSS-----
NOV68c  ------------------------------------------LKIVHHAEKTLSS-----
NOV68d  ------------------------------------------LKIVHHAEKTLSS-----
NOV68e  ------------------------------------------LKIVHHAEKTLSS-----
NOV68f  ------------------------------------------LKIVHHAEKTLSS-----
NOV68g  TTSETGTPRVEGMVTEKPANTLLPLGGDHQPEPSGKTANRNHLKLPNNMNQTKSSEPVLT
NOV68h  ------------------------------------------LKIVHHAEKTLSS-----
NOV68i  ------------------------------------------LKIVHHAEKTLSS-----
NOV68j  ------------------------------------------------------------
NOV68k  ------------------------------------------LKIVHHAEKTLSS-----
NOV68l  ------------------------------------------LKIVHHAEKTLSS-----
NOV68m  ------------------------------------------------------------

NOV68a  ------------------------------------------------PIDGQFSEKMLD
NOV68b  ------------------------------------------------FCKWQKSINPKS
NOV68c  ------------------------------------------------FCKWQKSINPKS
NOV68d  ------------------------------------------------FCKWQKSINPKS
NOV68e  ------------------------------------------------FCKWQKSINPKS
NOV68f  ------------------------------------------------FCKWQKSINPKS
NOV68g  EEDATSLITEGFLLNASNYKQLTNGYGSAHWIVGNWSECSTTCGLGAYWRRVECSTQMDS
NOV68h  ------------------------------------------------FCKWQKSINPKS
NOV68i  ------------------------------------------------FCKWQKSINPKS
NOV68j  ------------------------------------------------------------
NOV68k  ------------------------------------------------FCKWQKSINPKS
NOV68l  ------------------------------------------------FCKWQKSINPKS
NOV68m  ------------------------------------------------------------
```

```
NOV68a  AVIDGTPCFEGGNSRNVCINGICKMVGCDYEIDSNATEDRCGVCLGDG-SSCQTVRKMFK
NOV68b  DLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSCNINE-DSGLPLAFTIA
NOV68c  DLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSCNINE-DSGLPLAFTIA
NOV68d  DLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSCNINE-DSGLPLAFTIA
NOV68e  DLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSCNINE-DSGLPLAFTIA
NOV68f  DLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSCNINE-DSGLPLAFTIA
NOV68g  DCAAIQRPDPAKRCHLRPCAGWKVGNWSKCSRNCSGGFKIREIQCVDSRDHRNLRPFHCQ
NOV68h  DLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSCNINE-DSGLPLAFTIA
NOV68i  DLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSCNINE-DSGLPLAFTIA
NOV68j  ------------------------------------------------------------
NOV68k  DLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSCNINE-DSGLPLAFTIA
NOV68l  DLNPVHHDVAVLLTRKDICAGFNRPCETLGLSHLSGMCQPHRSCNINE-DSGLPLAFTIA
NOV68m  ------------------------------------------------------------

NOV68a  QKEGSGYVDIGLIPKGARDIRVMEIEGAGNFLAIRSE----------------DPEKYY
NOV68b  HELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPT----------------PLTWSK
NOV68c  HELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPT----------------PLTWSK
NOV68d  HELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPT----------------PLTWSK
NOV68e  HELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPT----------------PLTWSK
NOV68f  HELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPT----------------PLTWSK
NOV68g  FLAGIPPPLSMSCNPEPCEAWQVEPWSQCSRSCGGGVQERGVFCPGGLCDWTKRPTSTMS
NOV68h  HELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPT----------------PLTWSK
NOV68i  HELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPT----------------PLTWSK
NOV68j  ------------------------------------------------------------
NOV68k  HELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPT----------------PLTWSK
NOV68l  HELGHSFGIQHDGKENDCEPVGRHPYIMSRQLQYDPT----------------PLTWSK
NOV68m  ------------------------------------------------------------

NOV68a  LNGGFIIQWNGNYKLAGTVFQYDRKGDLEKLMATGPTNESVWIQLLFQVTNPGIKYEYTI
NOV68b  CSEEYITRFLDRGWGFCLDDIPLE------------------------------------
NOV68c  CSEEYITRFLDRGWGFCLDDIPLE------------------------------------
NOV68d  CSEEYITRFLDRGWGFCLDDIPLE------------------------------------
NOV68e  CSEEYITRFLDRGWGFCLDDIPKKKGLKSKVIAPGVIYDVHHQCQLQYGPNATLCQEVEN
NOV68f  CSEEYITRFLDRGWGFCLDDIPKKKGLKSKVIAPGVIYDVHHQCQLQYGPNATFCQEVEN
NOV68g  CNEHLCCHWATGNWDLCSTSCGGGFQKRTVQCVPSEGNKTEDQDQCLCDHKPRPPEFKKC
NOV68h  CSEEYITRFLDRGWGFCLDDIP--------------------------------------
NOV68i  CSEEYITRFLDRGWGFCLDDIP--------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  CSEEYITRFLDRGWGFCLDDIPKKKGLKSKVIAPGVIYDVHHQCQLQYGPNATFCQEVEN
NOV68l  CSEEYITRFLDRGWGFCLDDIPLE------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  QKDGLDNDVEQMYFWQYGHWTECSVTCGTGIRRQTAHCIKKGRGMVKATFCDPETQPNGR
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  VCQTLWCSVKGFCRSKLDAAADGTQCGEKKWCMAGKCITVGKKPESIPGGWGRWSPWSHC
NOV68f  VCQTLWCSVKGFCRSKLDAAADGTQCGEKKWCMAGKCITVGKKPESIPGGWGRWSPWSHC
NOV68g  NQQACKKSADLLCTKDKLSASFCQTLKAMKKCSVPTVRAECCFSCPQTHITHTQRQRRQR
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  VCQTLWCSVKGFCRSKLDAAADGTQCGEKKWCMAGKCITVGKKPESIPGGWGRWSPWSHC
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------
```

```
NOV68a  QKKCHEKACPPRWWAGEWEACSATCGPHGEKKRTVLCIQTMVSDEQALPPTDCQHLLKPK
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  SRTCGAGVQSAERLCNNPEPKFGGKYCTGERKRYRLCNVHPCRSEAPTFRQMQCSEFDTV
NOV68f  SRTCGAGVQSAERLCNNPEPKFGGKYCTGERKRYRLCNVHPCRSEAPTFRQMQCSEFDTV
NOV68g  LLQKSKELLE--------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  SRTCGAGVQSAERLCNNPEPKFGGKYCTGERKRYRLCNVHPCRSEAPTFRQMQCSEFDTV
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  TLLSCNRDILCPSDWTVGNWSECSVSCGGGVRIRSVTCAKNHDEPCDVTRKPNSRALCGL
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  PYKNELYHWFPIFNPAHPCELYCRPIDGQFSEKMLDAVIDGTPCFEGGNSRNVCINGICK
NOV68f  PYKNELYHWFPIFNPAHPCELYCRPIDGQFSEKMLDAVIDGTPCFEGGNSRNVCINGICK
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  PYKNELYHWFPIFNPAHPCELYCRPIDGQFSEKMLDAVIDGTPCFEGGNSRNVCINGICK
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  QQCPSSRRVLKPNKGTISNGKNPPTLKPVPPPTSRPRMLTTPTGPESMSTSTPAISSPSP
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  MVGCDYEIDSNATEDRCGVCLGDGSSCQTVRKMFKQKEGSGYVDIGLIPKGARDIRVMEI
NOV68f  MVGCDYEIDSNATEDRCGVCLGDGSSCQTVRKMFKQKEGSGYVDIGLIPKGARDIRVMEI
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  MVGCDYEIDSNATEDRCGVCLGDGSSCQTVRKMFKQKEGSGYVDIGLIPKGARDIRVMEI
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  TTASKEGDLGGKQWQDSSTQPELSSRYLISTGSTSQPILTSQSLSIQPSEENVSSSDTGP
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  EGAGNFLAIRSEDPEKYYLNGGFIIQWNGNYKLAGTVFQYDRKGDLEKLMATGPTNESVW
NOV68f  EGAGNFLAIRSEDPEKYYLNGGFIIQWNGNYKLAGTVFQYDRKGDLEKLMATGPTNESVW
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  EGAGNFLAIRSEDPEKYYLNGGFIIQWNGNYKLAGTVFQYDRKGDLEKLMATGPTNESVW
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  TSEGGLVATTTSGSGLSSSRNPITWPVTPFYNTLTKGPEMEIHSGSGEEREQPEDKDESN
```

```
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  IQLLFQVTNPGIKYEYTIQKDGLDNDVEQQMYFWQYGHWTECSVTCGTGIRRQTAHCIKK
NOV68f  IQLLFQVTNPGIKYEYTIQKDGLDNDVEQQMYFWQYGHWTECSVTCGTGIRRQTAHCIKK
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  IQLLFQVTNPGIKYEYTIQKDGLDNDVEQQMYFWQYGHWTECSVTCGTGIRRQTAHCIKK
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  PVIWTKIRVPGNDAPVESTEMPLAPPLTPDLSRESWWPPFSTVMEGLLPSQRPTTSETGT
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  GRGMVKATFCDPETQPNGRQKKCHEKACPPRWWAGEWEACSATCGPHGEKKRTVLCIQTM
NOV68f  GRGMVKATFCDPETQPNGRQKKCHEKACPPRWWAGEWEACSATCGPHGEKKRTVLCIQTM
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  GRGMVKATFCDPETQPNGRQKKCHEKACPPRWWAGEWEACSATCGPHGEKKRTVLCIQTM
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  PRVEGMVTEKPANTLLPLGGDHQPEPSGKTANRNHLKLPNNMNQTKSSEPVLTEEDATSL
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  VSDEQALPPTDCQHLLKPKTLLSCNRDILCPSDWTVGNWSECSVSCGGGVRIRSVTCAKN
NOV68f  VSDEQALPPTDCQHLLKPKTLLSCNRDILCPSDWTVGNWSECSVSCGGGVRIRSVTCAKN
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  VSDEQALPPTDCQHLLKPKTLLSCNRDILCPSDWTVGNWSECSVSCGGGVRIRSVTCAKN
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  ITEGFLLNASNYKQLTNGHGSAHWIVGNWSECSTTCGLGAYWKRVECTTQMDSDCAAIQR
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  HDEPCDVTRKPNSRALCGLQQCPSSRRVLKPNKGTISNGKNPPTLKPVPPPTSRPRMLTT
NOV68f  HDEPCDVTRKPNSRALCGLQQCPSSRRVLKPNKGTISNGKNPPTLKPVPPPTSRPRMLTT
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  HDEPCDVTRKPNSRALCGLQQCPSSRRVLKPNKGTISNGKNPPTLKPVPPPTSRPRMLTT
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  PDPAKRCHLRPCAGWKVGNWSKCSRNCSGGFKIREIQCVDSRDHRNLRPFHCQFLAGIPP
NOV68b  ------------------------------------------------------------
```

```
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  PTGPESMSTSTPAISSPSPTTASKEGDLGGKQWQDSSTQPELSSRYLISTGSTSQPILTS
NOV68f  PTGPESMSTSTPAISSPSPTTASKEGDLGGKQWQDSSTQPELSSRYLISTGSTSQPILTS
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  PTGPESMSTSTPAISSPSPTTASKEGDLGGKQWQDSSTQPELSSRYLISTGSTSQPILTS
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  PLSMSCNPEPCEAWQVEPWSQCSRSCGGGVQERGVFCPGGLCDWTKRPTSTMSCNEHLCC
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  QSLSIQPSEENVSSSDTGPTSEGGLVATTTSGSGLSSSRNPITWPVTPFYNTLTKGPEME
NOV68f  QSLSIQPSEENVSSSDTGPTSEGGLVATTTSGSGLSSSRNPITWPVTPFYNTLTKGPEME
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  QSLSIQPSEENVSSSDTGPTSEGGLVATTTSGSGLSSSRNPITWPVTPFYNTLTKGPEME
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  HWATGNWDLCSTSCGGGFQKRIVQCVPSEGNKTEDQDQCLCDHKPRPPEFKKCNQQACKK
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  IHSGSGEEREQPEDKDESNPVIWTKIRVPGNDAPVESTEMPLAPPLTPDLSRESWWPPFS
NOV68f  IHSGSGEEREQPEDKDESNPVIWTKIRVPGNDAPVESTEMPLAPPLTPDLSRESWWPPFS
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  IHSGSGEEREQPEDKDESNPVIWTKIRVPGNDAPVESTEMPLAPPLTPDLSRESWWPPFS
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  SADLLCTKDKLSASFCQTLKAMKKCSVPTVRAECCFSCPQTHITHTQRQRRQRLLQKSKE
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  TVMEGLLPSQRPTTSETGTPRVEGMVTEKPANTLLPLGGDHQPEPSGKTANRNHLKLPNN
NOV68f  TVMEGLLPSQRPTTSETGTPRVEGMVTEKPANTLLPLGGDHQPEPSGKTANRNHLKLPNN
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  TVMEGLLPSQRPTTSETGTPRVEGMVTEKPANTLLPLGGDHQPEPSGKTANRNHLKLPNN
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  L-----------------------------------------------------------
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
```

```
NOV68d  ------------------------------------------------------------
NOV68e  MNQTKSSEPVLTEEDATSLITEGFLLNASNYKQLTNGHGSAHWIVGNWSECSTTCGLGAY
NOV68f  MNQTKSSEPVLTEEDATSLITEGFLLNASNYKQLTNGHGSAHWIVGNWSECSTTCGLGAY
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  MNQTKSSEPVLTEEDATSLITEGFLLNASNYKQLTNGHGSAHWIVGNWSECSTTCGLGAY
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  ------------------------------------------------------------
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  WRRVECSTQMDSDCAAIQRPDPAKRCHLRPCAGWKVGNWSKCSRNCSGGFKIREIQCVDS
NOV68f  WRRVECSTQMDSDCAAIQRPDPAKRCHLRPCAGWKVGNWSKCSRNCSGGFKIREIQCVDS
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  WRRVECSTQMDSDCAAIQRPDPAKRCHLRPCAGWKVGNWSKCSRNCSGGFKIREIQCVDS
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  ------------------------------------------------------------
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  RDHRNLRPFHCQFLAGIPPPLSMSCNPEPCEAWQVEPWSQCSRSCGGGVQERGVFCPGGL
NOV68f  RDHRNLRPFHCQFLAGIPPPLSMSCNPEPCEAWQVEPWSQCSRSCGGGVQERGVFCPGGL
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  RDHRNLRPFHCQFLAGIPPPLSMSCNPEPCEAWQVEPWSQCSRSCGGGVQERGVFCPGGL
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  ------------------------------------------------------------
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
NOV68e  CDWTKRPTSTMSCNEHLCCHWATGNWDLCSTSCGGGFQKRTVQCVPSEGNKTEDQDQCLC
NOV68f  CDWTKRPTSTMSCNEHLCCHWATGNWDLCSTSCGGGFQKRTVQCVPSEGNKTEDQDQCLC
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  CDWTKRPTSTMSCNEHLCCHWATGNWDLCSTSCGGGFQKRTVQCVPSEGNKTEDQDQCLC
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  ------------------------------------------------------------
NOV68b  ------------------------------------------------------------
NOV68c  ------------------------------------------------------------
NOV68d  ------------------------------------------------------------
```

```
NOV68e  DHKPRPPEFKKCNQQACKKSADLLCTKDKLSASFCQTLKAMKKCSVPTVRAECCFSCPQT
NOV68f  DHKPRPPEFKKCNQQACKKSADLLCTKDKLSASFCQTLKAMKKCSVPTVRAECCFSCPQT
NOV68g  ------------------------------------------------------------
NOV68h  ------------------------------------------------------------
NOV68i  ------------------------------------------------------------
NOV68j  ------------------------------------------------------------
NOV68k  DHKPRPPEFKKCNQQACKKSADLLCTKDKLSASFCQTLKAMKKCSVPTVRAECCFSCPQT
NOV68l  ------------------------------------------------------------
NOV68m  ------------------------------------------------------------

NOV68a  -------------------------
NOV68b  -------------------------
NOV68c  -------------------------
NOV68d  -------------------------
NOV68e  HITHTQRQRRQRLLQKSKELLEEFF
NOV68f  HITHTQRQRRQRLLQKSKGLLEEFF
NOV68g  -------------------------
NOV68h  -------------------------
NOV68i  -------------------------
NOV68j  -------------------------
NOV68k  HITHTQRQRRQRLLQKSKGLLEEFF
NOV68l  -------------------------
NOV68m  -------------------------

NOV68a  (SEQ ID NO:  1006)
NOV68b  (SEQ ID NO:  1008)
NOV68c  (SEQ ID NO:  1010)
NOV68d  (SEQ ID NO:  1012)
NOV68e  (SEQ ID NO:  1014)
NOV68f  (SEQ ID NO:  1016)
NOV68g  (SEQ ID NO:  1018)
NOV68h  (SEQ ID NO:  1020)
NOV68i  (SEQ ID NO:  1022)
NOV68j  (SEQ ID NO:  1024)
NOV68k  (SEQ ID NO:  1026)
NOV68l  (SEQ ID NO:  1028)
NOV68m  (SEQ ID NO:  1030)
```

Further analysis of the NOV68a protein yielded the following properties shown in Table 68C.

Table 68C. Protein Sequence Properties NOV68a

SignalP analysis:	Cleavage site between residues 26 and 27

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 6;  pos.chg 1;  neg.chg 0
      H-region:  length 21;  peak value   6.23
      PSG score:   1.83

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -3.11
      possible cleavage site: between 27 and 28
```

```
>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   2
      Number of TMS(s) for threshold  0.5:   0
      PERIPHERAL  Likelihood =  4.98 (at 267)
      ALOM score:  -1.54  (number of TMSs: 0)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 6
      Charge difference:  1.0   C( 3.0) - N( 2.0)
      C > N: C-terminal side will be inside

>>>Caution: Inconsistent mtop result with signal peptide
MITDISC: discrimination of mitochondrial targeting seq
      R content:       3       Hyd Moment(75):  8.02
      Hyd Moment(95):  8.85    G content:       3
      D/E content:     1       S/T content:     2
      Score: -2.50

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 46  VRF|PD

NUCDISC: discrimination of nuclear localization signals
      pat4: RRKR (5) at   79
      pat4: PKKK (4) at  456
      pat7: PKKKGLK (5) at  456
      pat7: PNGRQKK (3) at  869
      pat7: PHGEKKR (3) at  899
      bipartite: none
      content of basic residues:  11.5%
      NLS Score:  1.04

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: KSKE

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: found
      KLPN at 1269

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none
```

```
Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 89

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       65.2 %: nuclear
       21.7 %: mitochondrial
       13.0 %: cytoplasmic

>> prediction for CG58504-01 is nuc (k=23)
```

A search of the NOV68a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 68D.

Table 68D. Geneseq Results for NOV68a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV68a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB74944	Human ADAM type metal protease MDTS1 protein SEQ ID NO:1 - Homo sapiens, 1686 aa. [JP2001008687-A, 16-JAN-2001]	24..1574 21..1673	733/1718 (42%) 941/1718 (54%)	0.0
AAE00913	Human 27875 ADAM-TS protein, alternative version - Homo sapiens, 1686 aa. [WO200131034-A1, 03-MAY-2001]	24..1574 21..1673	731/1718 (42%) 939/1718 (54%)	0.0
AAE00934	Human 27875 ADAM-TS (a disintegrin and metalloproteinase) - Homo sapiens, 1686 aa. [WO200131034-A1, 03-MAY-2001]	24..1574 21..1673	731/1718 (42%) 939/1718 (54%)	0.0
AAB86949	Human metalloprotease MPTS-19 protein - Homo sapiens, 1690 aa. [DE10107360-A1, 06-SEP-2001]	24..1574 25..1677	731/1718 (42%) 938/1718 (54%)	0.0
AAB72283	Human ADAMTS-7 amino acid sequence - Homo sapiens, 997 aa. [WO200111074-A2, 15-FEB-2001]	24..903 21..936	483/935 (51%) 609/935 (64%)	0.0

In a BLAST search of public sequence databases, the NOV68a protein was found to have homology to the proteins shown in the BLASTP data in Table 68E.

Table 68E. Public BLASTP Results for NOV68a				
Protein Accession Number	**Protein/Organism/Length**	**NOV68a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
P58397	ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12) - Homo sapiens (Human), 1593 aa.	1..1593 1..1593	1593/1593 (100%) 1593/1593 (100%)	0.0
CAD60967	Metalloprotease disintegrin 12 protein - Mus musculus (Mouse), 1600 aa.	1..1588 1..1595	1283/1600 (80%) 1399/1600 (87%)	0.0
Q8BKY1	ADAMTS-12 precursor - Mus musculus (Mouse), 1009 aa (fragment).	1..1004 1..1009	889/1009 (88%) 939/1009 (92%)	0.0
CAC38921	Sequence 2 from Patent WO0131034 - Homo sapiens (Human), 1686 aa.	24..1574 21..1673	731/1718 (42%) 939/1718 (54%)	0.0
Q9UKP4	ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7) - Homo sapiens (Human), 997 aa.	24..903 21..936	485/935 (51%) 611/935 (64%)	0.0

PFam analysis indicates that the NOV68a protein contains the domains shown in the Table 68F.

Table 68F. Domain Analysis of NOV68a			
Pfam Domain	**NOV68a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Pep_M12B_propep	105..222	27/128 (21%) 80/128 (62%)	9.2e-05
Reprolysin	246..456	66/224 (29%) 149/224 (67%)	1.3e-15
tsp_1	546..596	23/53 (43%) 38/53 (72%)	2.2e-13
tsp_1	827..881	15/64 (23%) 39/64 (61%)	0.054
tsp_1	945..995	17/58 (29%) 39/58 (67%)	7.1e-05
tsp_1	1314..1364	13/56 (23%) 29/56 (52%)	0.036
tsp_1	1426..1471	14/54 (26%) 32/54 (59%)	0.055
tsp_1	1474..1530	14/63 (22%) 40/63 (63%)	0.0045

Example 69.

The NOV69 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 69A.

Table 69A. NOV69 Sequence Analysis		
NOV69a, CG58510-01 DNA Sequence	SEQ ID NO: 1031	933 bp
	ORF Start: ATG at 8	ORF Stop: at 932

GGATTAAATGGAAAACGATCACAATCATAATCATACACATGGCGCGAACAAAAAGACTTTATTAATTA
GTTTTATTATTATTACAAGTTACATGATTGTAGAAGGGTTAGGTGGTTTCTTTACTAACAGTCTTGCG
TTAATTTCAGATGCTGGTCATATGTTAAGTGATTCTATTTCTTTAGGTATTGCTTTAATTGCATTTAC
TTTAGGAGCGAAGCAAGCTAATACAAATAAAACTTTTGGCTACAAAAGGTTTGAAATACTAGCAGCTG
TACTTAATGGTATTACTTTGATGTTAATAGCTATCTATATTTTTTATGAAGCTATTGAGAGATTTAAA
AACCCCCCTGAGGTAGCTTCTACAGGCATGTTAATTATCGCCTTGGTAGGCTTGTTTATTAATATTAT
TGTGGCTTGGATAATGCTGCGCGGGAGCGATGTAGAAGAAAACTTAAATATGCGTGGAGCATATTTGC
ATGTAATAAGCGACATGCTTGGATCTATAGGTGCAGTTATAGCAGCCCTTCTCATTATATTTTTTAGA
TGGGGGTGGGCGGATCCTTTAGCAAGTGTGATTGTAGCAATTTTAGTACTACGTAGCGGCTTTTATGT
AACAAAATCAAGTCTTCATGTATTAATGGAGGGAGCACCAAGCAATATAAATACAAAAGACATTATTA
AAACTATTAAAAAATTCAAAGAAGTTAAAAATATTCATGACTTTCATGTTTGGTCAGTAACTAGCGGA
TTAAACGCGTTATCTTGTCATATTGTTGTAGAAGATACAATGACCATTACTGAAAATGAGTTTTTACT
TAAACGTATAGAACATGAATTCAATCATCAAAATATTCAACATGTCACAATACAGACTGAAACTTCTA
ACAATAATCATAGTGAAAAATTGTTTTGTATCGTGAAAGAAGAAGACAG

NOV69a, CG58510-01 Protein Sequence	SEQ ID NO: 1032	308 aa	MW at 34326.6kD

MENDHNHNHTHGANKKTLLISFIIITSYMIVEGLGGFFTNSLALISDAGHMLSDSISLGIALIAFTLG
AKQANTNKTFGYKRFEILAAVLNGITLMLIAIYIFYEAIERFKNPPEVASTGMLIIALVGLFINIIVA
WIMLRGSDVEENLNMRGAYLHVISDMLGSIGAVIAALLIIFFRWGWADPLASVIVAILVLRSGFYVTK
SSLHVLMEGAPSNINTKDIIKTIKKFKEVKNIHDFHVWSVTSGLNALSCHIVVEDTMTITENEFLLKR
IEHEFNHQNIQHVTIQTETSNNNHSEKLFCIVKEED

Further analysis of the NOV69a protein yielded the following properties shown in Table 69B.

Table 69B. Protein Sequence Properties NOV69a

SignalP analysis:	Cleavage site between residues 34 and 35

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 4;  pos.chg 0;  neg.chg 2
      H-region:  length 10;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -6.24
      possible cleavage site: between 43 and 44

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   6
      INTEGRAL    Likelihood = -4.46   Transmembrane   18 -  34
      INTEGRAL    Likelihood = -1.65   Transmembrane   51 -  67
      INTEGRAL    Likelihood = -6.32   Transmembrane   85 - 101
      INTEGRAL    Likelihood =-11.41   Transmembrane  122 - 138
      INTEGRAL    Likelihood = -9.55   Transmembrane  162 - 178
      INTEGRAL    Likelihood = -4.73   Transmembrane  186 - 202
      PERIPHERAL  Likelihood =  2.54 (at 241)
      ALOM score: -11.41  (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 25
      Charge difference: -3.0   C(-1.0) - N( 2.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  6.56
      Hyd Moment(95):  3.47    G content:       0
      D/E content:     2       S/T content:     0
      Score: -7.29

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found
```

```
NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   7.5%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: VKEE

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       66.7 %: endoplasmic reticulum
       22.2 %: mitochondrial
       11.1 %: vesicles of secretory system

>> prediction for CG58510-01 is end (k=9)
```

A search of the NOV69a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 69C.

Table 69C. Geneseq Results for NOV69a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV69a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB48454	Listeria monocytogenes protein #1158 - Listeria monocytogenes, 303 aa. [WO200177335-A2, 18-OCT-2001]	5..296 11..299	163/294 (55%) 225/294 (76%)	2e-91
ABP39345	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4190 - Staphylococcus epidermidis, 342 aa. [US6380370-B1, 30-APR-2002]	3..302 35..334	157/300 (52%) 220/300 (73%)	3e-91
ABP40773	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5618 - Staphylococcus epidermidis, 359 aa. [US6380370-B1, 30-APR-2002]	1..299 47..351	112/305 (36%) 188/305 (60%)	4e-56
AAU61849	Propionibacterium acnes immunogenic protein #22745 - Propionibacterium acnes, 310 aa. [WO200181581-A2, 01-NOV-2001]	6..298 10..306	108/299 (36%) 185/299 (61%)	8e-52
AAB76797	Corynebacterium glutamicum MCT protein SEQ ID NO:576 - Corynebacterium glutamicum, 318 aa. [WO200100805-A2, 04-JAN-2001]	3..300 18..317	107/302 (35%) 180/302 (59%)	2e-51

In a BLAST search of public sequence databases, the NOV69a protein was found to have homology to the proteins shown in the BLASTP data in Table 69D.

Table 69D. Public BLASTP Results for NOV69a				
Protein Accession Number	**Protein/Organism/Length**	**NOV69a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
O07084	Cation transport protein YRDO - Bacillus subtilis, 311 aa.	7..307 3..303	206/301 (68%) 253/301 (83%)	e-122
P71023	CzcD - Bacillus subtilis, 295 aa (fragment).	21..307 1..287	194/287 (67%) 242/287 (83%)	e-115
Q8NVF2	CzrB protein - Staphylococcus aureus (strain MW2), 326 aa.	3..302 8..307	161/300 (53%) 225/300 (74%)	2e-95
Q99SB4	CzrB protein (Cation-efflux system membrane protein homolog) - Staphylococcus aureus (strain Mu50 / ATCC 700699), and, 325 aa.	3..302 7..306	161/300 (53%) 225/300 (74%)	2e-95
Q9ZNF5	CzrB protein - Staphylococcus aureus, 325 aa.	3..302 7..306	161/300 (53%) 225/300 (74%)	2e-95

PFam analysis indicates that the NOV69a protein contains the domains shown in the Table 69E.

Table 69E. Domain Analysis of NOV69a			
Pfam Domain	**NOV69a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Cation_efflux	18..296	102/303 (34%) 245/303 (81%)	1.5e-106

Example 70.

The NOV70 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 70A.

Table 70A. NOV70 Sequence Analysis		
NOV70a, CG59309-01 DNA Sequence	SEQ ID NO: 1033	1375 bp
	ORF Start: ATG at 96	ORF Stop: TAA at 1362

GGGACGCCGGACGCCGTCCGGACATTCGGCGCGCTTGCCACGATCTTGGACGGGTCTCGGGCCTCGAC
CTTTGAATTCCCCGCTCCGGCTCCAAGATGTCAGCAACGCTGATCCTGGAGCCCCCAGGCCGCTGCTG
CTGGAACGAGCCGGTGCGCATTGCCGTGCGCGGCCTGGCCCCGGAGCAGCGGGTTACGCTGCGCGCGT
CCCTGCGCGACGAGAAGGGCGCGCTCTTCCGGGCCCACGCGCGCTACTGCGCCGACGCCCGCGGCGAG
CTGGACCTGGAGCGCGCACCCGCGCTGGGCGGCAGCTTCGCGGGACTCGAGCCCATGGGGCTGCTCTG
GGCCCTGGAACCCGAGAAGCCTTTTTGGCGCTTCCTGAAGCGGGACGTACAGATTCCTTTTGTCGTGG
AGTTGGAGGTGCTGGACGGCCACGACCCCGAGCCTGGACGGCTGCTGTGCCAGGCGCAGCACGAGCGC
CACTTCCTCCCGCCAGGGGTGCGGCGCCAGTCGGTGCGAGCGGGCCGGGTGCGCGCCACGCTCTTCCT

GCCGCCAGGTGAGCCTGGACCCTTCCCAGGGATCATTGACATCTTTGGTATTGGAGGGGGCCTCTTGG
AATATCGAGCCAGCCTCCTTGCTGGCCATGGCTTTGCCACGTTGGCTCTAGCTTATTATAACTTTGAA
GATCTCCCCAATAACATGGACAACATATCCCTGGAGTACTTCGAAGAAGCCGTATGCTACATGCTTCA
ACATCCCCAGGTTAAAGGCCCAGGCATTGGGCTTTTGGGCATTTCTCTAGGAGCTGATATTTGTCTCT
CAATGGCCTCATTCTTGAAGAATGTCTCAGCCACAGTTTCCATCAATGGATCTGGGATCAGTGGGAAC
ACAGCCATCAACTATAAGCACAGTAGCATTCCACCATTGGGCTATGACCTGAGGAGAATCAAGGTAGC
TTTCTCAGGCCTCGTGGACATTGTGGATATAAGGAATGCTCTCGTAGGAGGGTACAAGAACCCCAGCA
TGATTCCAATAGAGAAGGCCCAGGGGCCCATCCTGCTCATTGTTGGTCAGGATGACCATAACTGGAGA
AGTGAGTTGTATGCCCAAACAGTCTCTGAACGGTTACAGGCCCATGGAAAGGAAAAACCCCAGATCAT
CTGTTACCCTGGGACTGGGCATTACATCGAGCCTCCTTACTTCCCCCTGTGCCCAGCTTCCCTTCACA
GATTACTGAACAAACATGTTATATGGGGTGGGGAGCCCAGGGCTCATTCTAAGGCCCAGGAAGATGCC
TGGAAGCAAATTCTAGCCTTCTTCTGCAAACACCTGGGAGGTACCCAGAAAACAGCTGTCCCTAAATT
GTAATGCATTTGTCT

NOV70a, CG59309-01 Protein Sequence	SEQ ID NO: 1034	422 aa	MW at 46455.1kD

MSATLILEPPGRCCWNEPVRIAVRGLAPEQRVTLRASLRDEKGALFRAHARYCADARGELDLERAPAL
GGSFAGLEPMGLLWALEPEKPFWRFLKRDVQIPFVVELEVLDGHDPEPGRLLCQAQHERHFLPPGVRR
QSVRAGRVRATLFLPPGEPGPFPGIIDIFGIGGGLLEYRASLLAGHGFATLALAYYNFEDLPNNMDNI
SLEYFEEAVCYMLQHPQVKGPGIGLLGISLGADICLSMASFLKNVSATVSINGSGISGNTAINYKHSS
IPPLGYDLRRIKVAFSGLVDIVDIRNALVGGYKNPSMIPIEKAQGPILLIVGQDDHNWRSELYAQTVS
ERLQAHGKEKPQIICYPGTGHYIEPPYFPLCPASLHRLLNKHVIWGGEPRAHSKAQEDAWKQILAFFC
KHLGGTQKTAVPKL

NOV70b, 278901386 DNA Sequence	SEQ ID NO: 1035	1285 bp
	ORF Start: at 2	ORF Stop: end of sequence

CACCGGATCCACCATGTCAGCAACGCTGATCCTGGAGCCCCCAGGCCGCTGCTGCTGGAACGAGCCGG
TGCGCATTGCCGTGCGCGGCCTGGCCCCGGAGCAGCGGGTTACGCTGCGCGCGTCCCTGCGCGACGAG
AAGGGCGCGCTCTTCCGGGCCCACGCGCGCTACTGCGCCGACGCCTGCGGCGAGCTGGACCTGGAGCG
CGCACCCGCGCTGGGCGGCAGCTTCGCGGGACTCGAGCCCATGGGGCTGCTCTGGGCCCTGGAACCCG
AGAAGCCTTTTTGGCGCTTCCTGAAGCGGGACGTACAGATTCCTTTTGTCGTGGAGTTGGAGGTGCTG
GACGGCCACGACCCCGAGCCTGGACGGCTGCTGTGCCAGGCGCAGCACGAGCGCCACTTCCTCCCGCC
AGGGGTGCGGCGCCAGTCGGTGCGAGCGGGCCGGGTGCGCGCCACGCTCTTCCTGCCGCCAGGACCTG
GACCCTTCCCAGGGATCATTGACATCTTTGGTATTGGAGGGGGCCTCTTGGAATATCGAGCCAGCCTC
CTTGCTGGCCATGGCTTTGCCACGTTGGCTCTAGCTTATTATAACTTTGAAGATCTCCCCAATAACAT
GGACAACATATCCCTGGAGTACTTCGAAGAAGCCGTATGCTACATGCTTCAACATCCCCAGGTAAAAG
GCCCAGGCATTGGGCTTTTGGGCATTTCTCTAGGAGCTGATATTTGTCTCTCAATGGCCTCATTCTTG
AAGAATGTCTCAGCCACAGTTTCCATCAATGGATCTGGGATCAGTGGGAACACAGCCATCAACTATAA
GCACAGTAGCATTCCACCATTGGGCTATGACCTGAGGAGAATCAAGGTAGCTTTCTCAGGCCTCGTGG
ACATCGTGGATATAAGGAATGCTCTCGTAGGAGGGTACAAGAACCCCAGCATGATTCCAATAGAGAAG
GCCCAGGGGCCCATCCTGCTCATTGTTGGTCAGGATGACCATAACTGGAGAAGTGAGTTGTATGCCCA
AACAGTCTCTGAACGGTTACAGGCCCATGGAAAGGAAAAACCCCAGATCATCTGTTACCCTGGGACTG
GGCATTACATCGAGCCTCCTTACTTCCCCCTGTGCCCAGCTTCCCTTCACAGATTACTGAACAAACAT
GTTATATGGGGTGGGGAGCCCAGGGCTCATTCTAAGGCCCAGGAAGATGCCTGGAAGCAAATTCTAGC
CTTCTTCTGCAAACACCTGGGAGGTACCCAGAAAACAGCTGTCCCTAAATTGGTCGACGGC

NOV70b, 278901386 Protein Sequence	SEQ ID NO: 1036	428 aa	MW at 46890.5kD

TGSTMSATLILEPPGRCCWNEPVRIAVRGLAPEQRVTLRASLRDEKGALFRAHARYCADACGELDLER
APALGGSFAGLEPMGLLWALEPEKPFWRFLKRDVQIPFVVELEVLDGHDPEPGRLLCQAQHERHFLPP
GVRRQSVRAGRVRATLFLPPGPGPFPGIIDIFGIGGGLLEYRASLLAGHGFATLALAYYNFEDLPNNM
DNISLEYFEEAVCYMLQHPQVKGPGIGLLGISLGADICLSMASFLKNVSATVSINGSGISGNTAINYK
HSSIPPLGYDLRRIKVAFSGLVDIVDIRNALVGGYKNPSMIPIEKAQGPILLIVGQDDHNWRSELYAQ
TVSERLQAHGKEKPQIICYPGTGHYIEPPYFPLCPASLHRLLNKHVIWGGEPRAHSKAQEDAWKQILA
FFCKHLGGTQKTAVPKLVDG

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 70B.

Table 70B. Comparison of the NOV70 protein sequences.

```
NOV70a  ----MSATLILEPPGRCCWNEPVRIAVRGLAPEQRVTLRASLRDEKGALFRAHARYCADA
NOV70b  TGSTMSATLILEPPGRCCWNEPVRIAVRGLAPEQRVTLRASLRDEKGALFRAHARYCADA

NOV70a  RGELDLERAPALGGSFAGLEPMGLLWALEPEKPFWRFLKRDVQIPFVVELEVLDGHDPEP
NOV70b  CGELDLERAPALGGSFAGLEPMGLLWALEPEKPFWRFLKRDVQIPFVVELEVLDGHDPEP

NOV70a  GRLLCQAQHERHFLPPGVRRQSVRAGRVRATLFLPPGEPGPFPGIIDIFGIGGGLLEYRA
NOV70b  GRLLCQAQHERHFLPPGVRRQSVRAGRVRATLFLPPG-PGPFPGIIDIFGIGGGLLEYRA

NOV70a  SLLAGHGFATLALAYYNFEDLPNNMDNISLEYFEEAVCYMLQHPQVKGPGIGLLGISLGA
NOV70b  SLLAGHGFATLALAYYNFEDLPNNMDNISLEYFEEAVCYMLQHPQVKGPGIGLLGISLGA

NOV70a  DICLSMASFLKNVSATVSINGSGISGNTAINYKHSSIPPLGYDLRRIKVAFSGLVDIVDI
NOV70b  DICLSMASFLKNVSATVSINGSGISGNTAINYKHSSIPPLGYDLRRIKVAFSGLVDIVDI

NOV70a  RNALVGGYKNPSMIPIEKAQGPILLIVGQDDHNWRSELYAQTVSERLQAHGKEKPQIICY
NOV70b  RNALVGGYKNPSMIPIEKAQGPILLIVGQDDHNWRSELYAQTVSERLQAHGKEKPQIICY

NOV70a  PGTGHYIEPPYFPLCPASLHRLLNKHVIWGGEPRAHSKAQEDAWKQILAFFCKHLGGTQK
NOV70b  PGTGHYIEPPYFPLCPASLHRLLNKHVIWGGEPRAHSKAQEDAWKQILAFFCKHLGGTQK

NOV70a  TAVPKL---
NOV70b  TAVPKLVDG

NOV70a  (SEQ ID NO:  1034)
NOV70b  (SEQ ID NO:  1036)
```

Further analysis of the NOV70a protein yielded the following properties shown in Table 70C.

Table 70C. Protein Sequence Properties NOV70a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 8;  pos.chg 0;  neg.chg 1
      H-region:  length 3;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -7.81
      possible cleavage site: between 13 and 14

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   0
      PERIPHERAL  Likelihood =  1.75 (at 285)
      ALOM score:  -1.86  (number of TMSs: 0)
```

```
MITDISC: discrimination of mitochondrial targeting seq
      R content:        1       Hyd Moment(75):  7.13
      Hyd Moment(95):  4.74     G content:       1
      D/E content:     2        S/T content:     2
      Score: -6.03

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  10.4%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:
      KKXX-like motif in the C-terminus: AVPK

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------
```

```
Final Results (k = 9/23):

       65.2 %: cytoplasmic
       13.0 %: nuclear
        8.7 %: mitochondrial
        8.7 %: peroxisomal
        4.3 %: plasma membrane

>> prediction for CG59309-01 is cyt (k=23)
```

A search of the NOV70a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 70D.

Table 70D. Geneseq Results for NOV70a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV70a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE25382	Human NZMS-6 protein - Homo sapiens, 421 aa. [WO200246385-A2, 13-JUN-2002]	1..422 1..421	421/422 (99%) 421/422 (99%)	0.0
AAU76350	Human Acyl-CoA thioesterase 56939 - Homo sapiens, 421 aa. [WO200208274-A2, 31-JAN-2002]	1..422 1..421	420/422 (99%) 420/422 (99%)	0.0
AAM41490	Human polypeptide SEQ ID NO 6421 - Homo sapiens, 494 aa. [WO200153312-A1, 26-JUL-2001]	1..422 74..494	296/422 (70%) 341/422 (80%)	e-178
AAM39704	Human polypeptide SEQ ID NO 2849 - Homo sapiens, 483 aa. [WO200153312-A1, 26-JUL-2001]	1..422 63..483	296/422 (70%) 341/422 (80%)	e-178
AAY71112	Human Hydrolase protein-10 (HYDRL-10) - Homo sapiens, 483 aa. [WO200028045-A2, 18-MAY-2000]	1..422 63..483	296/422 (70%) 341/422 (80%)	e-178

In a BLAST search of public sequence databases, the NOV70a protein was found to have homology to the proteins shown in the BLASTP data in Table 70E.

Table 70E. Public BLASTP Results for NOV70a				
Protein Accession Number	**Protein/Organism/Length**	**NOV70a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
CAD35497	Sequence 1 from Patent WO0208274 - Homo sapiens (Human), 421 aa.	1..422 1..421	420/422 (99%) 420/422 (99%)	0.0
Q8N9L9	Hypothetical protein FLJ36904 - Homo sapiens (Human), 421 aa.	1..422 1..421	417/422 (98%) 417/422 (98%)	0.0
CAD62354	Human full-length cDNA clone CS0DI029YH06 of Placenta of Homo sapiens (human) - Homo sapiens (Human), 352 aa (fragment).	70..422 1..352	352/353 (99%) 352/353 (99%)	0.0
Q8BWN8	Peroxisomal long chain acyl-CoA thioesterase IB - Mus musculus (Mouse), 421 aa.	1..422 1..421	313/422 (74%) 363/422 (85%)	0.0
Q8BL20	Peroxisomal long chain acyl-CoA thioesterase IB - Mus musculus (Mouse), 421 aa.	1..422 1..421	313/422 (74%) 362/422 (85%)	0.0

PFam analysis indicates that the NOV70a protein contains the domains shown in the Table 70F.

Table 70F. Domain Analysis of NOV70a			
Pfam Domain	**NOV70a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Bile_Hydr_Trans	3..152	84/157 (54%) 140/157 (89%)	1.9e-82
DLH	144..411	65/309 (21%) 170/309 (55%)	0.33

Example 71.

The NOV71 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 71A.

Table 71A. NOV71 Sequence Analysis		
NOV71a, CG59490-01 DNA Sequence	SEQ ID NO: 1037	660 bp
	ORF Start: at 1	ORF Stop: end of sequence
TCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGGGAGGAGTTGGAGGCTTGCGCGTT TAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGATCGTCC		

GTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAGCTGGAG
GCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCGGGACGTGCCCTC
GGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGCCCCTCA
GCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGTAACCAGACCTGTCGCCGCCGCTTT
CCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGACGGGAA
CCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGGTCCAGG
TGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTATCCCGGCATGTACACCCGCGTGACG
AGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAGACGC

NOV71a, CG59490-01 Protein Sequence	SEQ ID NO: 1038	220 aa	MW at 24527.8kD

SLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLKLE
APVPLSELIHPVSLPSASRDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSNVLCNQTCRRRF
PSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTRVT
SYVSWIRQYVPPFPRR

NOV71b, 207639512 DNA Sequence	SEQ ID NO: 1039	672 bp
	ORF Start: at 1	ORF Stop: end of sequence

AGATCTTCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGGGAGGAGTTGGAGGCTTG
CGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGA
TCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAG
CTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGT
GCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGC
CCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAGCGTCCTCTGTAACCAGACCTGTCGCCGC
CGCTTTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGA
CGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGG
TCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTATCCCGGCATGTACACCCGC
GTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAGACGCCTCGAG

NOV71b, 207639512 Protein Sequence	SEQ ID NO: 1040	224 aa	MW at 24943.3kD

RSSLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLK
LEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSSVLCNQTCRR
RFPSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTR
VTSYVSWIRQYVPPFPRRLE

NOV71c, 207639476 DNA Sequence	SEQ ID NO: 1041	672 bp
	ORF Start: at 1	ORF Stop: end of sequence

AGATCTTCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGGGAGGAGTTGGAGGCTTG
CGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGA
TCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAG
CTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGT
GCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGC
CCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGTAACCAGACCTGTCGCCGC
CGCTTTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGA
CGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGG
TCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTATCCCGGCATGTACACCCGC
GTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAGACGCCTCGAG

NOV71c, 207639476 Protein Sequence	SEQ ID NO: 1042	224 aa	MW at 24970.3kD

RSSLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLK
LEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSNVLCNQTCRR
RFPSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTR
VTSYVSWIRQYVPPFPRRLE

NOV71d, 207639523 DNA Sequence	SEQ ID NO: 1043	672 bp
	ORF Start: at 1	ORF Stop: end of sequence

AGATCTTCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGGGAGGAGTTGGAGGCTTG
CGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGA
TCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGCCGGTGCGGACATCGCCCTGCTGAAG

CTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGT
GCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGC
CCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGTAACCAGACCTGTCGCCGC
CGCTTTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGA
CGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGG
TCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTATCCCGGCATGTACACCCGC
GTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAGACGCCTCGAG

NOV71d, 207639523 Protein Sequence	SEQ ID NO: 1044	224 aa	MW at 24984.3kD

RSSLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQAGADIALLK
LEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSNVLCNQTCRR
RFPSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTR
VTSYVSWIRQYVPPFPRRLE

NOV71e, CG59490-02 DNA Sequence	SEQ ID NO: 1045	672 bp
	ORF Start: at 7	ORF Stop: at 667

AGATCTTCACTGGGGGCTGCGACGTCTCGGCCAGGAGGCACCCCTGGCAGGGAGGAGTTGGAGGCTTG
CGCGTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGACGACCAGCGGACGAAGGTGGTTGAGA
TCGTCCGTCACCCCCAGTACAACGAGAGCCTGTCTGCCCAGGGCGGTGCGGACATCGCCCTGCTGAAG
CTGGAGGCCCCGGTGCCGCTGTCTGAGCTCATCCACCCGGTCTCGCTCCCGTCTGCCTCCCTGGACGT
GCCCTCGGGGAAGACCTGCTGGGTGACCGGCTGGGGTGTCATTGGACGTGGAGAACTACTGCCCTGGC
CCCTCAGCTTGTGGGAGGCGACGGTGAAGGTCAGGAGCAACGTCCTCTGTAACCAGACCTGTCGCCGC
CGCTTTCCTTCCAACCACACTGAGCGGTTTGAGCGGCTCATCAAGGACGACATGCTGTGTGCCGGGGA
CGGGAACCACGGCTCCTGGCCAGGCGACAACGGGGGCCCCCTCCTGTGCAGGCGGAATTGCACCTGGG
TCCAGGTGGAGGTGGTGAGCTGGGGCAAACTCTGCGGCCTTCGCGGCTATCCCGGCATGTACACCCGC
GTGACGAGCTACGTGTCCTGGATCCGCCAGTACGTCCCGCCGTTCCCCAGACGCCTCGAG

NOV71e, CG59490-02 Protein Sequence	SEQ ID NO: 1046	220 aa	MW at 24484.8kD

SLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQGGADIALLKLE
APVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEATVKVRSNVLCNQTCRRRF
PSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCTWVQVEVVSWGKLCGLRGYPGMYTRVT
SYVSWIRQYVPPFPRR

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 71B.

Table 71B. Comparison of the NOV71 protein sequences.

```
NOV71a  --SLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQG
NOV71b  RSSLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQG
NOV71c  RSSLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQG
NOV71d  RSSLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQA
NOV71e  --SLGAATSRPGGTPGREELEACAFRVQVGQLRLYEDDQRTKVVEIVRHPQYNESLSAQG

NOV71a  GADIALLKLEAPVPLSELIHPVSLPSASRDVPSGKTCWVTGWGVIGRGELLPWPLSLWEA
NOV71b  GADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEA
NOV71c  GADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEA
NOV71d  GADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEA
NOV71e  GADIALLKLEAPVPLSELIHPVSLPSASLDVPSGKTCWVTGWGVIGRGELLPWPLSLWEA

NOV71a  TVKVRSNVLCNQTCRRRFPSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCT
NOV71b  TVKVRSSVLCNQTCRRRFPSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCT
NOV71c  TVKVRSNVLCNQTCRRRFPSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCT
NOV71d  TVKVRSNVLCNQTCRRRFPSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCT
NOV71e  TVKVRSNVLCNQTCRRRFPSNHTERFERLIKDDMLCAGDGNHGSWPGDNGGPLLCRRNCT

NOV71a  WVQVEVVSWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR--
```

```
NOV71b  WVQVEVVSWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRRLE
NOV71c  WVQVEVVSWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRRLE
NOV71d  WVQVEVVSWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRRLE
NOV71e  WVQVEVVSWGKLCGLRGYPGMYTRVTSYVSWIRQYVPPFPRR--

NOV71a  (SEQ ID NO:  1038)
NOV71b  (SEQ ID NO:  1040)
NOV71c  (SEQ ID NO:  1042)
NOV71d  (SEQ ID NO:  1044)
NOV71e  (SEQ ID NO:  1046)
```

Further analysis of the NOV71a protein yielded the following properties shown in Table 71C.

Table 71C. Protein Sequence Properties NOV71a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 8;  pos.chg 1;  neg.chg 0
      H-region:  length 6;  peak value  -8.72
      PSG score: -13.12

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -7.58
      possible cleavage site: between 14 and 15

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  4.88 (at 62)
      ALOM score:   4.88  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       2       Hyd Moment(75):  6.94
      Hyd Moment(95):  7.43    G content:       4
      D/E content:     1       S/T content:     4
      Score: -3.60

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 18  SRP|GG

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:  12.3%
      NLS Score: -0.47
```

```
KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 89

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       39.1 %: cytoplasmic
       30.4 %: mitochondrial
       17.4 %: nuclear
        8.7 %: vesicles of secretory system
        4.3 %: vacuolar

>> prediction for CG59490-01 is cyt (k=23)
```

A search of the NOV71a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 71D.

Table 71D. Geneseq Results for NOV71a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV71a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAE08591	Human NOV12 protein - Homo sapiens, 220 aa. [WO200161009-A2, 23-AUG-2001]	1..220 1..220	220/220 (100%) 220/220 (100%)	e-133
AAE14347	Human protease PRTS-12 protein - Homo sapiens, 262 aa. [WO200183775-A2, 08-NOV-2001]	15..220 57..262	205/206 (99%) 205/206 (99%)	e-123
AAE08590	Human NOV11 protein - Homo sapiens, 285 aa. [WO200161009-A2, 23-AUG-2001]	16..220 81..285	204/205 (99%) 204/205 (99%)	e-123
AAU82736	Amino acid sequence of novel human protease #35 - Homo sapiens, 948 aa. [WO200200860-A2, 03-JAN-2002]	15..215 211..411	198/201 (98%) 198/201 (98%)	e-118
AAE08587	Human NOV8 protein - Homo sapiens, 290 aa. [WO200161009-A2, 23-AUG-2001]	14..220 84..290	200/207 (96%) 200/207 (96%)	e-118

In a BLAST search of public sequence databases, the NOV71a protein was found to have homology to the proteins shown in the BLASTP data in Table 71E.

Table 71E. Public BLASTP Results for NOV71a				
Protein Accession Number	**Protein/Organism/Length**	**NOV71a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAP21675	Mast cell protease-11 - Mus musculus (Mouse), 318 aa.	16..217 86..284	118/202 (58%) 142/202 (69%)	2e-59
AAA30855	Mastin precursor - Canis sp, 251 aa (fragment).	16..218 53..249	110/203 (54%) 127/203 (62%)	9e-51
P19236	Mastocytoma protease precursor (EC 3.4.21.-) - Canis familiaris (Dog), 269 aa.	16..218 71..267	110/203 (54%) 127/203 (62%)	9e-51
Q8SQ44	Tryptase precursor - Sus scrofa (Pig), 277 aa.	24..218 90..275	105/195 (53%) 127/195 (64%)	7e-49
Q9XSM1	Tryptase (EC 3.4.21.59) - Ovis aries (Sheep), 273 aa.	14..218 75..273	89/205 (43%) 120/205 (58%)	1e-43

PFam analysis indicates that the NOV71a protein contains the domains shown in the Table 71F.

Table 71F. Domain Analysis of NOV71a			
Pfam Domain	**NOV71a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
trypsin	8..210	81/272 (30%) 152/272 (56%)	1.9e-32

Example 72.

The NOV72 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 72A.

Table 72A. NOV72 Sequence Analysis			
NOV72a, CG59693-01 DNA Sequence	SEQ ID NO: 1047		972 bp
	ORF Start: ATG at 1		ORF Stop: TAA at 970
ATGGATTCGAAATATCAGTGTGTGAAGCTGAATGATGGTCACTTCATGCCTGTCCTGGGATTTGGCAC CTATGCGCCTGCAGAGGTTCCTAAAAGTAAAGCTTTAGAGGCCACCAAATTGGCAATTGAAGCTGGCT TCCGCCATATTGATTCTGCTCATTTATACAATAATGAGGAGCAGGTTGGACTGGCCATCCGAAGCAAG ATTGCAGATGGCAGTGTGAAGAGAGAAGACATATTCTACACTTCAAAGCTTTGGTGCAATTCCCATCG ACCAGAGTTGGTCCGACCAGCCTTGGAAAGGTCACTGAAAAATCTTCAATTGGATTATGTTGACCTCT ACCTTATTCATTTTCCAGTGTCTGTAAAGCCAGGTGAGGAAGTGATCCCAAAAGATGAAAATGGAAAA ATACTATTTGACACAGTGGATCTCTGTGCCACGTGGGAGGCCGTGGAGAAGTGTAAAGATGCAGGATT GGCCAAGTCCATCGGGGTGTCCAACTTCAACCGCAGGCAGCTGGAGATGATCCTCAACAAGCCAGGGC TCAAGTACAAGCCTGTCTGCAACCAGGTGGAATGTCATCCTTACTTCAACCAGAGAAAACTGCTGGAT TTCTGCAAGTCAAAAGACATTGTTCTGGTTGCCTATAGTGCTCTGGGATCCCACCGAGAAGAACCATG GGTGGACCCGAACTCCCCGGTGCTCTTGGAGGACCCAGTCCTTTGTGCCTTGGCAAAAAAGCACAAGC GAACCCCAGCCCTGATTGCCCTGCGCTACCAGCTACAGCGTGGGGTTGTGGTCCTGGCCAAGAGCTAC AATGAGCAGCGCATCAGACAGAACGTGCAGGTGTTTGAATTCCAGTTGACTTCAGAGGAGATGAAAGC CATAGATGGCCTAAACAGAAATGTGCGATATTTGACCCTTGATATTTTTGCTGGCCCCCCTAATTATC CATTTTCTGATGAATAT**TAA**			
NOV72a, CG59693-01 Protein Sequence	SEQ ID NO: 1048	323 aa	MW at 36787.9kD
MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHIDSAHLYNNEEQVGLAIRSK IADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLDYVDLYLIHFPVSVKPGEEVIPKDENGK ILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFNRRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLD FCKSKDIVLVAYSALGSHREEPWVDPNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSY NEQRIRQNVQVFEFQLTSEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY			
NOV72b, CG59693-03 DNA Sequence	SEQ ID NO: 1049		972 bp
	ORF Start: ATG at 1		ORF Stop: TAA at 970
ATGGATTCGAAATATCAGTGTGTGAAGCTGAATGATGGTCACTTCATGCCTGTCCTGGGATTTGGCAC CTATGCGCCTGCAGAGGTTCCTAAAAGTAAAGCTTTAGAGGCCACCAAATTGGCAATTGAAGCTGGCT TCCGCCATATTGATTCTGCTCATTTATACAATAATGAGGAGCAGGTTGGACTGGCCATCCGAAGCAAG ATTGCAGATGGCAGTGTGAAGAGAGAAGACATATTCTACACTTCAAAGCTTTGGTGCAATTCCCATCG ACCAGAGTTGGTCCGACCAGCCTTGGAAAGGTCACTGAAAAATCTTCAATTGGATTATGTTGACCTCT ACCTTATTCATTTTCCAGTGTCTGTAAAGCCAGGTGAGGAAGTGATCCCAAAAGATGAAAATGGAAAA ATACTATTTGACACAGTGGATCTCTGTGCCACGTGGGAGGCCGTGGAGAAGTGTAAAGATGCAGGATT GGCCAAGTCCATCGGGGTGTCCAACTTCAACCGCAGGCAGCTGGAGATGATCCTCAACAAGCCAGGGC TCAAGTACAAGCCTGTCTGCAACCAGGTGGAATGTCATCCTTACTTCAACCAGAGAAAACTGCTGGAT TTCTGCAAGTCAAAAGACATTGTTCTGGTTGCCTATAGTGCTCTGGGATCCCACCGAGAAGAACCATG GGTGGACCCGAACTCCCCGGTGCTCTTGGAGGACCCAGTCCTTTGTGCCTTGGCAAAAAAGCACAAGC			

GAACCCCAGCCCTGATTGCCCTGCGCTACCAGCTACAGCGTGGGGTTGTGGTCCTGGCCAAGAGATAC
AATGAGCAGCGCATCAGACAGAACGTGCAGGTGTTTGAATTCCAGTTGACTTCAGAGGAGATGAAAGC
CATAGATGGCCTAAACAGAAATGTGCGATATTTGACCCTTGATATTTTTGCTGGCCCCCCTAATTATC
CATTTTCTGATGAATATTAA

NOV72b, CG59693-03 Protein Sequence	SEQ ID NO: 1050	323 aa	MW at 36857.0kD

MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHIDSAHLYNNEEQVGLAIRSK
IADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLDYVDLYLIHFPVSVKPGEEVIPKDENGK
ILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFNRRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLD
FCKSKDIVLVAYSALGSHREEPWVDPNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKRY
NEQRIRQNVQVFEFQLTSEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY

NOV72c, 277637252 DNA Sequence	SEQ ID NO: 1051	1001 bp
	ORF Start: at 8	ORF Stop: TAA at 989

CATCTAGGCCACCATGGCCATGGATTCGAAATATCAGTGTGTGAAGCTGAATGATGGTCACTTCATGC
CTGTCCTGGGATTTGGCACCTATGCGCCTGCAGAGGTTCCTAAAAGTAAAGCTTTAGAGGCCACCAAA
TTGGCAATTGAAGCTGGCTTCCGCCATATTGATTCTGCTCATTTATACAATAATGAGGAGCAGGTTGG
ACTGGCCATCCGAAGCAAGATTGCAGATGGCAGTGTGAAGAGAGAAGACATATTCTACACTTCAAAGC
TTTGGTGCAATTCCCATCGACCAGAGTTGGTCCGACCAGCCTTGGAAAGGTCACTGAAAAATCTTCAA
TTGGATTATGTTGACCTCTACCTTATTCATTTTCCAGTGTCTGTAAAGCCAGGTGAGGAAGTGATCCC
AAAAGATGAAAATGGAAAAATACTATTTGACACAGTGGATCTCTGTGCCACGTGGGAGGCCGTGGAGA
AGTGTAAAGATGCAGGATTGGCCAAGTCCATCGGGGGTGTCCAACTTCAACCGCAGGCAGCTGGAGATG
ATCCTCAACAAGCCAGGGCTCAAGTACAAGCCTGTCTGCAACCAGGTGGAATGTCATCCTTACTTCAA
CCAGAGAAAACTGCTGGATTTCTGCAAGTCAAAAGACATTGTTCTGGTTGCCTATAGTGCTCTGGGAT
CCCACCGAGAAGAACCATGGGTGGACCCGAACTCCCCGGTGCTCTTGGAGGACCCAGTCCTTTGTGCC
TTGGCAAAAAAGCACAAGCGAACCCCAGCCCTGATTGCCCTGCGCTACCAGCTACAGCGTGGGGTTGT
GGTCCTGGCCAAGAGCTACAATGAGCAGCGCATCAGACAGAACGTGCAGGTGTTTGAATTCCAGTTGA
CTTCAGAGGAGATGAAAGCCATAGATGGCCTAAACAGAAATGTGCGATATTTGACCCTTGATATTTTT
GCTGGCCCCCCTAATTATCCATTTTCTGATGAATATTAAACGCGTGATC

NOV72c, 277637252 Protein Sequence	SEQ ID NO: 1052	327 aa	MW at 37162.4kD

ATMAMDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHIDSAHLYNNEEQVGLA
IRSKIADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLDYVDLYLIHFPVSVKPGEEVIPKD
ENGKILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFNRRQLEMILNKPGLKYKPVCNQVECHPYFNQR
KLLDFCKSKDIVLVAYSALGSHREEPWVDPNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVL
AKSYNEQRIRQNVQVFEFQLTSEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY

NOV72d, CG59693-02 DNA Sequence	SEQ ID NO: 1053	983 bp
	ORF Start: ATG at 30	ORF Stop: TAA at 981

ATGGATTCGATATCAGTGTGTGAAGCTGAATGATGGTCACTTCGTGCCTGTCCTGGGATTTGGCACCT
ATGCGCCTGCAGAGGTTACTCCCCCAGGTTCCTAAAAGTAAAGCTTTAGAGGCCACCAAATTGGCAAT
TGAAGCTGGCTTCCGCCATATTGATTCTGCTCATTTATACAATAATGAGGAGCAGGTTGGACTGGCCA
TCCGAAGCAAGATTGCAGATGGCAGTGTGAAGAGAGAAGACATATTCTACACTTCAAAGCTTTGGTGC
AATTCCCATCGACCAGAGTTGGTCCGACCAGCCTTGGAAAGGTCACTGAAAAATCTTCAATTGGATTA
TGTTGACCTCTACCTTATTCATTTTCCAGTGTCTGTAAAGCCAGGTGAGGAAGTGATCCCAAAAGATG
AAAGTGGAAAAATACTATTTGACACAGTGGATCTCTGTGCCACGTGGGAGGCCGTGGAGAAGTGTAAA
GATGCAGGATTGGCCAAGTCCATCGGGGTGTCCAACTTCAACCGCAGGCAGCTGGAGATGATCCTCAA
CAAGCCAGGGCTCAAGTACAAGCCTGTCTGCAACCAGGTGGAATGTCATCCTTACTTCAACCAGAGAA
AACTGCTGGATTTCTGCAAGTCAAAAGACATTGTTCTGGTTGCCTATAGTGCTCTGGGATCCCACCGA
GAAGAACCATGGGTGGACCCGAACTCCCCGGTGCTCTTGGAGGACCCAGTCCTTTGTGCCTTGGCAAA
AAAGCACAAGCGAACCCCAGCCCTGGTTGCCCTGCGCTACCAGCTACAGCGTGGGGTTGTGGTCCTGG
CCAAGAGCTACAATGAGCAGCGCATCAGACAGAACGTGCAGGTGTTTGAATTCCAGTTGACTTCAGAG
GAGATGAAAGCCATAGATGGCCTAAACAGAAATGTGCGATATTTGACCCTTGATATTTTTGCTGGCCC
CCCTAATTATCCATTTTCTGATGAATATTAA

NOV72d, CG59693-02 Protein Sequence	SEQ ID NO: 1054	317 aa	MW at 36217.5kD

MMVTSCLSWDLAPMRLQRLLPQVPKSKALEATKLAIEAGFRHIDSAHLYNNEEQVGLAIRSKIADGSV
KREDIFYTSKLWCNSHRPELVRPALERSLKNLQLDYVDLYLIHFPVSVKPGEEVIPKDESGKILFDTV

DLCATWEAVEKCKDAGLAKSIGVSNFNRRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLDFCKSKD
IVLVAYSALGSHREEPWVDPNSPVLLEDPVLCALAKKHKRTPALVALRYQLQRGVVVLAKSYNEQRIR
QNVQVFEFQLTSEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY

NOV72e, CG59693-04 DNA Sequence	SEQ ID NO: 1055	994 bp
	ORF Start: ATG at 16	ORF Stop: at 979

GCCAGATCTCCCACCATGGATTCGAAATATCAGTGTGTGAAGCTGAATGATGGTCACTTCATGCCTGT
CCTGGGATTTGGCACCTATGCGCCTGCAGAGGTTCCTAAAAGTAAAGCTTTAGAGGCCACCAAATTGG
CAATTGAAGCTGGCTTCCGCCATATTGATTCTGCTCATTTATACAATAATGAGGAGCAGGTTGGACTG
GCCATCCGAAGCAAGATTGCAGATGGCAGTGTGAAGAGAGAAGACATATTCTACACTTCAAAGCTTTG
GTGCAATTCCCATCGACCAGAGTTGGTCCGACCAGCCTTGGAAAGGTCACTGAAAAATCTTCAATTGG
ATTATGTTGACCTCTACCTTATTCATTTTCCAGTGTCTGTAAAGCCAGGTGAGGAAGTGATCCCAAAA
GATGAAAATGGAAAAATACTATTTGACACAGTGGATCTCTGTGCCACGTGGGAGGCCGTGGAGAAGTG
TAAAGATGCAGGATTGGCCAAGTCCATCGGGGTGTCCAACTTCAACCGCAGGCAGCTGGAGATGATCC
TCAACAAGCCAGGGCTCAAGTACAAGCCTGTCTGCAACCAGGTGGAATGTCATCCTTACTTCAACCAG
AGAAAACTGCTGGATTTCTGCAAGTCAAAAGACATTGTTCTGGTTGCCTATAGTGCTCTGGGATCCCA
CCGAGAAGAACCATGGGTGGACCCGAACTCCCCGGTGCTCTTGGAGGACCCAGTCCTTTGTGCCTTGG
CAAAAAAGCACAAGCGAACCCCAGCCCTGATTGCCCTGCGCTACCAGCTACAGCGTGGGGTTGTGGTC
CTGGCCAAGAGCTACAATGAGCAGCGCATCAGACAGAACGTGCAGGTGTTTGAATTCCAGTTGACTTC
AGAGGAGATGAAAGCCATAGATGGCCTAAACAGAAATGTGCGATATTTGACCCTTGATATTTTTGCTG
GCCCCCCTAATTATCCATTTTCTGATGAATATCTCGAGGGTG

NOV72e, CG59693-04 Protein Sequence	SEQ ID NO: 1056	321 aa	MW at 36495.6kD

MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHIDSAHLYNNEEQVGLAIRSK
IADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLDYVDLYLIHFPVSVKPGEEVIPKDENGK
ILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFNRRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLD
FCKSKDIVLVAYSALGSHREEPWVDPNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSY
NEQRIRQNVQVFEFQLTSEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSD

NOV72f, CG59693-05 DNA Sequence	SEQ ID NO: 1057	1219 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 993

TGCTAACCAGGCCAGTGACAGAAATGGATTCGAAATACCAGTGTGTGAAGCTGAATGATGGTCACTTC
ATGCCTGTCCTGGGATTTGGCACCTATGCGCCTGCAGAGGTTCCTAAAAGTAAAGCTCTAGAGGCCGT
CAAATTGGCAATAGAAGCCGGGTTCCACCATATTGATTCTGCACATGTTTACAATAATGAGGAGCAGG
TTGGACTGGCCATCCGAAGCAAGATTGCAGATGGCAGTGTGAAGAGAGAAGACATATTCTACACTTCA
AAGCTTTGGAGCAATTCCCATCGACCAGAGTTGGTCCGACCAGCCTTGGAAAGGTCACTGAAAAATCT
TCAATTGGACTATGTTGACCTCTATCTTATTCATTTTCCAGTGTCTGTAAAGCCAGGTGAGGAAGTGA
TCCCAAAAGATGAAAATGGAAAAATACTATTTGACACAGTGGATCTCTGTGCCACGTGGGAGGCCATG
GAGAAGTGTAAAGATGCAGGATTGGCCAAGTCCATCGGGGTGTCCAACTTCAACCACAGGCTGCTGGA
GATGATCCTCAACAAGCCAGGGCTCAAGTACAAGCCTGTCTGCAACCAGGTGGAATGTCATCCTTACT
TCAACCAGAGAAAACTGCTGGATTTCTGCAAGTCAAAAGACATTGTTCTGGTTGCCTATAGTGCTCTG
GGATCCCATCGAGAAGAACCATGGGTGGACCCGAACTCCCCGGTGCTCTTGGAGGACCCAGTCCTTTG
TGCCTTGGCAAAAAAGCACAAGCGAACCCCAGCCCTGATTGCCCTGCGCTACCAGCTGCAGCGTGGGG
TTGTGGTCCTGGCCAAGAGCTACAATGAGCAGCGCATCAGACAGAACGTGCAGGTGTTTGAATTCCAG
TTGACTTCAGAGGAGATGAAAGCCATAGATGGCCTAAACAGAAATGTGCGATATTTGACCCTTGATAT
TTTTGCTGGCCCCCCTAATTATCCATTTTCTGATGAATATTAACATGGAGGGCATTGCATGAGGTCTG
CCAGAAGGCCCTGCGTGTGGATGGTGACACAGAGGATGGCTCTATGCTGGTGACTGGACACATCGCCT
CTGGTTAAATCTCTCCTGCTTGGCGACTTCAGTAAGCTACAGCTAAGCCCATCGGCCGGAAAAGAAAG
ACAATAATTTTGTTTTTTCATTTTGAAAAAATTAAATGCTCTCTCCTAAAGATTCTTCACCTA

NOV72f, CG59693-05 Protein Sequence	SEQ ID NO: 1058	323 aa	MW at 36734.9kD

MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEAVKLAIEAGFHHIDSAHVYNNEEQVGLAIRSK
IADGSVKREDIFYTSKLWSNSHRPELVRPALERSLKNLQLDYVDLYLIHFPVSVKPGEEVIPKDENGK
ILFDTVDLCATWEAMEKCKDAGLAKSIGVSNFNHRLLEMILNKPGLKYKPVCNQVECHPYFNQRKLLD
FCKSKDIVLVAYSALGSHREEPWVDPNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSY
NEQRIRQNVQVFEFQLTSEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY

NOV72g, CG59693-06 DNA Sequence	SEQ ID NO: 1059	1001 bp
	ORF Start: at 11	ORF Stop: at 983

CATCTAGGCCACCATGGCCATGGATTCGAAATATCAGTGTGTGAAGCTGAATGATGGTCACTTCATGC
CTGTCCTGGGATTTGGCACCTATGCGCCTGCAGAGGTTCCTAAAAGTAAAGCTTTAGAGGCCACCAAA
TTGGCAATTGAAGCTGGCTTCCGCCATATTGATTCTGCTCATTTATACAATAATGAGGAGCAGGTTGG
ACTGGCCATCCGAAGCAAGATTGCAGATGGCAGTGTGAAGAGAGAAGACATATTCTACACTTCAAAGC
TTTGGTGCAATTCCCATCGACCAGAGTTGGTCCGACCAGCCTTGGAAAGGTCACTGAAAAATCTTCAA
TTGGATTATGTTGACCTCTACCTTATTCATTTTCCAGTGTCTGTAAAGCCAGGTGAGGAAGTGATCCC
AAAAGATGAAAATGGAAAAATACTATTTGACACAGTGGATCTCTGTGCCACGTGGGAGGCCGTGGAGA
AGTGTAAAGATGCAGGATTGGCCAAGTCCATCGGGGTGTCCAACTTCAACCGCAGGCAGCTGGAGATG
ATCCTCAACAAGCCAGGGCTCAAGTACAAGCCTGTCTGCAACCAGGTGGAATGTCATCCTTACTTCAA
CCAGAGAAAACTGCTGGATTTCTGCAAGTCAAAAGACATTGTTCTGGTTGCCTATAGTGCTCTGGGAT
CCCACCGAGAAGAACCATGGGTGGACCCGAACTCCCCGGTGCTCTTGGAGGACCCAGTCCTTTGTGCC
TTGGCAAAAAAGCACAAGCGAACCCCAGCCCTGATTGCCCTGCGCTACCAGCTACAGCGTGGGGTTGT
GGTCCTGGCCAAGAGCTACAATGAGCAGCGCATCAGACAGAACGTGCAGGTGTTTGAATTCCAGTTGA
CTTCAGAGGAGATGAAAGCCATAGATGGCCTAAACAGAAATGTGCGATATTTGACCCTTGATATTTTT
GCTGGCCCCCCTAATTATCCATTTTCTGATGAATATTAAACGCGTGATC

NOV72g, CG59693-06 Protein Sequence	SEQ ID NO: 1060	324 aa	MW at 36799.0kD

TMAMDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHIDSAHLYNNEEQVGLAI
RSKIADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLDYVDLYLIHFPVSVKPGEEVIPKDE
NGKILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFNRRQLEMILNKPGLKYKPVCNQVECHPYFNQRK
LLDFCKSKDIVLVAYSALGSHREEPWVDPNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLA
KSYNEQRIRQNVQVFEFQLTSEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSD

NOV72h, CG59693-07 DNA Sequence	SEQ ID NO: 1061	1012 bp
	ORF Start: at 1	ORF Stop: at 1012

GCCGGTACCACCATGGGCCACCATCACCACCATCACGATTCGAAATATCAGTGTGTGAAGCTGAATGA
TGGTCACTTCATGCCTGTCCTGGGATTTGGCACCTATGCGCCTGCAGAGGTTCCTAAAAGTAAAGCTT
TAGAGGCCACCAAATTGGCAATTGAAGCTGGCTTCCGCCATATTGATTCTGCTCATTTATACAATAAT
GAGGAGCAGGTTGGACTGGCCATCCGAAGCAAGATTGCAGATGGCAGTGTGAAGAGAGAAGACATATT
CTACACTTCAAAGCTTTGGTGCAATTCCCATCGACCAGAGTTGGTCCGACCAGCCTTGGAAAGGTCAC
TGAAAAATCTTCAATTGGATTATGTTGACCTCTACCTTATTCATTTTCCAGTGTCTGTAAAGCCAGGT
GAGGAAGTGATCCCAAAAGATGAAAATGGAAAAATACTATTTGACACAGTGGATCTCTGTGCCACGTG
GGAGGCCGTGGAGAAGTGTAAAGATGCAGGATTGGCCAAGTCCATCGGGGTGTCCAACTTCAACCGCA
GGCAGCTGGAGATGATCCTCAACAAGCCAGGGCTCAAGTACAAGCCTGTCTGCAACCAGGTGGAATGT
CATCCTTACTTCAACCAGAGAAAACTGCTGGATTTCTGCAAGTCAAAAGACATTGTTCTGGTTGCCTA
TAGTGCTCTGGGATCCCACCGAGAAGAACCATGGGTGGACCCGAACTCCCCGGTGCTCTTGGAGGACC
CAGTCCTTTGTGCCTTGGCAAAAAAGCACAAGCGAACCCCAGCCCTGATTGCCCTGCGCTACCAGCTA
CAGCGTGGGGTTGTGGTCCTGGCCAAGAGCTACAATGAGCAGCGCATCAGACAGAACGTGCAGGTGTT
TGAATTCCAGTTGACTTCAGAGGAGATGAAAGCCATAGATGGCCTAAACAGAAATGTGCGATATTTGA
CCCTTGATATTTTTGCTGGCCCCCCTAATTATCCATTTTCTGATGAATATCTCGAGGGTG

NOV72h, CG59693-07 Protein Sequence	SEQ ID NO: 1062	337 aa	MW at 38297.5kD

AGTTMGHHHHHHDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHIDSAHLYNN
EEQVGLAIRSKIADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLDYVDLYLIHFPVSVKPG
EEVIPKDENGKILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFNRRQLEMILNKPGLKYKPVCNQVEC
HPYFNQRKLLDFCKSKDIVLVAYSALGSHREEPWVDPNSPVLLEDPVLCALAKKHKRTPALIALRYQL
QRGVVVLAKSYNEQRIRQNVQVFEFQLTSEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEYLEG

NOV72i, CG59693-08 DNA Sequence	SEQ ID NO: 1063	1225 bp
	ORF Start: ATG at 24	ORF Stop: TAA at 993

TGCTAACCAGGCCAGTGACAGAAATGGATTCGAAATACCAGTGTGTGAAGCTGAATGATGGTCACTTC
ATGCCTGTCCTGGGATTTGGCACCTATGCGCCTGCAGAGGTTCCTAAAAGTAAAGCTCTAGAGGCCGT
CAAATTGGCAATAGAAGCCGGGTTCCACCATATTGATTCTGCACATGTTTACAATAATGAGGAGCAGG
TTGGACTGGCCATCCGAAGCAAGATTGCAGATGGCAGTGTGAAGAGAGAAGACATATTCTACACTTCA
AAGCTTTGGAGCAATTCCCATCGACCAGAGTTGGTCCGACCAGCCTTGGAAAGGTCACTGAAAAATCT

TCAATTGGACTATGTTGACCTCTATCTTATTCATTTTCCAGTGTCTGTAAAGCCAGGTGAGGAAGTGA
TCCCAAAAGATGAAAATGGAAAAATACTATTTGACACAGTGGATCTCTGTGCCACGTGGGAGGCCATG
GAGAAGTGTAAAGATGCAGGATTGGCCAAGTCCATCGGGGTGTCCAACTTCAACCACAGGCTGCTGGA
GATGATCCTCAACAAGCCAGGGCTCAAGTACAAGCCTGTCTGCAACCAGGTGGAATGTCATCCTTACT
TCAACCAGAGAAAACTGCTGGATTTCTGCAAGTCAAAAGACATTGTTCTGGTTGCCTATAGTGCTCTG
GGATCCCATCGAGAAGAACCATGGGTGGACCCGAACTCCCCGGTGCTCTTGGAGGACCCAGTCCTTTG
TGCCTTGGCAAAAAAGCACAAGCGAACCCCAGCCCTGATTGCCCTGCGCTACCAGCTGCAGCGTGGGG
TTGTGGTCCTGGCCAAGAGCTACAATGAGCAGCGCATCAGACAGAACGTGCAGGTGTTTGAATTCCAG
TTGACTTCAGAGGAGATGAAAGCCATAGATGGCCTAAACAGAAATGTGCGATATTTGACCCTTGATAT
TTTTGCTGGCCCCCCTAATTATCCATTTTCTGATGAATATTAACATGGAGGGCATTGCATGAGGTCTG
CCAGAAGGCCCTGCGTGTGGATGGTGACACAGAGGATGGCTCTATGCTGGTGACTGGACACATCGCCT
CTGGTTAAATCTCTCCTGCTTGGCGACTTCAGTAAGCTACAGCTAAGCCCATCGGCCGGAAAAGAAAG
ACAATAATTTTGTTTTTTCATTTTGAAAAAATTAAATGCTCTCTCCTAAAGATTCTTCACCTAAAAAA
A

NOV72i, CG59693-08 Protein Sequence	SEQ ID NO: 1064	323 aa	MW at 36734.9kD

MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEAVKLAIEAGFHHIDSAHVYNNEEQVGLAIRSK
IADGSVKREDIFYTSKLWSNSHRPELVRPALERSLKNLQLDYVDLYLIHFPVSVKPGEEVIPKDENGK
ILFDTVDLCATWEAMEKCKDAGLAKSIGVSNFNHRLLEMILNKPGLKYKPVCNQVECHPYFNQRKLLD
FCKSKDIVLVAYSALGSHREEPWVDPNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSY
NEQRIRQNVQVFEFQLTSEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY

NOV72j, CG59693-09 DNA Sequence	SEQ ID NO: 1065	996 bp
	ORF Start: ATG at 16	ORF Stop: TAG at 985

CACCGCGGCCGCACC**ATG**GATTCGAAATATCAGTGTGTGAAGCTGAATGATGGTCACTTCATGCCTGT
CCTGGGATTTGGCACCTATGCGCCTGCAGAGGTTCCTAAAAGTAAAGCTTTAGAGGCCACCAAATTGG
CAATTGAAGCTGGCTTCCGCCATATTGATTCTGCTCATTTATACAATAATGAGGAGCAGGTTGGACTG
GCCATCCGAAGCAAGATTGCAGATGGCAGTGTGAAGAGAGAAGACATATTCTACACTTCAAAGCTTTG
GTGCAATTCCCATCGACCAGAGTTGGTCCGACCAGCCTTGGAAAGGTCACTGAAAAATCTTCAATTGG
ATTATGTTGACCTCTACCTTATTCATTTTCCAGTGTCTGTAAAGCCAGGTGAGGAAGTGATCCCAAAA
GATGAAAATGGAAAAATACTATTTGACACAGTGGATCTCTGTGCCACGTGGGAGGCCGTGGAGAAGTG
TAAAGATGCAGGATTGGCCAAGTCCATCGGGGTGTCCAACTTCAACCGCAGGCAGCTGGAGATGATCC
TCAACAAGCCAGGGCTCAAGTACAAGCCTGTCTGCAACCAGGTGGAATGTCATCCTTACTTCAACCAG
AGAAAACTGCTGGATTTCTGCAAGTCAAAAGACATTGTTCTGGTTGCCTATAGTGCTCTGGGATCCCA
CCGAGAAGAACCATGGGTGGACCCGAACTCCCCGGTGCTCTTGGAGGACCCAGTCCTTTGTGCCTTGG
CAAAAAAGCACAAGCGAACCCCAGCCCTGATTGCCCTGCGCTACCAGCTACAGCGTGGGGTTGTGGTC
CTGGCCAAGAGCTACAATGAGCAGCGCATCAGACAGAACGTGCAGGTGTTTGAATTCCAGTTGACTTC
AGAGGAGATGAAAGCCATAGATGGCCTAAACAGAAATGTGCGATATTTGACCCTTGATATTTTTGCTG
GCCCCCCTAATTATCCATTTTCTGATGAATAT**TAG**GTCGACGGC

NOV72j, CG59693-09 Protein Sequence	SEQ ID NO: 1066	323 aa	MW at 36787.9kD

MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHIDSAHLYNNEEQVGLAIRSK
IADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLDYVDLYLIHFPVSVKPGEEVIPKDENGK
ILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFNRRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLD
FCKSKDIVLVAYSALGSHREEPWVDPNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSY
NEQRIRQNVQVFEFQLTSEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 72B.

Table 72B. Comparison of the NOV72 protein sequences.

```
NOV72a  ----------MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHI
NOV72b  ----------MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHI
NOV72c  -------ATMAMDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHI
NOV72d  ----------------MMVTSCLSWDLAPMRLQRLLP-QVPKSKALEATKLAIEAGFRHI
NOV72e  ----------MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHI
NOV72f  ----------MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEAVKLAIEAGFHHI
```

```
NOV72g  --------TMAMDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHI
NOV72h  AGTTMGHHHHHHDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHI
NOV72i  -----------MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEAVKLAIEAGFHHI
NOV72j  -----------MDSKYQCVKLNDGHFMPVLGFGTYAPAEVPKSKALEATKLAIEAGFRHI

NOV72a  DSAHLYNNEEQVGLAIRSKIADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLD
NOV72b  DSAHLYNNEEQVGLAIRSKIADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLD
NOV72c  DSAHLYNNEEQVGLAIRSKIADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLD
NOV72d  DSAHLYNNEEQVGLAIRSKIADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLD
NOV72e  DSAHLYNNEEQVGLAIRSKIADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLD
NOV72f  DSAHVYNNEEQVGLAIRSKIADGSVKREDIFYTSKLWSNSHRPELVRPALERSLKNLQLD
NOV72g  DSAHLYNNEEQVGLAIRSKIADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLD
NOV72h  DSAHLYNNEEQVGLAIRSKIADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLD
NOV72i  DSAHVYNNEEQVGLAIRSKIADGSVKREDIFYTSKLWSNSHRPELVRPALERSLKNLQLD
NOV72j  DSAHLYNNEEQVGLAIRSKIADGSVKREDIFYTSKLWCNSHRPELVRPALERSLKNLQLD

NOV72a  YVDLYLIHFPVSVKPGEEVIPKDENGKILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFN
NOV72b  YVDLYLIHFPVSVKPGEEVIPKDENGKILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFN
NOV72c  YVDLYLIHFPVSVKPGEEVIPKDENGKILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFN
NOV72d  YVDLYLIHFPVSVKPGEEVIPKDESGKILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFN
NOV72e  YVDLYLIHFPVSVKPGEEVIPKDENGKILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFN
NOV72f  YVDLYLIHFPVSVKPGEEVIPKDENGKILFDTVDLCATWEAMEKCKDAGLAKSIGVSNFN
NOV72g  YVDLYLIHFPVSVKPGEEVIPKDENGKILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFN
NOV72h  YVDLYLIHFPVSVKPGEEVIPKDENGKILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFN
NOV72i  YVDLYLIHFPVSVKPGEEVIPKDENGKILFDTVDLCATWEAMEKCKDAGLAKSIGVSNFN
NOV72j  YVDLYLIHFPVSVKPGEEVIPKDENGKILFDTVDLCATWEAVEKCKDAGLAKSIGVSNFN

NOV72a  RRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLDFCKSKDIVLVAYSALGSHREEPWVD
NOV72b  RRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLDFCKSKDIVLVAYSALGSHREEPWVD
NOV72c  RRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLDFCKSKDIVLVAYSALGSHREEPWVD
NOV72d  RRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLDFCKSKDIVLVAYSALGSHREEPWVD
NOV72e  RRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLDFCKSKDIVLVAYSALGSHREEPWVD
NOV72f  HRLLEMILNKPGLKYKPVCNQVECHPYFNQRKLLDFCKSKDIVLVAYSALGSHREEPWVD
NOV72g  RRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLDFCKSKDIVLVAYSALGSHREEPWVD
NOV72h  RRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLDFCKSKDIVLVAYSALGSHREEPWVD
NOV72i  HRLLEMILNKPGLKYKPVCNQVECHPYFNQRKLLDFCKSKDIVLVAYSALGSHREEPWVD
NOV72j  RRQLEMILNKPGLKYKPVCNQVECHPYFNQRKLLDFCKSKDIVLVAYSALGSHREEPWVD

NOV72a  PNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSYNEQRIRQNVQVFEFQLT
NOV72b  PNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKRYNEQRIRQNVQVFEFQLT
NOV72c  PNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSYNEQRIRQNVQVFEFQLT
NOV72d  PNSPVLLEDPVLCALAKKHKRTPALVALRYQLQRGVVVLAKSYNEQRIRQNVQVFEFQLT
NOV72e  PNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSYNEQRIRQNVQVFEFQLT
NOV72f  PNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSYNEQRIRQNVQVFEFQLT
NOV72g  PNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSYNEQRIRQNVQVFEFQLT
NOV72h  PNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSYNEQRIRQNVQVFEFQLT
NOV72i  PNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSYNEQRIRQNVQVFEFQLT
NOV72j  PNSPVLLEDPVLCALAKKHKRTPALIALRYQLQRGVVVLAKSYNEQRIRQNVQVFEFQLT

NOV72a  SEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY---
NOV72b  SEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY---
NOV72c  SEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY---
NOV72d  SEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY---
NOV72e  SEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSD-----
NOV72f  SEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY---
NOV72g  SEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSD-----
NOV72h  SEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEYLEG
```

```
NOV72i  SEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY---
NOV72j  SEEMKAIDGLNRNVRYLTLDIFAGPPNYPFSDEY---

NOV72a  (SEQ ID NO:  1048)
NOV72b  (SEQ ID NO:  1050)
NOV72c  (SEQ ID NO:  1052)
NOV72d  (SEQ ID NO:  1054)
NOV72e  (SEQ ID NO:  1056)
NOV72f  (SEQ ID NO:  1058)
NOV72g  (SEQ ID NO:  1060)
NOV72h  (SEQ ID NO:  1062)
NOV72i  (SEQ ID NO:  1064)
NOV72j  (SEQ ID NO:  1066)
```

Further analysis of the NOV72a protein yielded the following properties shown in Table 72C.

Table 72C. Protein Sequence Properties NOV72a

SignalP analysis:	No Known Signal Sequence Indicated

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 9;  pos.chg 2;  neg.chg 1
      H-region:  length 2;  peak value  -3.30
      PSG score:  -7.70

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -5.32
      possible cleavage site: between 27 and 28

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  3.66 (at 253)
      ALOM score:   3.66  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:        0       Hyd Moment(75):  0.10
      Hyd Moment(95):  4.27     G content:       0
      D/E content:     2        S/T content:     1
      Score: -7.79

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: KKHK (3) at  246
      pat4: KHKR (3) at  247
      pat7: none
      bipartite: none
      content of basic residues:  13.3%
```

```
      NLS Score: -0.10

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 76.7

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

-------------------------

Final Results (k = 9/23):

       34.8 %: cytoplasmic
       30.4 %: mitochondrial
       30.4 %: nuclear
        4.3 %: vacuolar

>> prediction for CG59693-01 is cyt (k=23)
```

A search of the NOV72a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 72D.

Table 72D. Geneseq Results for NOV72a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV72a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB43444	Human cancer associated protein sequence SEQ ID NO:889 - Homo sapiens, 336 aa. [WO200055350-A1, 21-SEP-2000]	1..323 14..336	318/323 (98%) 318/323 (98%)	0.0
AAU85559	Clone #59314 (L1426P) of lung tumour protein - Homo sapiens, 323 aa. [WO200204514-A2, 17-JAN-2002]	1..323 1..323	311/323 (96%) 316/323 (97%)	0.0
ABB75050	Human lung tumour L773P recombinant protein sequence SEQ ID NO:433 - Homo sapiens, 371 aa. [WO200200174-A2, 03-JAN-2002]	29..323 77..371	288/295 (97%) 290/295 (97%)	e-168
ABB74958	Human lung tumour L773P protein sequence SEQ ID NO:172 - Homo sapiens, 364 aa. [WO200200174-A2, 03-JAN-2002]	29..323 70..364	288/295 (97%) 290/295 (97%)	e-168
AAU85520	L773P lung tumour protein - Homo sapiens, 364 aa. [WO200204514-A2, 17-JAN-2002]	29..323 70..364	288/295 (97%) 290/295 (97%)	e-168

In a BLAST search of public sequence databases, the NOV72a protein was found to have homology to the proteins shown in the BLASTP data in Table 72E.

Table 72E. Public BLASTP Results for NOV72a				
Protein Accession Number	**Protein/Organism/Length**	**NOV72a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q04828	Aldo-keto reductase family 1 member C1 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (High-affinity hepatic bile acid-binding protein) (HBAB) (Chlordecone reductase homolog HAKRC) (Dihydrodiol dehydrogenase 2) (DD2) (20 alpha- hydroxysteroid dehydrogenase) - Homo sapiens (Human), 323 aa.	1..323 1..323	323/323 (100%) 323/323 (100%)	0.0
P52895	Aldo-keto reductase family 1 member C2 (EC 1.1.1.-) (Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase) (EC 1.3.1.20) (Chlordecone reductase homolog HAKRD) (Dihydrodiol dehydrogenase/bile acid-binding protein) (DD/BABP) (Dihydrodiol dehydrogenase 2) (DD2) - Homo sapiens (Human), 323 aa.	1..323 1..323	316/323 (97%) 318/323 (97%)	0.0
I73676	chlordecone reductase homolog (clone HAKRd) - human, 323 aa.	1..323 1..323	313/323 (96%) 317/323 (97%)	0.0
I73675	chlordecone reductase homolog (clone HAKRc) - human, 320 aa (fragment).	4..323 1..320	312/320 (97%) 313/320 (97%)	0.0
Q95JH6	3(20)alpha-hydroxysteroid/dihydrodiol/indanol dehydrogenase (EC 1.1.1.112) - Macaca fuscata (Japanese macaque), 323 aa.	1..323 1..323	304/323 (94%) 319/323 (98%)	0.0

PFam analysis indicates that the NOV72a protein contains the domains shown in the Table 72F.

Table 72F. Domain Analysis of NOV72a			
Pfam Domain	**NOV72a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
aldo_ket_red	10..303	164/369 (44%) 274/369 (74%)	3.4e-156

Example 73.

The NOV73 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 73A.

Table 73A. NOV73 Sequence Analysis		
NOV73a, CG59839-02 DNA Sequence	SEQ ID NO: 1067	4052 bp
	ORF Start: ATG at 786	ORF Stop: TAG at 4023

CTTCTGCCTCAGCCTCCCGAGTAGCTGGGATGGGACTACAGATGCACGCCACCACACCTGGCAATTTT
TTTTAGTTTTGTGGAGACAAGGTCCCATGTGAGCCACTGTGCCCCACCTTTTAAAGAAATTACCATAA
GGAAGCAACTGTCTTTAGCAAATAGAAGATCATAAACATCCTTCTGGAGATCTTGAGATGTTGACTGC
GGTTTATAGGGCAATGAAAATCTGCCTACCTGAGTGGGCAATGAGACGGGAGTGCCACCTGTGATCCT
TCCAGGGAGCAATTCTTTCTGGTGCCAGGAGCCCAAAGGAAACTCAGTCAACCAGCTACAGATGGGCT
TGCTCCGAGTTTGTTCTAGAAGTCAGAACTGCCAGCTCTGAAAATAGGTCACAAGTGACAGGCCGTCC
CTGGGCTGCTCACAGGAAGTAGCTGCAAAGGTCAGGGAAAGTTTTTAACGACTATACAGGCCAGTGAA
TGACAGGCTTTGAGGAGCTTCTCTTCTTGGCATCTGAAACAGCATACTTTCTGACTACCGCTGTGTGC
TGTGAGGGGCTGAGGATCCAGACGTGGCTGAGGAAGGTCCCGGCTCATGAGGAGGTGGAAGCCTGGGT
GCTTAAACAGATAATCCAGAGTTTGGATGTGAGGTGGTGAATCACTCTGGCTTCATTTGGTTATACTG
GTCAAGAGTTTGGCTCTGGAGCAACGTGTCTGGGTTTGAATCCTGGCTGTGCTTTTGGTGATCCTCCA
TTGTTTTGAAGCACAAGACCTTTAAACCACTGTAGGTATGGACAAGGAAGAAAGGAAGATCATCAATC
AGGGTCAAGAAGATGAAATGGAGATTTATGGTTACAATTTGAGTCGCTGGAAGCTTGCCATAGTTTCT
TTAGGAGTGATTTGCACTGGTGGGTTTCTCCTCCTCCTCCTCTATTGGATGCCTGAGTGGCGGGTGAA
AGCGACCTGTGTCAGAGCTGCAATTAAAGACTGTGACGTAGTGCTGCTGAGGACTACTGATGAATTCA
AAATGTGGTTTTGTGCAAAAATTCGCGTTCTTTCTTTGGAAACTCACCCAATTTCAAGTCCAAAATCT
ATGTCTAATAAGCTTTCAAATGGCCATGCAGTTTGTTTAACTGAGAATCCCACTGGAGAAAATAGGCA
CGGGATCAGTAAATATTCACAGGCTGAATCACAACAGATTCGTTATTTCACCCATCATAGTGTAAAAT
ATTTCTGGAATGATACCATTCACAATTTTGATTTCTTAAAGGGACTGGATGAAGGTGTTTCTTGTACG
TCAATTTATGAAAAGCATAGTGCAGGACTGACAAAGGGGATGCATGCCTACAGAAAACTGCTTTATGG
AGTAAATGAAATTGCTGTAAAAGTGCCTTCTGTTTTTAAGCTTCTAATTAAAGAGGTTCTCAACCCAT
TTTACATTTTCCAGCTGTTCAGTGTTATACTGTGGAGCACTGATGAATACTATTACTATGCTCTAGCT
ATTGTGGTTATGTCCATAGTATCAATCGTAAGCTCACTGTATTCCATTAGAAAGCAATATGTTATGTT
GCATGACATGGTGGCAACTCATAGTACTGTAAGAGTTTCAGTTTGTAGAGTAAATGAAGAAATAGAAG
AAATCTTTTCTACTGACCTTGTGCCAGGAGATGTCATGGTCATTCCATTAAATGGGACAATAATGCCT
TGCGATGCTGTACTTATTAATGGTACCTGCATTGTAAATGAAAGCATGTTAACAGGAGAAAGTGTTCC
AGTGACAAAGACTAACTTGCCAAATCCTTCGGTGGATGTGAAAGGAATAGGAGATGAATTATATAATC
CAGAAACACATAAACGACATACTTTGTTTTGTGGGACAACTGTTATTCAGACTCGTTTCTACACCGGA
GAACTCGTCAAAGCCATAGTTGTTAGAACAGGATTTAGTACTTCCAAAGGACAGCTTGTTCGTTCCAT
ATTGTATCCCAAACCAACTGATTTTAAACTCTACAGAGATGCCTACTTGTTTCTACTATGTCTTGTGG
CAGTTGCTGGCATTGGGTTTATCTACACTATTATTAATAGCATTTTAAATGAGGTACAAGTTGGGGTC
ATAATTATCGAGTCTCTTGATATTATCACAATTACTGTGCCCCCTGCACTTCCTGCTGCAATGACTGC
TGGTATTGTGTATGCTCAGAGAAGACTGAAAAAAATCGGTATTTTCTGTATCAGTCCTCAAAGAATAA
ATATTTGTGGACAGCTCAATCTTGTTTGCTTTGACAAGACTGGAACTCTAACTGAAGATGGTTTAAAT
CTTTGGGGGATTCAACGAGTGGAAAATGCACGGTTTCTTTCACCAGAAGAAAATGTGTGCAATGAGAT
GTTGGTAAAATCCCAGTTTGTTGCTTGTATGGCTACTTGTCATTCACTTACAAAAATTGAAGGAGTGC
TCTCTGGTGATCCACTTGATCTGAAAATGTTTGAGGCTATTGGATGGATTCTGGAAGAAGCAACTGAA
GAAGAAACAGCACTTCATAATCGAATTATGCCCACAGTGGTTCGTCCTCCCAAACAACTGCTTCCTGA
ATCTACCCCTGCAGGAAACCAAGAAATGGCTACTTATGAGATAGGAATTGTTCGCCAGTTCCCATTTT
CTTCTGCTTTGCAACGTATGAGTGTGGTTGCCAGGGTGCTGGGGGATAGGAAAATGGACGCCTACATG

AAAGGAGCGCCCGAGGCCATTGCCGGTCTCTGTAAACCTGAAACAGTTCCTGTCGATTTTCAAAACGT
TTTGGAAGACTTCACTAAACAGGGCTTCCGTGTGATTGCTCTTGCACACAGAAAATTGGAGTCAAAAC
TGACATGGCATAAAGTACAGAATATTAGCAGGGATGCAATTGAGAACAACATGGATTTTATGGGATTA
ATTATAATGCAGAACAAATTAAAGCAAGAAACCCCTGCAGTACTTGAAGATTTGCATAAAGCCAACAT
TCGCACCGTCATGGTCACAGGTGACAGTATGTTGACTGCTGTCTCTGTGGCCAGAGATTGTGGAATGA
TTCTACCTCAGGATAAAGTGATTATTGCTGAAGCATTACCTCCAAAGGATGGGAAAGTTGCCAAAATA
AATTGGCATTATGCAGACTCCCTCACGCAGTGCAGTCATCCATCAGCAATTGACCCAGAGGCTATTCC
GGTTAAATTGGTCCATGATAGCTTAGAGGATCTTCAAATGACTCGTTATCATTTTGCAATGAATGGAA
AATCATTCTCAGTGATACTGGAGCATTTTCAAGACCTTGTTCCTAAGTTGATGTTGCATGGCACCGTG

TTTGCCCGTATGGCACCTGATCAGAAGACACAGTTGATAGAAGCATTGCAAAATGTTGAGTATTTTGT
TGGGATGTGTGGTGATGGCGCAAATGATTGTGGTGCTTTGAAGAGGGCACACGGAGGCATTTCCTTAT
CGGAGCTCGAAGCTTCAGTGGCATCTCCCTTTACCTCTAAGACTCCTAGTATTTCCTGTGTGCCAAAC
CTTATCAGGGAAGGCCGTGCTGCTTTAATAACTTCCTTCTGTGTGTTTAAATTCATGGCATTGTACAG
CATTATCCAGTACTTCAGTGTTACTCTGCTGTATTCTATCTTAAGTAACCTAGGAGACTTCCAGTTTC
TCTTCATTGATCTGGCAATCATTTTGGTAGTGGTATTTACAGTGAGTTTAAATCCTGCCTGGAAAGAA
CTTGTGGCACAAAGACCACCTTCGGGTCTTATATCTGGGGCCCTTCTCTTCTCCGTTTTGTCTCAGAT
TATCATCTGCATTGGATTTCAATCTTTGGGTTTTTTTTGGGTCAAACAGCAACCTTGGTATGAAGTGT
GGCATCCAAAATCAGAGGCTTGTAATACAACAGGAAGCGGGTTTTGGAATTCTTCACACGTAGACAAT
GAAACCGAACTTGATGAAATAATATACAAAATTATGAAAATACCACAGTGTTTTTTATTTCCAGTTTT
CAGTACCTCATAGTGGCAATTGCCTTTTCAAAAGGAAAAC

NOV73a, CG59839-02 Protein Sequence	SEQ ID NO: 1068	1079 aa	MW at 120773.2kD

MDKEERKIINQGQEDEMEIYGYNLSRWKLAIVSLGVICTGGFLLLLLYWMPEWRVKATCVRAAIKDCD
VVLLRTTDEFKMWFCAKIRVLSLETHPISSPKSMSNKLSNGHAVCLTENPTGENRHGISKYSQAESQQ
IRYFTHHSVKYFWNDTIHNFDFLKGLDEGVSCTSIYEKHSAGLTKGMHAYRKLLYGVNEIAVKVPSVF
KLLIKEVLNPFYIFQLFSVILWSTDEYYYYALAIVVMSIVSIVSSLYSIRKQYVMLHDMVATHSTVRV
SVCRVNEEIEEIFSTDLVPGDVMVIPLNGTIMPCDAVLINGTCIVNESMLTGESVPVTKTNLPNPSVD
VKGIGDELYNPETHKRHTLFCGTTVIQTRFYTGELVKAIVVRTGFSTSKGQLVRSILYPKPTDFKLYR
DAYLFLLCLVAVAGIGFIYTIINSILNEVQVGVIIIESLDIITITVPPALPAAMTAGIVYAQRRLKKI
GIFCISPQRINICGQLNLVCFDKTGTLTEDGLNLWGIQRVENARFLSPEENVCNEMLVKSQFVACMAT
CHSLTKIEGVLSGDPLDLKMFEAIGWILEEATEEETALHNRIMPTVVRPPKQLLPESTPAGNQEMATY
EIGIVRQFPFSSALQRMSVVARVLGDRKMDAYMKGAPEAIAGLCKPETVPVDFQNVLEDFTKQGFRVI
ALAHRKLESKLTWHKVQNISRDAIENNMDFMGLIIMQNKLKQETPAVLEDLHKANIRTVMVTGDSMLT
AVSVARDCGMILPQDKVIIAEALPPKDGKVAKINWHYADSLTQCSHPSAIDPEAIPVKLVHDSLEDLQ
MTRYHFAMNGKSFSVILEHFQDLVPKLMLHGTVFARMAPDQKTQLIEALQNVEYFVGMCGDGANDCGA
LKRAHGGISLSELEASVASPFTSKTPSISCVPNLIREGRAALITSFCVFKFMALYSIIQYFSVTLLYS
ILSNLGDFQFLFIDLAIILVVVFTVSLNPAWKELVAQRPPSGLISGALLFSVLSQIIICIGFQSLGFF
WVKQQPWYEVWHPKSEACNTTGSGFWNSSHVDNETELDEIIYKIMKIPQCFLFPVFSTS

NOV73b, CG59839-01 DNA Sequence	SEQ ID NO: 1069	2649 bp
	ORF Start: ATG at 183	ORF Stop: TAG at 2055

CAACTGATTTTAAACTCTACAGAGATGCCTACTTGTTTCTACTATGTCTTGTGGCAGTTGCTGGCATT
GGGTTTATCTACACTATTATTAATAGCATTTTAAATGAGGTACAAGTTGGGGTCATAATTATCGAGTC
TCTTGATATTATCACAATTACTGTGCCCCCTGCACTTCCTGCTGCAATGACTGCTGGTATTGTGTATG
CTCAGAGAAGACTGAAAAAAATCGGTATTTTCTGTATCAGTCCTCAAAGAATAAATATTTGTGGACAG
CTCAATCTTGTTTGCTTTGACAAGACTGGAACTCTAACTGAAGATGGTTTAGATCTTTGGGGGATTCA
ACGAGTGGAAAATGCACGATTTCTTTCACCAGAAGAAAATGTGTGCAATGAGATGTTGGTAAAATCCC
AGTTTGTTGCTTGTATGGCTACTTGTCATTCACTTACAAAAATTGAAGGAGTGCTCTCTGGTGATCCA
CTTGATCTGAAAATGTTTGAGGCTATTGGATGGATTCTGGAAGAAGCAACTGAAGAAGAAACAGCACT
TCATAATCGAATTATGCCCACAGTGGTTCGTCCTCCCAAACAACTGCTTCCTGAATCTACCCCTGCAG
GAAACCAAGAAATGGAGCTGTTTGAACTTCCAGCTACTTATGAGATAGGAATTGTTCGCCAGTTCCCA
TTTTCTTCTGCTTTGCAACGTATGAGTGTGGTTGCCAGGGTGCTGGGGGATAGGAAAATGGACGCCTA
CATGAAAGGAGCGCCCGAGGCCATTGCCGGTCTCTGTAAACCTGAAACAGTTCCTGTCGATTTTCAAA
ACGTTTTGGAAGACTTCACTAAACAGGGCTTCCGTGTGATTGCTCTTGCACACAGAAAATTGGAGTCA
AAACTGACATGGCATAAAGTACAGAATATTAGCAGAGATGCAATTGAGAACAACATGGATTTTATGGG
ATTAATTATAATGCAGAACAAATTAAAGCAAGAAACCCCTGCAGTACTTGAAGATTTGCATAAAGCCA
ACATTCGCACCGTCATGGTCACAGGTGACAGTATGTTGACTGCTGTCTCTGTGGCCAGAGATTGTGGA
ATGATTCTACCTCAGGATAAAGTGATTATTGCTGAAGCATTACCTCCAAAGGATGGGAAAGTTGCCAA
AATAAATTGGCATTATGCAGACTCCCTCACGCAGTGCAGTCATCCATCAGCAATTGACCCAGAGGCTA
TTCCGGTTAAATTGGTCCATGATAGCTTAGAGGATCTTCAAATGACTCGTTATCATTTTGCAATGAAT
GGAAAATCATTCTCAGTGATACTGGAGCATTTTCAAGACCTTGTTCCTAAGTTGATGTTGCATGGCAC

CGTGTTTGCCCGTATGGCACCTGATCAGAAGACACAGTTGATAGAAGCATTGCAAAATGTTGATTATT
TTGTTGGGATGTGTGGTGATGGCGCAAATGATTGTGGTGCTTTGAAGAGGGCACACGGAGGCATTTCC
TTATCGGAGCTCGAAGCTTCAGTGGCATCTCCCTTTACCTCTAAGACTCCTAGTATTTCCTGTGTGCC
AAACCTTATCAGGGAAGGCCGTGCTGCTTTAATAACTTCCTTCTGTGTGTTTAAATTCATGGCATTGT
ACAGCATTATCCAGTACTTCAGTGTTACTCTGCTGTATTCTATCTTAAGTAACCTAGGAGACTTCCAG
TTTCTCTTCATTGATCTGGCAATCATTTTGGTAGTGGTATTTACAATGAGTTTAAATCCTGCCTGGAA
AGAACTTGTGGCACAAAGACCACCTTCGGGTCTTATATCTGGGGCCCTTCTCTTCTCCGTTTTGTCTC
AGATTATCATCTGCATTGGATTTCAATCTTTGGGTTTTTTTTGGGTCAAACAGCAACCTTGGTATGAA
GTGTGGCATCCAAAATCAGATGCTTGTAATACAACAGGAAGCGGGTTTTGGAATTCTTCACACGTAGA
CAATGAAACCGAACTTGATGAACTAATATACAAAATTATGAAAATACCACAGTGTTTTTTATTTCCAG
TTTTCAGTACCTCATAGTGGCAATTGCCTTTTCAAAAGGAAAACCCTTCAGGCAACCTTGCTACAAAA
ATTATTTTTTTGTTTTTTCTGTGATTTTTTTATATATTTTTATATTATTCATCATGTTGTATCCAGTT
GCCTCTGTTGACCAGGTTCTTCAGATAGTGTGTGTACCATATCAGTGGCGTGTAACTATGCTCATCAT
TGTTCTTGTCAATGCCTTTGTGTCTATCACAGTGGAGGAGTCAGTGGATCGGTGGGGAAAATGCTGCT
TACCCTGGGCCCTGGGCTGTAGAAAGAAGACACCAAAGGCAAAGTACATGTATCTGGCGCAGGAGCTC
TTGGTTGATCCAGAATGGCCACCAAAACCTCAGACAACCACAGAAGCTAAAGCTTTAGTTAAGGAGAA
TGGATCATGTCAAATCATCACCATAACATAGCAGTGAATCAGTCTCAGTGGTATTGCTGATAGCAGTA
TTCAGGAATATGTGATTTTAGGAGTTTCTGATCCTGTGTGTCAGAATGGCACTAGTTCAGTTTATGTC
CCTTCTGATATAGTAGCTTATTTGACAGCTTTGCTCTTCCTTAAAATAAAAAAAAAAAAAAAAAA

NOV73b, CG59839-01 Protein Sequence	SEQ ID NO: 1070	624 aa	MW at 69590.1kD

MTAGIVYAQRRLKKIGIFCISPQRINICGQLNLVCFDKTGTLTEDGLDLWGIQRVENARFLSPEENVC
NEMLVKSQFVACMATCHSLTKIEGVLSGDPLDLKMFEAIGWILEEATEEETALHNRIMPTVVRPPKQL
LPESTPAGNQEMELFELPATYEIGIVRQFPFSSALQRMSVVARVLGDRKMDAYMKGAPEAIAGLCKPE
TVPVDFQNVLEDFTKQGFRVIALAHRKLESKLTWHKVQNISRDAIENNMDFMGLIIMQNKLKQETPAV
LEDLHKANIRTVMVTGDSMLTAVSVARDCGMILPQDKVIIAEALPPKDGKVAKINWHYADSLTQCSHP
SAIDPEAIPVKLVHDSLEDLQMTRYHFAMNGKSFSVILEHFQDLVPKLMLHGTVFARMAPDQKTQLIE
ALQNVDYFVGMCGDGANDCGALKRAHGGISLSELEASVASPFTSKTPSISCVPNLIREGRAALITSFC
VFKFMALYSIIQYFSVTLLYSILSNLGDFQFLFIDLAIILVVVFTMSLNPAWKELVAQRPPSGLISGA
LLFSVLSQIIICIGFQSLGFFWVKQQPWYEVWHPKSDACNTTGSGFWNSSHVDNETELDELIYKIMKI
PQCFLFPVFSTS

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 73B.

Table 73B. Comparison of the NOV73 protein sequences.

```
NOV73a  MDKEERKIINQGQEDEMEIYGYNLSRWKLAIVSLGVICTGGFLLLLLYWMPEWRVKATCV
NOV73b  ------------------------------------------------------------

NOV73a  RAAIKDCDVVLLRTTDEFKMWFCAKIRVLSLETHPISSPKSMSNKLSNGHAVCLTENPTG
NOV73b  ------------------------------------------------------------

NOV73a  ENRHGISKYSQAESQQIRYFTHHSVKYFWNDTIHNFDFLKGLDEGVSCTSIYEKHSAGLT
NOV73b  ------------------------------------------------------------

NOV73a  KGMHAYRKLLYGVNEIAVKVPSVFKLLIKEVLNPFYIFQLFSVILWSTDEYYYYALAIVV
NOV73b  ------------------------------------------------------------

NOV73a  MSIVSIVSSLYSIRKQYVMLHDMVATHSTVRVSVCRVNEEIEEIFSTDLVPGDVMVIPLN
NOV73b  ------------------------------------------------------------

NOV73a  GTIMPCDAVLINGTCIVNESMLTGESVPVTKTNLPNPSVDVKGIGDELYNPETHKRHTLF
NOV73b  ------------------------------------------------------------

NOV73a  CGTTVIQTRFYTGELVKAIVVRTGFSTSKGQLVRSILYPKPTDFKLYRDAYLFLLCLVAV
NOV73b  ------------------------------------------------------------
```

```
NOV73a  AGIGFIYTIINSILNEVQVGVIIIESLDIITITVPPALPAAMTAGIVYAQRRLKKIGIFC
NOV73b  ----------------------------------------MTAGIVYAQRRLKKIGIFC

NOV73a  ISPQRINICGQLNLVCFDKTGTLTEDGLNLWGIQRVENARFLSPEENVCNEMLVKSQFVA
NOV73b  ISPQRINICGQLNLVCFDKTGTLTEDGLDLWGIQRVENARFLSPEENVCNEMLVKSQFVA

NOV73a  CMATCHSLTKIEGVLSGDPLDLKMFEAIGWILEEATEEETALHNRIMPTVVRPPKQLLPE
NOV73b  CMATCHSLTKIEGVLSGDPLDLKMFEAIGWILEEATEEETALHNRIMPTVVRPPKQLLPE

NOV73a  STPAGNQ------EMATYEIGIVRQFPFSSALQRMSVVARVLGDRKMDAYMKGAPEAIAG
NOV73b  STPAGNQEMELFELPATYEIGIVRQFPFSSALQRMSVVARVLGDRKMDAYMKGAPEAIAG

NOV73a  LCKPETVPVDFQNVLEDFTKQGFRVIALAHRKLESKLTWHKVQNISRDAIENNMDFMGLI
NOV73b  LCKPETVPVDFQNVLEDFTKQGFRVIALAHRKLESKLTWHKVQNISRDAIENNMDFMGLI

NOV73a  IMQNKLKQETPAVLEDLHKANIRTVMVTGDSMLTAVSVARDCGMILPQDKVIIAEALPPK
NOV73b  IMQNKLKQETPAVLEDLHKANIRTVMVTGDSMLTAVSVARDCGMILPQDKVIIAEALPPK

NOV73a  DGKVAKINWHYADSLTQCSHPSAIDPEAIPVKLVHDSLEDLQMTRYHFAMNGKSFSVILE
NOV73b  DGKVAKINWHYADSLTQCSHPSAIDPEAIPVKLVHDSLEDLQMTRYHFAMNGKSFSVILE

NOV73a  HFQDLVPKLMLHGTVFARMAPDQKTQLIEALQNVEYFVGMCGDGANDCGALKRAHGGISL
NOV73b  HFQDLVPKLMLHGTVFARMAPDQKTQLIEALQNVDYFVGMCGDGANDCGALKRAHGGISL

NOV73a  SELEASVASPFTSKTPSISCVPNLIREGRAALITSFCVFKFMALYSIIQYFSVTLLYSIL
NOV73b  SELEASVASPFTSKTPSISCVPNLIREGRAALITSFCVFKFMALYSIIQYFSVTLLYSIL

NOV73a  SNLGDFQFLFIDLAIILVVVFTVSLNPAWKELVAQRPPSGLISGALLFSVLSQIIICIGF
NOV73b  SNLGDFQFLFIDLAIILVVVFTMSLNPAWKELVAQRPPSGLISGALLFSVLSQIIICIGF

NOV73a  QSLGFFWVKQQPWYEVWHPKSEACNTTGSGFWNSSHVDNETELDEIIYKIMKIPQCFLFP
NOV73b  QSLGFFWVKQQPWYEVWHPKSDACNTTGSGFWNSSHVDNETELDELIYKIMKIPQCFLFP

NOV73a  VFSTS
NOV73b  VFSTS

NOV73a  (SEQ ID NO:  1068)
NOV73b  (SEQ ID NO:  1070)
```

Further analysis of the NOV73a protein yielded the following properties shown in Table 73C.

Table 73C. Protein Sequence Properties NOV73a

SignalP analysis:	Cleavage site between residues 58 and 59

PSORT II analysis:

```
PSG:  a new signal peptide prediction method
      N-region:  length 7;  pos.chg 3;  neg.chg 3
      H-region:  length 6;  peak value -12.14
      PSG score: -16.54

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -4.30
      possible cleavage site: between 41 and 42
```

```
>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   8
      INTEGRAL    Likelihood = -8.44   Transmembrane   31 -  47
      INTEGRAL    Likelihood = -6.37   Transmembrane  234 - 250
      INTEGRAL    Likelihood =  0.32   Transmembrane  295 - 311
      INTEGRAL    Likelihood = -9.08   Transmembrane  410 - 426
      INTEGRAL    Likelihood = -2.81   Transmembrane  439 - 455
      INTEGRAL    Likelihood = -2.92   Transmembrane  926 - 942
      INTEGRAL    Likelihood =-11.15   Transmembrane  963 - 979
      INTEGRAL    Likelihood = -6.37   Transmembrane  996 -1012
      PERIPHERAL  Likelihood =  1.11 (at 211)
      ALOM score: -11.15  (number of TMSs: 8)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 38
      Charge difference:  4.0   C( 2.0) - N(-2.0)
      C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  8.50
      Hyd Moment(95):  5.62    G content:       0
      D/E content:     2       S/T content:     0
      Score: -6.77

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   9.1%
      NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none
```

```
Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       55.6 %: endoplasmic reticulum
       11.1 %: mitochondrial
       11.1 %: vacuolar
       11.1 %: vesicles of secretory system
       11.1 %: Golgi

>> prediction for CG59839-02 is end (k=9)
```

A search of the NOV73a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 73D.

Table 73D. Geneseq Results for NOV73a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV73a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU91185	Human HEAT-2 polypeptide - Homo sapiens, 1256 aa. [WO200216591-A2, 28-FEB-2002]	1..1059 1..1065	1044/1065 (98%) 1053/1065 (98%)	0.0
AAO14197	Human transporter and ion channel TRICH-14 - Homo sapiens, 1256 aa. [WO200204520-A2, 17-JAN-2002]	1..1059 1..1065	1039/1065 (97%) 1049/1065 (97%)	0.0
ABU11063	Human protein NOV27 - Homo sapiens, 973 aa. [WO200281629-A2, 17-OCT-2002]	183..1059 1..812	784/883 (88%) 792/883 (88%)	0.0
ABP69451	Human polypeptide SEQ ID NO 1498 - Homo sapiens, 765 aa. [WO200270539-A2, 12-SEP-2002]	462..1059 1..604	593/604 (98%) 598/604 (98%)	0.0
AAB40996	Human ORFX ORF760 polypeptide sequence SEQ ID NO:1520 - Homo sapiens, 692 aa. [WO200058473-A2, 05-OCT-2000]	509..1059 2..558	547/557 (98%) 551/557 (98%)	0.0

In a BLAST search of public sequence databases, the NOV73a protein was found to have homology to the proteins shown in the BLASTP data in Table 73E.

Table 73E. Public BLASTP Results for NOV73a				
Protein Accession Number	**Protein/Organism/Length**	**NOV73a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9H7F0	Probable cation-transporting ATPase 3 (EC 3.6.3.-) (ATPase family homolog up-regulated in senescence cells 1) - Homo sapiens (Human), 1130 aa.	1..1059 1..1065	1045/1065 (98%) 1054/1065 (98%)	0.0
CAD29017	Sequence 5 from Patent WO0216591 - Homo sapiens (Human), 1256 aa.	1..1059 1..1065	1044/1065 (98%) 1053/1065 (98%)	0.0
Q96KS1	Hypothetical protein - Homo sapiens (Human), 701 aa.	280..954 3..680	655/681 (96%) 662/681 (97%)	0.0
Q95JN5	Hypothetical 56.9 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 504 aa (fragment).	1..492 1..492	492/492 (100%) 492/492 (100%)	0.0
Q95JN5	Probable cation-transporting ATPase 3 (EC 3.6.3.-) (ATPase family homolog up-regulated in senescence cells 1) - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 492 aa (fragment).	1..492 1..492	492/492 (100%) 492/492 (100%)	0.0

PFam analysis indicates that the NOV73a protein contains the domains shown in the Table 73F.

Table 73F. Domain Analysis of NOV73a			
Pfam Domain	**NOV73a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Cation_ATPase_N	160..227	15/87 (17%) 41/87 (47%)	0.041
E1-E2_ATPase	239..488	58/267 (22%) 162/267 (61%)	4.3e-07
Hydrolase	492..898	40/417 (10%) 243/417 (58%)	0.0073

Example 74.

The NOV74 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 74A.

Table 74A. NOV74 Sequence Analysis		
NOV74a, CG90866-04 DNA Sequence	SEQ ID NO: 1071	2937 bp
	ORF Start: ATG at 1	ORF Stop: TAA at 2935

ATGCAGCCCCTGGACTTTAGTTCAGGGGGAAGTGACCCCAACATCAGCCTCTCAGAAAAGATCCGAGA
TCAGCTTGTTGTTGGACAGCTGATTCCAGACTGCTATGTAGAACTTGAAAAAATCATTTTATCGGAGC
GTAAAAATGTGCCAATTGAATTTCCCGTAATTGACCGGAAACGATTATTACAACTAGTGAGAGAAAAT
CAGCTGCAGTTAGATGAAAATGAGCTTCCTCACGCAGTTCACTTTCTAAATGAATCAGGAGTCCTTCT
TCATTTTCAAGACCCAGCACTGCAGTTAAGTGACTTGTACTTTGTGGAACCCAAGTGGCTTTGTAAAA
TCATGGCACAGGATGTTAGCAGCATTTTTGGCCTTTATATTCGAGACATTTTGACAGTGAAAGTGGAA
GGTTGTCCAAAACACCCTAAGGGCATTATTTCGCGTAGAGATGTGGAAAAATTTCTTTCAAAAAAAAG
GAAATTTCCAAAGAACTACATGTCACAGTATTTTAAGCTCCTAGAAAAATTCCAGATTGCTTTGCCAA
TAGGAGAAGAATATTTGCTGGTTCCAAGCAGTTTGTCTGACCACAGGCCTGTGATAGAGCTTCCCCAT
TGTGAGAACTCTGAAATTATCATCCGACTATATGAAATGCCTTATTTTCCAATGGGATTTTGGTCAAG
ATTAATCAATCGATTACTTGAGATTTCACCTTACATGCTTTCAGGGAGAGGCTGTATTCTTTTGGGCC
AAGTTGTGGACCACATTGATTCTCTCATGGAAGAATGGTTTCCTGGGTTGCTGGAGATTGATATTTGT
GGTGAAGGAGAAACTCTGTTGAAGAAATGGGCATTATATAGTTTTAATGATGGTGAAGAACATCAAAA
AATCTTACTTGATGACTTGATGAAGAAAGCAGAGGAAGGAGATCTCTTAGTAAATCCAGATCAACCAA
GGCTCACCATTCCAATATCTCAGATTGCCCCTGACTTGATTTTGGCTGACCCGCCTAGAAATATTATG
TTGAATAATGATGAGTTGGAATTTGAACAAGCTCCAGAGTTTCTCCTAGACTGTTTTGTGTGTATTCA
CTTATATCCATCAAGTGACTACATTTCAAGGCACTATATGAGAACCATAAATATTGTACAAACAGGAT
TTGCTAAATGTCGGTGGAGAGTAACAGTCCACGGGGCTGATCATGGTGATGGCAGTTTTGGATCAGTT
TACCGAGCAGCCTATGAAGGAGAAGAAGTGGCTGTGAAGATTTTTAATAAACATACATCACTCAGGCT
GTTAAGACAAGAGCTTGTGGTGCTTTGCCACCTCCACCACCCCAGTTTGATATCTTTGCTGGCAGCTG
GGATTCGTCCCCGGATGTTGGTGATGGAGTTAGCCTCCAAGGGTTCCTTGGATCGCCTGCTTCAGCAG
GACAAAGCCAGCCTCACTAGAACCCTACAGCACAGGATTGCACTCCACGTAGCTGATGGTTTGAGATA
CCTCCACTCAGCCATGATTATATACCGAGACCTGAAACCCCACAATGTGCTGCTTTTCACACTGTATC
CCAATGCTGCCATCATTGCAAAGATTGCTGACTACGGCATTGCTCAGTACTGCTGTAGAATGGGGATA
AAAACATCAGAGGGCACACCAGGTTTTCGTGCACCTGAAGTTGCCAGAGGAAATGTCATTTATAACCA
ACAGGCTGATGTTTATTCATTTGGTTTACTACTCTATGACATTTTGACAACTGGAGGTAGAATAGTAG
AGGGTTTGAAGTTTCCAAATGAGTTTGATGAATTAGAAATACAAGGAAAATTACCTGATCCAGTTAAA
GAATATGGTTGTGCCCCATGGCCTATGGTTGAGAAATTAATTAAACAGTGTTTGAAAGAAAATCCTCA
AGAAAGGCCTACTTCTGCCCAGGTATTCGACATTTTGAATTCAGCTGAATTAGTCTGTCTGACGAGAC
GCATTTTATTACCTAAAAACGTAATTGTTGAATGCATGGTTGCTACACATCACAACAGCAGGAATGCA
AGCATTTGGCTGGGCTGTGGGCACACCGACAGAGGACAGCTCTCATTTCTTGACTTAAATACTGAAGG
ATACACTTCTGAGAGCAAACAAAAAAATTTTCTTTTGGTTGGAACCGCTGATGGCAAGTTAGCAATTT

TTGAAGATAAGACTGTTAAGCTTAAAGGAGCTGCTCCTTTGAAGATACTAAATATAGGAAATGTCAGT
ACTCCATTGATGTGTTTGAGTGAATCCACAAATTCAACGGAAAGAAATGTAATGTGGGGAGGATGTGG
CACAAAGATTTTCTCCTTTTCTAATGATTTCACCATTCAGAAACTCATTGAGACAAGAACAAGCCAAC
TGTTTTCTTATGCAGCTTTCAGTGATTCCAACATCATAACAGTGGTGGTAGACACTGCTCTCTATATT
GCTAAGCAAAATAGCCCTGTTGTGGAAGTGTGGGATAAGAAAACTGAAAAACTCTGTGGACTAATAGA
CTGCGTGCACTTTTTAAGGGAGGTAATGGTAAAAGAAAACAAGGAATCAAAACACAAAATGTCTTATT
CTGGGAGAGTGAAAACCCTCTGCCTTCAGAAGAACACTGCTCTTTGGATAGGAACTGGAGGAGGCCAT
ATTTTACTCCTGGATCTTTCAACTCGTCGACTTATACGTGTAATTTACAACTTTTGTAATTCGGTCAG
AGTCATGATGACAGCACAGCTAGGCAGCCTTAAAAATGTCATGCTGGTATTGGGCTACAACCGGAAAA
ATACTGAAGGTACACAAAAGCAGAAAGAGATACAATCTTGCTTGACCGTTTGGGACATCAATCTTCCA
CATGAAGTGCAAAATTTAGAAAAACACATTGAAGTGAGAAAAGAATTAGCTGAAAAAATGAGACGAAC
ATCTGTTGAG**TAA**

NOV74a, CG90866-04 Protein Sequence	SEQ ID NO: 1072	978 aa	MW at 111173.8kD

MQPLDFSSGGSDPNISLSEKIRDQLVVGQLIPDCYVELEKIILSERKNVPIEFPVIDRKRLLQLVREN
QLQLDENELPHAVHFLNESGVLLHFQDPALQLSDLYFVEPKWLCKIMAQDVSSIFGLYIRDILTVKVE
GCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIALPIGEEYLLVPSSLSDHRPVIELPH
CENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYMLSGRGCILLGQVVDHIDSLMEEWFPGLLEIDIC
GEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDLILADPPRNIM
LNNDELEFEQAPEFLLDCFVCIHLYPSSDYISRHYMRTINIVQTGFAKCRWRVTVHGADHGDGSFGSV
YRAAYEGEEVAVKIFNKHTSLRLLRQELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQ
DKASLTRTLQHRIALHVADGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYGIAQYCCRMGI
KTSEGTPGFRAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPVK
EYGCAPWPMVEKLIKQCLKENPQERPTSAQVFDILNSAELVCLTRRILLPKNVIVECMVATHHNSRNA
SIWLGCGHTDRGQLSFLDLNTEGYTSESKQKNFLLVGTADGKLAIFEDKTVKLKGAAPLKILNIGNVS
TPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTIQKLIETRTSQLFSYAAFSDSNIITVVVDTALYI
AKQNSPVVEVWDKKTEKLCGLIDCVHFLREVMVKENKESKHKMSYSGRVKTLCLQKNTALWIGTGGGH
ILLLDLSTRRLIRVIYNFCNSVRVMMTAQLGSLKNVMLVLGYNRKNTEGTQKQKEIQSCLTVWDINLP
HEVQNLEKHIEVRKELAEKMRRTSVE

NOV74b, CG90866-03 DNA Sequence	SEQ ID NO: 1073	2687 bp
	ORF Start: ATG at 108	ORF Stop: TAA at 2658

ACGCAGTTCACTTTCTAAATGAATCAGGAGTCCTTCTTCATTTTCAAGACCCAGCACTGCAGTTAAGT
GACTTGTACTTTGTGGAACCCAAGTGGCTTTGTAAAATC**ATG**GCACAGATTTTGACAGTGAAAGTGGA
AGGTTGTCCAAAACACCCTAAGGGCATTATTTCGCGTAGAGATGTGGAAAAATTTCTTTCAAAAAAAA
GGAAATTTCCAAAGAACTACATGTCACAGTATTTTAAGCTCCTAGAAAAATTCCAGATTGCTTTGCCA
ATAGGAGAAGAATATTTGCTGGTTCCAAGCAGTTTGTCTGACCACAGGCCTGTGATAGAGCTTCCCCA
TTGTGAGAACTCTGAAATTATCATCCGACTATATGAAATGCCTTATTTTCCAATGGGATTTTGGTCAA
GATTAATCAATCGATTACTTGAGATTTCACCTTACATGCTTTCAGGGAGAGAACGAGCACTTCGCCCA
AACAGAATGTATTGGCGACAAGGCATTTACTTAAATTGGTCTCCTGAAGCTTATTGTCTGGTAGGATC
TGAAGTCTTAGACAATCATCCAGAGAGTTTCTTAAAAATTACAGTTCCTTCTTGTAGAAAAGGCTGTA
TTCTTTTGGGCCAAGTTGTGGACCACATTGATTCTCTCATGGAAGAATGGTTTCCTGGGTTGCTGGAG
ATTGATATTTGTGGTGAAGGAGAAACTCTGTTGAAGAAATGGGCATTATATAGTTTTAATGATGGCGA
AGAACATCAAAAAATCTTACTTGATGACTTGATGAAGAAAGCAGAGGAAGGAGATCTCTTAGTAAATC
CAGATCAACCAAGGCTCACCATTCCAATATCTCAGATTGCCCCTGACTTGATTTTGGCTGACCTGCCT
AGAAATATTATGTTGAATAATGATGAGTTGGAATTTGAACAAGCTCCAGAGTTTCTCCTAGGTGATGG
CAGTTTTGGATCAGTTTACCGAGCAGCCTATGAAGGAGAAGAAGTGGCTGTGAAGATTTTTAATAAAC
ATACATCACTCAGGCTGTTAAGACAAGAGCTTGTGGTGCTTTGCCACCTCCACCACCCCAGTTTGATA
TCTTTGCTGGCAGCTGGGATTCGTCCCCGGATGTTGGTGATGGAGTTAGCCTCCAAGGGTTCCTTGGA
TCGCCTGCTTCAGCAGGACAAAGCCAGCCTCACTAGAACCCTACAGCACAGGATTGCACTCCACGTAG
CTGATGGTTTGAGATACCTCCACTCAGCCATGATTATATACCGAGACCTGAAACCCCACAATGTGCTG
CTTTTCACACTGTATCCCAATGCTGCCATCATTGCAAAGATTGCTGACTACGGCATTGCTCAGTACTG
CTGTAGAATGGGGATAAAAACATCAGAGGGCACACCAGGTTTTCGTGCACCTGAAGTTGCCAGAGGAA
ATGTCATTTATAACCAACAGGCTGATGTTTATTCATTTGGTTTACTACTCTATGACATTTTGACAACT
GGAGGTAGAATAGTAGAGGGTTTGAAGTTTCCAAATGAGTTTGATGAATTAGAAATACAAGGAAAATT
ACCTGATCCAGTTAAAGAATATGGTTGTGCCCCATGGCCTATGGTTGAGAAATTAATTAAACAGTGTT
TGAAAGAAAATCCTCAAGAAAGGCCTACTTCTGCCCAGGTATTCGACATTTTGAATTCAGCTGAATTA
GTCTGTCTGACGAGACGCATTTTATTACCTAAAAACGTAATTGTTGAATGCATGGTTGCTACACATCA
CAACAGCAGGAATGCAAGCATTTGGCTGGGCTGTGGGCACACCGACAGAGGACAGCTCTCATTTCTTG
ACTTAAATACTGAAGGATACACTTCTGAGGAAGTTGCTGATAGTAGAATATTGTGCTTAGCCTTGGTG
CATCTTCCTGTTGAAAAGGAAAGCTGGATTGTGTCTGGGACACAGTCTGGTACTCTCCTGGTCATCAA

TACCGAAGATGGGAAAAAGAGACATACCCTAGAAAAGATGACTGATTCTGTCACTTGTTTGTATTGCA
ATTCCTTTTCCAAGCAAAGCAAACAAAAAAATTTTCTTTTGGTTGGAACCGCTGATGGCAAGTTAGCA
ATTTTTGAAGATAAGACTGTTAAGCTTAAAGGAGCTGCTCCTTTGAAGATACTAAATATAGGAAATGT
CAGTACTCCATTGATGTGTTTGAGTGAATCCACAAATTCAACGGAAAGAAATGTAATGTGGGGAGGAT
GTGGCACAAAGATTTTCTCCTTTTCTAATGATTTCACCATTCAGAAACTCATTGAGACAAGAACAAGC
CAACTGTTTTCTTATGCAGCTTTCAGTGATTCCAACATCATAACAGTGGTGGTAGACACTGCTCTCTA
TATTGCTAAGCAAAATAGCCCTGTTGTGGAAGTGTGGGATAAGAAAACTGAAAAACTCTGTGGACTAA
TAGACTGCGTGCACTTTTTAAGCCTTAAAAATGTCATGCTGGTATTGGGCTACAACCGGAAAAATACT
GAAGGTACACAAAAGCAGAAAGAGATACAATCTTGCTTGACCGTTTGGGACATCAATCTTCCACATGA
AGTGCAAAATTTAGAAAAACACATTGAAGTGAGAAAAGAATTAGCTGAAAAAATGAGACGAACATCTG
TTGAGTAAGAGAGAAATAGGAATTGTCTTTGGATA

NOV74b, CG90866-03 Protein Sequence	SEQ ID NO: 1074	850 aa	MW at 96332.5kD

MAQILTVKVEGCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIALPIGEEYLLVPSSLS
DHRPVIELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYMLSGRERALRPNRMYWRQGIYLNW
SPEAYCLVGSEVLDNHPESFLKITVPSCRKGCILLGQVVDHIDSLMEEWFPGLLEIDICGEGETLLKK
WALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDLILADLPRNIMLNNDELEFE
QAPEFLLGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRLLRQELVVLCHLHHPSLISLLAAGIRPRMLV
MELASKGSLDRLLQQDKASLTRTLQHRIALHVADGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAK
IADYGIAQYCCRMGIKTSEGTPGFRAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNE
FDELEIQGKLPDPVKEYGCAPWPMVEKLIKQCLKENPQERPTSAQVFDILNSAELVCLTRRILLPKNV
IVECMVATHHNSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKESWIVSG
TQSGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKLAIFEDKTVKLKGAA
PLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTIQKLIETRTSQLFSYAAFSDSNI
ITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHFLSLKNVMLVLGYNRKNTEGTQKQKEIQSCL
TVWDINLPHEVQNLEKHIEVRKELAEKMRRTSVE

NOV74c, CG90866-01 DNA Sequence	SEQ ID NO: 1075	3052 bp
	ORF Start: ATG at 108	ORF Stop: TAA at 2853

ACGCAGTTCACTTTCTAAATGAATCAGGAGTCCTTCTTCATTTTCAAGACCCAGCACTGCAGTTAAGT
GACTTGTACTTTGTGGAACCCAAGTGGCTTTGTAAAATCATGGCACAGATTTTGACAGTGAAAGTGGA
AGGTTGTCCAAAACACCCTAAGGGCATTATTTCGCGTAGAGATGTGGAAAAATTTCTTTCAAAAAAAA
GGAAATTTCCAAAGAACTACATGTCACAGTATTTTAAGCTCCTAGAAAAATTCCAGATTGCTTTGCCA
ATAGGAGAAGAATATTTGCTGGTTCCAAGCAGTTTGTCTGACCACAGGCCTGTGATAGAGCTTCCCCA
TTGTGAGAACTCTGAAATTATCATCCGACTATATGAAATGCCTTATTTTCCAATGGGATTTTGGTCAA
GATTAATCAATCGATTACTTGAGATTTCACCTTACATGCTTTCAGGGAGAGAACGAGCACTTCGCCCA
AACAGAATGTATTGGCGACAAGGCATTTACTTAAATTGGTCTCCTGAAGCTTATTGTCTGGTAGGATC
TGAAGTCTTAGACAATCATCCAGAGAGTTTCTTAAAAATTACAGTTCCTTCTTGTAGAAAAGGCTGTA
TTCTTTTGGGCCAAGTTGTGGACCACATTGATTCTCTCATGGAAGAATGGTTTCCTGGGTTGCTGGAG
ATTGATATTTGTGGTGAAGGAGAAACTCTGTTGAAGAAATGGGCATTATATAGTTTTAATGATGGCGA
AGAACATCAAAAAATCTTACTTGATGACTTGATGAAGAAAGCAGAGGAAGGAGATCTCTTAGTAAATC
CAGATCAACCAAGGCTCACCATTCCAATATCTCAGATTGCCCCTGACTTGATTTTGGCTGACCTGCCT
AGAAATATTATGTTGAATAATGATGAGTTGGAATTTGAACAAGCTCCAGAGTTTCTCCTAGGTGATGG
CAGTTTTGGATCAGTTTACCGAGCAGCCTATGAAGGAGAAGAAGTGGCTGTGAAGATTTTTAATAAAC
ATACATCACTCAGGCTGTTAAGACAAGAGCTTGTGGTGCTTTGCCACCTCCACCACCCCAGTTTGATA
TCTTTGCTGGCAGCTGGGATTCGTCCCCGGATGTTGGTGATGGAGTTAGCCTCCAAGGGTTCCTTGGA
TCGCCTGCTTCAGCAGGACAAAGCCAGCCTCACTAGAACCCTACAGCACAGGATTGCACTCCACGTAG
CTGATGGTTTGAGATACCTCCACTCAGCCATGATTATATACCGAGACCTGAAACCCCACAATGTGCTG
CTTTTCACACTGTATCCCAATGCTGCCATCATTGCAAAGATTGCTGACTACGGCATTGCTCAGTACTG
CTGTAGAATGGGGATAAAAACATCAGAGGGCACACCAGGGTTTCGTGCACCTGAAGTTGCCAGAGGAA
ATGTCATTTATAACCAACAGGCTGATGTTTATTCATTTGGTTTACTACTCTATGACATTTTGACAACT
GGAGGTAGAATAGTAGAGGGTTTGAAGTTTCCAAATGAGTTTGATGAATTAGAAATACAAGGAAAATT
ACCTGATCCAGTTAAAGAATATGGTTGTGCCCCATGGCCTATGGTTGAGAAATTAATTAAACAGTGTT
TGAAAGAAAATCCTCAAGAAAGGCCTACTTCTGCCCAGGTATTCTCTCAGGTCTTTGACATTTTGAAT
TCAGCTGAATTAGTCTGTCTGACGAGACGCATTTTATTACCTAAAAACGTAATTGTTGAATGCATGGT
TGCTACACATCACAACAGCAGGAATGCAAGCATTTGGCTGGGCTGTGGGCACACCGACAGAGGACAGC
TCTCATTTCTTGACTTAAATACTGAAGGATACACTTCTGAGGAAGTTGCTGATAGTAGAATATTGTGC
TTAGCCTTGGTGCATCTTCCTGTTGAAAAGGAAAGCTGGATTGTGTCTGGGACACAGTCTGGTACTCT
CCTGGTCATCAATACCGAAGATGGGAAAAAGAGACATACCCTAGAAAAGATGACTGATTCTGTCACTT
GTTTGTATTGCAATTCCTTTTCCAAGCAAAGCAAACAAAAAAATTTTCTTTTGGTTGGAACCGCTGAT
GGCAAGTTAGCAATTTTTGAAGATAAGACTGTTAAGCTTAAAGGAGCTGCTCCTTTGAAGATACTAAA

TATAGGAAATGTCAGTACTCCATTGATGTGTTTGAGTGAATCCACAAATTCAACGGAAAGAAATGTAA
TGTGGGGAGGATGTGGCACAAAGATTTTCTCCTTTTCTAATGATTTCACCATTCAGAAACTCATTGAG
ACAAGAACAAGCCAACTGTTTTCTTATGCAGCTTTCAGTGATTCCAACATCATAACAGTGGTGGTAGA
CACTGCTCTCTATATTGCTAAGCAAAATAGCCCTGTTGTGGAAGTGTGGGATAAGAAAACTGAAAAAC
TCTGTGGACTAATAGACTGCGTGCACTTTTTAAGGTTAGTAAAACCAAATAGAAAAAAATTATCTAAC
CTTATGATGTCTTTGGCTTTACATCCTATATGTTTAAAATCAAAGTTAAGATGCAGTTCATCCAAAGG
AAGATCCCATATTTTGCTTCGTGTAATTTACAACTTTTGTAATTCGGTCAGAGTCATGATGACAGCAC
AGCTAGGCGGAAGCCTTAAAAATGTCATGCTGGTATTGGGCTACAACCGGAAAAATACTGAAGGTACA
CAAAAGCAGAAAGAGATACAATCTTGCTTGACCGTTTGGGACATCAATCTTCCACATGAAGTGCAAAA
TTTAGAAAAACACATTGAAGTGAGAAAAGAATTAGCTGAAAAAATGAGACGAACATCTGTTGAGTAAG
AGAGAAATAGGAATTGTCTTTGGATAGGAAAATTATTCTCTCCTCTTGTAAATATTTATTTTAAAAAT
GTTCACATGGAAAGGGTACTCACATTTTTTGAAATAGCTCGTGTGTATGAAGGAATGTTATTATTTTT
AATTTAAATATATGTAAAAATACTTACCAGTAAATGTGTATTTTAAAGAACTATTTAAAA

NOV74c, CG90866-01 Protein Sequence	SEQ ID NO: 1076	915 aa	MW at 103676.4kD

MAQILTVKVEGCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIALPIGEEYLLVPSSLS
DHRPVIELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYMLSGRERALRPNRMYWRQGIYLNW
SPEAYCLVGSEVLDNHPESFLKITVPSCRKGCILLGQVVDHIDSLMEEWFPGLLEIDICGEGETLLKK
WALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDLILADLPRNIMLNNDELEFE
QAPEFLLGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRLLRQELVVLCHLHHPSLISLLAAGIRPRMLV
MELASKGSLDRLLQQDKASLTRTLQHRIALHVADGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAK
IADYGIAQYCCRMGIKTSEGTPGFRAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNE
FDELEIQGKLPDPVKEYGCAPWPMVEKLIKQCLKENPQERPTSAQVFSQVFDILNSAELVCLTRRILL
PKNVIVECMVATHHNSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKESW
IVSGTQSGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKLAIFEDKTVKL
KGAAPLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTIQKLIETRTSQLFSYAAFS
DSNIITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHFLRLVKPNRKKLSNLMMSLALHPICLK
SKLRCSSSKGRSHILLRVIYNFCNSVRVMMTAQLGGSLKNVMLVLGYNRKNTEGTQKQKEIQSCLTVW
DINLPHEVQNLEKHIEVRKELAEKMRRTSVE

NOV74d, CG90866-02 DNA Sequence	SEQ ID NO: 1077	3040 bp
	ORF Start: ATG at 108	ORF Stop: TAA at 2841

ACGCAGTTCACTTTCTAAATGAATCAGGAGTCCTTCTTCATTTTCAAGACCCAGCACTGCAGTTAAGT
GACTTGTACTTTGTGGAACCCAAGTGGCTTTGTAAAATC**ATG**GCACAGATTTTGACAGTGAAAGTGGA
AGGTTGTCCAAAACACCCTAAGGGCATTATTTCGCGTAGAGATGTGGAAAAATTTCTTTCAAAAAAAA
GGAAATTTCCAAAGAACTACATGTCACAGTATTTTAAGCTCCTAGAAAAATTCCAGATTGCTTTGCCA
ATAGGAGAAGAATATTTGCTGGTTCCAAGCAGTTTGTCTGACCACAGGCCTGTGATAGAGCTTCCCCA
TTGTGAGAACTCTGAAATTATCATCCGACTATATGAAATGCCTTATTTTCCAATGGGATTTTGGTCAA
GATTAATCAATCGATTACTTGAGATTTCACCTTACATGCTTTCAGGGAGAGAACGAGCACTTCGCCCA
AACAGAATGTATTGGCGACAAGGCATTTACTTAAATTGGTCTCCTGAAGCTTATTGTCTGGTAGGATC
TGAAGTCTTAGACAATCATCCAGAGAGTTTCTTAAAAATTACAGTTCCTTCTTGTAGAAAAGGCTGTA
TTCTTTTGGGCCAAGTTGTGGACCACATTGATTCTCTCATGGAAGAATGGTTTCCTGGGTTGCTGGAG
ATTGATATTTGTGGTGAAGGAGAAACTCTGTTGAAGAAATGGGCATTATATAGTTTTAATGATGGCGA
AGAACATCAAAAAATCTTACTTGATGACTTGATGAAGAAAGCAGAGGAAGGAGATCTCTTAGTAAATC
CAGATCAACCAAGGCTCACCATTCCAATATCTCAGATTGCCCCTGACTTGATTTTGGCTGACCTGCCT
AGAAATATTATGTTGAATAATGATGAGTTGGAATTTGAACAAGCTCCAGAGTTTCTCCTAGGTGATGG
CAGTTTTGGATCAGTTTACCGAGCAGCCTATGAAGGAGAAGAAGTGGCTGTGAAGATTTTTAATAAAC
ATACATCACTCAGGCTGTTAAGACAAGAGCTTGTGGTGCTTTGCCACCTCCACCACCCCAGTTTGATA
TCTTTGCTGGCAGCTGGGATTCGTCCCCGGATGTTGGTGATGGAGTTAGCCTCCAAGGGTTCCTTGGA
TCGCCTGCTTCAGCAGGACAAAGCCAGCCTCACTAGAACCCTACAGCACAGGATTGCACTCCACGTAG
CTGATGGTTTGAGATACCTCCACTCAGCCATGATTATATACCGAGACCTGAAACCCCACAATGTGCTG
CTTTTCACACTGTATCCCAATGCTGCCATCATTGCAAAGATTGCTGACTACGGCATTGCTCAGTACTG
CTGTAGAATGGGGATAAAAACATCAGAGGGCACACCAGGGTTTCGTGCACCTGAAGTTGCCAGAGGAA
ATGTCATTTATAACCAACAGGCTGATGTTTATTCATTTGGTTTACTACTCTATGACATTTTGACAACT
GGAGGTAGAATAGTAGAGGGTTTGAAGTTTCCAAATGAGTTTGATGAATTAGAAATACAAGGAAAATT
ACCTGATCCAGTTAAAGAATATGGTTGTGCCCCATGGCCTATGGTTGAAAAATTAATTAAACAGTGTT
TGAAAGAAAATCCTCAAGAAAGGCCTACTTCTGCCCAGGTCTTTGACATTTTGAATTCAGCTGAATTA
GTCTGTCTGACGAGACGCATTTTATTACCTAAAAACGTAATTGTTGAATGCATGGTTGCTACACATCA
CAACAGCAGGAATGCAAGCATTTGGCTGGGCTGTGGGCACACCGACAGAGGACAGCTCTCATTTCTTG
ACTTAAATACTGAAGGATACACTTCTGAGGAAGTTGCTGATAGTAGAATATTGTGCTTAGCCTTGGTG
CATCTTCCTGTTGAAAAGGAAAGCTGGATTGTGTCTGGGACACAGTCTGGTACTCTCCTGGTCATCAA

TACCGAAGATGGGAAAAAGAGACATACCCTAGAAAAGATGACTGATTCTGTCACTTGTTTGTATTGCA
ATTCCTTTTCCAAGCAAAGCAAACAAAAAAATTTTCTTTTGGTTGGAACCGCTGATGGCAAGTTAGCA
ATTTTTGAAGATAAGACTGTTAAGCTTAAAGGAGCTGCTCCTTTGAAGATACTAAATATAGGAAATGT
CAGTACTCCATTGATGTGTTTGAGTGAATCCACAAATTCAACGGAAAGAAATGTAATGTGGGGAGGAT
GTGGCACAAAGATTTTCTCCTTTTCTAATGATTTCACCATTCAGAAACTCATTGAGACAAGAACAAGC
CAACTGTTTTCTTATGCAGCTTTCAGTGATTCCAACATCATAACAGTGGTGGTAGACACTGCTCTCTA
TATTGCTAAGCAAAATAGCCCTGTTGTGGAAGTGTGGGATAAGAAAACTGAAAAACTCTGTGGACTAA
TAGACTGCGTGCACTTTTTAAGGTTAGTAAAACCAAATAGAAAAAAATTATCTAACCTTATGATGTCT
TTGGCTTTACATCCTATATGTTTAAAATCAAAGTTAAGATGCAGTTCATCCAAAGGAAGATCCCATAT
TTTGCTTCGTGTAATTTACAACTTTTGTAATTCGGTCAGAGTCATGATGACAGCACAGCTAGGCGGAA
GCCTTAAAAATGTCATGCTGGTATTGGGCTACAACCGGAAAAATACTGAAGGTACACAAAAGCAGAAA
GAGATACAATCTTGCTTGACCGTTTGGGACATCAATCTTCCACATGAAGTGCAAAATTTAGAAAAACA
CATTGAAGTGAGAAAAGAATTAGCTGAAAAAATGAGACGAACATCTGTTGAGTAAGAGAGAAATAGGA
ATTGTCTTTGGATAGGAAAATTATTCTCTCCTCTTGTAAATATTTATTTTAAAAATGTTCACATGGAA
AGGGTACTCACATTTTTTGAAATAGCTCGTGTGTATGAAGGAATGTTATTATTTTTAATTTAAATATA
TGTAAAAATACTTACCAGTAAATGTGTATTTTAAAGAACTATTTAAAA

NOV74d, CG90866-02 Protein Sequence	SEQ ID NO: 1078	911 aa	MW at 103214.9kD

MAQILTVKVEGCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIALPIGEEYLLVPSSLS
DHRPVIELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYMLSGRERALRPNRMYWRQGIYLNW
SPEAYCLVGSEVLDNHPESFLKITVPSCRKGCILLGQVVDHIDSLMEEWFPGLLEIDICGEGETLLKK
WALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDLILADLPRNIMLNNDELEFE
QAPEFLLGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRLLRQELVVLCHLHHPSLISLLAAGIRPRMLV
MELASKGSLDRLLQQDKASLTRTLQHRIALHVADGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAK
IADYGIAQYCCRMGIKTSEGTPGFRAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNE
FDELEIQGKLPDPVKEYGCAPWPMVEKLIKQCLKENPQERPTSAQVFDILNSAELVCLTRRILLPKNV
IVECMVATHHNSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKESWIVSG
TQSGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKLAIFEDKTVKLKGAA
PLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTIQKLIETRTSQLFSYAAFSDSNI
ITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHFLRLVKPNRKKLSNLMMSLALHPICLKSKLR
CSSSKGRSHILLRVIYNFCNSVRVMMTAQLGGSLKNVMLVLGYNRKNTEGTQKQKEIQSCLTVWDINL
PHEVQNLEKHIEVRKELAEKMRRTSVE

NOV74e, CG90866-05 DNA Sequence	SEQ ID NO: 1079	2955 bp
	ORF Start: ATG at 81	ORF Stop: TAA at 2844

GAGTCCTTCTTCATTTTCAAGACCCAGCACTGCAGTTAAGTGACTTGTACTTTGTGGAACCCAAGTGG
CTTTGTAAAATCATGGCACAGATTTTGACAGTGAAAGTGGAAGGTTGTCCAAAACACCCTAAGGGAAT
TATTTCGCGTAGAGATGTGGAAAAATTTCTTTCAAAAAAAAGGAAATTTCCAAAGAACTACATGTCAC
AGTATTTTAAGCTCCTAGAAAAATTCCAGATTGCTTTGCCAATAGGAGAAGAATATTTGCTGGTTCCA
AGCAGTTTGTCTGACCACAGGCCTGTGATAGAGCTTCCCCATTGTGAGAACTCTGAAATTATCATCCG
ACTATATGAAATGCCTTATTTTCCAATGGGATTTTGGTCAAGATTAATCAATCGATTACTTGAGATTT
CACCTTACATGCTTTCAGGGAGAGAACGAGCACTTCGCCCAAACAGAATGTATTGGCGACAAGGCATT
TACTTAAATTGGTCTCCTGAAGCTTATTGTCTGGTAGGATCTGAAGTCTTAGACAATCATCCAGAGAG
TTTCTTAAAAATTACAGTTCCTTCTTGTAGAAAAGGCTGTATTCTTTTGGGCCAAGTTGTGGACCACA
TTGATTCTCTCATGGAAGAATGGTTTCCTGGGTTGCTGGAGATTGATATTTGTGGTGAAGGAGAAACT
CTGTTGAAGAAATGGGCATTATATAGTTTTAATGATGGTGAAGAACATCAAAAAATCTTACTTGATGA
CTTGATGAAGAAAGCAGAGGAAGGAGATCTCTTAGTAAATCCAGATCAACCAAGGCTCACCATTCCAA
TATCTCAGATTGCCCCTGACTTGATTTTGGCTGACCTGCCTAGAAATATTATGTTGAATAATGATGAG
TTGGAATTTGAACAAGCTCCAGAGTTTCTCCTAGGTGATGGCAGTTTTGGATCAGTTTACCGAGCAGC
CTATGAAGGAGAAGAAGTGGCTGTGAAGATTTTTAATAAACATACATCACTCAGGCTGTTAAGACAAG
AGCTTGTGGTGCTTTGCCACCTCCACCACCCCAGTTTGATATCTTTGCTGGCAGCTGGGATTCGTCCC
CGGATGTTGGTGATGGAGTTAGCCTCCAAGGGTTCCTTGGATCGCCTGCTTCAGCAGGACAAAGCCAG
CCTCACTAGAACCCTACAGCACAGGATTGCACTCCACGTAGCTGATGGTTTGAGATACCTCCACTCAG
CCATGATTATATACCGAGACCTGAAACCCCACAATGTGCGGCTTTTCACACTGTATCCCAATGCTGCC
ATCATTGCAAAGATTGCTGACTACGGCATTGCTCAGTACTGCTGTAGAATGGGGATAAAAACATCAGA
GGGCACACCAGGGTTTCGTGCACCTGAAGTTGCCAGAGGAAATGTCATTTATAACCAACAGGCTGATG
TTTATTCATTTGGTTTACTACTCTATGACATTTTGACAACTGGAGGTAGAATAGTAGAGGGTTTGAAG
TTTCCAAATGAGTTTGATGAATTAGAAATACAAGGAAAATTACCTGATCCAGTTAAAGAATATGGTTG
TGCCCCATGGCCTATGGTTGAGAAATTAATTAAACAGTGTTTGAAAGAAAATCCTCAAGAAAGGCCTA
CTTCTGCCCAGGTCTTTGACATTTTGAATTCAGCTGAATTAGTCTGTCTGACGAGACGCATTTTATTA
CCTAAAAACGTAATTGTTGAATGCATGGTTGCTACACATCACAACAGCAGGAATGCAAGCATTTGGCT

GGGCTGTGGGCACACCGACAGAGGACAGCTCTCATTTCTTGACTTAAATACTGAAGGATACACTTCTG
AGGAAGTTGCTGATAGTAGAATATTGTGCTTAGCCTTGGTGCATCTTCCTGTTGAAAAGGAAAGCTGG
ATTGTGTCTGGGACACAGTCTGGTACTCTCCTGGTCATCAATACCGAAGATGGGAAAAAGAGACATAC
CCTAGAAAAGATGACTGATTCTGTCACTTGTTTGTATTGCAATTCCTTTTCCAAGCAAAGCAAACAAA
AAAATTTTCTTTTGGTTGGAACCGCTGATGGCAAGTTAGCAATTTTTGAAGATAAGACTGTTAAGCTT
AAAGGAGCTGCTCCTTTGAAGATACTAAATATAGGAAATGTCAGTACTCCATTGATGTGTTTGAGTGA
ATCCACAAATTCAACGGAAAGAAATGTAATGTGGGGAGGATGTGGCACAAAGATTTTCTCCTTTTCTA
ATGATTTCACCATTCAGAAACTCATTGAGACAAGAACAAGCCAACTGTTTTCTTATGCAGCTTTCAGT
GATTCCAACATCATAACAGTGGTGGTAGACACTGCTCTCTATATTGCTAAGCAAAATAGCCCTGTTGT
GGAAGTGTGGGATAAGAAAACTGAAAAACTCTGTGGGCTAATAGACTGCGTGCACTTTTTAAGGGAGG
TAACGGTAAAAGAAAACAAGGAATCAAAACACAAAATGTCTTATTCTGGGAGAGTGAAAACCCTCTGC
CTTCAGAAGAACACTGCTCTTTGGATAGGAACTGGAGGAGGCCATATTTTACTCCTGGATCTTTCAAC
TCGTCGACTTATACGTGTAATTTACAACTTTTGTAATTCGGTCAGAGTCATGATGACAGCACAGCTAG
GAAGCCTTAAAAATGTCATGCTGGTATTGGGCTACAACCGGAAAAATACTGAAGGTACACAAAAGCAG
AAAGAGATACAATCTTGCTTGACCGTTTGGGACATCAATCTTCCACATGAAGTGCAAAATTTAGAAAA
ACACATTGAAGTGAGAAAAGAATTAGCTGAAAAAATGAGACGAACATCTGTTGAGTAAGAGAGAAATA
GGAATTGTCTTTGGATAGGAAAATTATTCTCTCCTCTTGTAAATATTTATTTTAAAAATGTTCACATG
GAAAGGGTACTCACATTTTTAAGGGCGAATC

NOV74e, CG90866-05 Protein Sequence	SEQ ID NO: 1080	921 aa	MW at 104423.0kD

MAQILTVKVEGCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIALPIGEEYLLVPSSLS
DHRPVIELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYMLSGRERALRPNRMYWRQGIYLNW
SPEAYCLVGSEVLDNHPESFLKITVPSCRKGCILLGQVVDHIDSLMEEWFPGLLEIDICGEGETLLKK
WALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDLILADLPRNIMLNNDELEFE
QAPEFLLGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRLLRQELVVLCHLHHPSLISLLAAGIRPRMLV
MELASKGSLDRLLQQDKASLTRTLQHRIALHVADGLRYLHSAMIIYRDLKPHNVRLFTLYPNAAIIAK
IADYGIAQYCCRMGIKTSEGTPGFRAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNE
FDELEIQGKLPDPVKEYGCAPWPMVEKLIKQCLKENPQERPTSAQVFDILNSAELVCLTRRILLPKNV
IVECMVATHHNSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKESWIVSG
TQSGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKLAIFEDKTVKLKGAA
PLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTIQKLIETRTSQLFSYAAFSDSNI
ITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHFLREVTVKENKESKHKMSYSGRVKTLCLQKN
TALWIGTGGGHILLLDLSTRRLIRVIYNFCNSVRVMMTAQLGSLKNVMLVLGYNRKNTEGTQKQKEIQ
SCLTVWDINLPHEVQNLEKHIEVRKELAEKMRRTSVE

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 74B.

Table 74B. Comparison of the NOV74 protein sequences.

```
NOV74a  MQPLDFSSGGSDPNISLSEKIRDQLVVGQLIPDCYVELEKIILSERKNVPIEFPVIDRKR
NOV74b  ------------------------------------------------------------
NOV74c  ------------------------------------------------------------
NOV74d  ------------------------------------------------------------
NOV74e  ------------------------------------------------------------

NOV74a  LLQLVRENQLQLDENELPHAVHFLNESGVLLHFQDPALQLSDLYFVEPKWLCKIMAQDVS
NOV74b  ----------------------------------------------------MAQ---
NOV74c  ----------------------------------------------------MAQ---
NOV74d  ----------------------------------------------------MAQ---
NOV74e  ----------------------------------------------------MAQ---

NOV74a  SIFGLYIRDILTVKVEGCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIAL
NOV74b  ---------ILTVKVEGCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIAL
NOV74c  ---------ILTVKVEGCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIAL
NOV74d  ---------ILTVKVEGCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIAL
NOV74e  ---------ILTVKVEGCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIAL
```

```
NOV74a  PIGEEYLLVPSSLSDHRPVIELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYML
NOV74b  PIGEEYLLVPSSLSDHRPVIELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYML
NOV74c  PIGEEYLLVPSSLSDHRPVIELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYML
NOV74d  PIGEEYLLVPSSLSDHRPVIELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYML
NOV74e  PIGEEYLLVPSSLSDHRPVIELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYML

NOV74a  SG--R----------------------------------------------GCILLGQV
NOV74b  SGRERALRPNRMYWRQGIYLNWSPEAYCLVGSEVLDNHPESFLKITVPSCRKGCILLGQV
NOV74c  SGRERALRPNRMYWRQGIYLNWSPEAYCLVGSEVLDNHPESFLKITVPSCRKGCILLGQV
NOV74d  SGRERALRPNRMYWRQGIYLNWSPEAYCLVGSEVLDNHPESFLKITVPSCRKGCILLGQV
NOV74e  SGRERALRPNRMYWRQGIYLNWSPEAYCLVGSEVLDNHPESFLKITVPSCRKGCILLGQV

NOV74a  VDHIDSLMEEWFPGLLEIDICGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLL
NOV74b  VDHIDSLMEEWFPGLLEIDICGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLL
NOV74c  VDHIDSLMEEWFPGLLEIDICGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLL
NOV74d  VDHIDSLMEEWFPGLLEIDICGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLL
NOV74e  VDHIDSLMEEWFPGLLEIDICGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLL

NOV74a  VNPDQPRLTIPISQIAPDLILADPPRNIMLNNDELEFEQAPEFLLDCFVCIHLYPSSDYI
NOV74b  VNPDQPRLTIPISQIAPDLILADLPRNIMLNNDELEFEQAPEFLL---------------
NOV74c  VNPDQPRLTIPISQIAPDLILADLPRNIMLNNDELEFEQAPEFLL---------------
NOV74d  VNPDQPRLTIPISQIAPDLILADLPRNIMLNNDELEFEQAPEFLL---------------
NOV74e  VNPDQPRLTIPISQIAPDLILADLPRNIMLNNDELEFEQAPEFLL---------------

NOV74a  SRHYMRTINIVQTGFAKCRWRVTVHGADHGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRL
NOV74b  ----------------------------GDGSFGSVYRAAYEGEEVAVKIFNKHTSLRL
NOV74c  ----------------------------GDGSFGSVYRAAYEGEEVAVKIFNKHTSLRL
NOV74d  ----------------------------GDGSFGSVYRAAYEGEEVAVKIFNKHTSLRL
NOV74e  ----------------------------GDGSFGSVYRAAYEGEEVAVKIFNKHTSLRL

NOV74a  LRQELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIAL
NOV74b  LRQELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIAL
NOV74c  LRQELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIAL
NOV74d  LRQELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIAL
NOV74e  LRQELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIAL

NOV74a  HVADGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYGIAQYCCRMGIKTSEGTP
NOV74b  HVADGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYGIAQYCCRMGIKTSEGTP
NOV74c  HVADGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYGIAQYCCRMGIKTSEGTP
NOV74d  HVADGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYGIAQYCCRMGIKTSEGTP
NOV74e  HVADGLRYLHSAMIIYRDLKPHNVRLFTLYPNAAIIAKIADYGIAQYCCRMGIKTSEGTP

NOV74a  GFRAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPV
NOV74b  GFRAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPV
NOV74c  GFRAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPV
NOV74d  GFRAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPV
NOV74e  GFRAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPV

NOV74a  KEYGCAPWPMVEKLIKQCLKENPQERPTSAQVF----DILNSAELVCLTRRILLPKNVIV
NOV74b  KEYGCAPWPMVEKLIKQCLKENPQERPTSAQVF----DILNSAELVCLTRRILLPKNVIV
NOV74c  KEYGCAPWPMVEKLIKQCLKENPQERPTSAQVFSQVFDILNSAELVCLTRRILLPKNVIV
NOV74d  KEYGCAPWPMVEKLIKQCLKENPQERPTSAQVF----DILNSAELVCLTRRILLPKNVIV
NOV74e  KEYGCAPWPMVEKLIKQCLKENPQERPTSAQVF----DILNSAELVCLTRRILLPKNVIV

NOV74a  ECMVATHHNSRNASIWLGCGHTDRGQLSFLDLNTEGYTSE-------------------
NOV74b  ECMVATHHNSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKE
NOV74c  ECMVATHHNSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKE
```

```
NOV74d  ECMVATHHNSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKE
NOV74e  ECMVATHHNSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKE

NOV74a  ----------------------------------------SKQKNFLLVGTADGKL
NOV74b  SWIVSGTQSGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKL
NOV74c  SWIVSGTQSGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKL
NOV74d  SWIVSGTQSGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKL
NOV74e  SWIVSGTQSGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKL

NOV74a  AIFEDKTVKLKGAAPLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTI
NOV74b  AIFEDKTVKLKGAAPLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTI
NOV74c  AIFEDKTVKLKGAAPLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTI
NOV74d  AIFEDKTVKLKGAAPLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTI
NOV74e  AIFEDKTVKLKGAAPLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTI

NOV74a  QKLIETRTSQLFSYAAFSDSNIITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHF
NOV74b  QKLIETRTSQLFSYAAFSDSNIITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHF
NOV74c  QKLIETRTSQLFSYAAFSDSNIITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHF
NOV74d  QKLIETRTSQLFSYAAFSDSNIITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHF
NOV74e  QKLIETRTSQLFSYAAFSDSNIITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHF

NOV74a  LREVMVKENKESKHKMSYSGRVKTLCLQKNTALWIGTGGGHILLLDLSTRRLIRVIYNFC
NOV74b  L----------S-------------------------------------------------
NOV74c  LRLVKPNRKKLSNLMMSLA--LHPICLKSKLRCSSSKGRSHILL---------RVIYNFC
NOV74d  LRLVKPNRKKLSNLMMSLA--LHPICLKSKLRCSSSKGRSHILL---------RVIYNFC
NOV74e  LREVTVKENKESKHKMSYSGRVKTLCLQKNTALWIGTGGGHILLLDLSTRRLIRVIYNFC

NOV74a  NSVRVMMTAQLG-SLKNVMLVLGYNRKNTEGTQKQKEIQSCLTVWDINLPHEVQNLEKHI
NOV74b  -------------LKNVMLVLGYNRKNTEGTQKQKEIQSCLTVWDINLPHEVQNLEKHI
NOV74c  NSVRVMMTAQLGGSLKNVMLVLGYNRKNTEGTQKQKEIQSCLTVWDINLPHEVQNLEKHI
NOV74d  NSVRVMMTAQLGGSLKNVMLVLGYNRKNTEGTQKQKEIQSCLTVWDINLPHEVQNLEKHI
NOV74e  NSVRVMMTAQLG-SLKNVMLVLGYNRKNTEGTQKQKEIQSCLTVWDINLPHEVQNLEKHI

NOV74a  EVRKELAEKMRRTSVE
NOV74b  EVRKELAEKMRRTSVE
NOV74c  EVRKELAEKMRRTSVE
NOV74d  EVRKELAEKMRRTSVE
NOV74e  EVRKELAEKMRRTSVE

NOV74a  (SEQ ID NO:  1072)
NOV74b  (SEQ ID NO:  1074)
NOV74c  (SEQ ID NO:  1076)
NOV74d  (SEQ ID NO:  1078)
NOV74e  (SEQ ID NO:  1080)
```

Further analysis of the NOV74a protein yielded the following properties shown in Table 74C.

Table 74C. Protein Sequence Properties NOV74a

SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	
PSG: a new signal peptide prediction method N-region: length 5; pos.chg 0; neg.chg 1	

```
      H-region:  length 6;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1): -11.05
      possible cleavage site: between 13 and 14

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   2
      Number of TMS(s) for threshold  0.5:   0
      PERIPHERAL  Likelihood =  1.96 (at 794)
      ALOM score:  -0.64  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  1.98
      Hyd Moment(95):  4.87    G content:       2
      D/E content:     2       S/T content:     3
      Score: -7.81

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: KKRK (5) at  157
      pat7: PVIDRKR (3) at   54
      bipartite: RKNVPIEFPVIDRKRLL at   46
      content of basic residues:  11.6%
      NLS Score:  0.59

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none
```

```
checking 63 PROSITE DNA binding motifs:
      Leucine zipper pattern (PS00029):  *** found ***
         LQLDENELPHAVHFLNESGVLL at 70

      Regulator of chromosome condensation (RCC1) signature 2
(PS00626):  *** found ***
         IGTGGGHILLL at 878

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 89

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       52.2 %: cytoplasmic
       39.1 %: nuclear
        8.7 %: mitochondrial

>> prediction for CG90866-04 is cyt (k=23)
```

A search of the NOV74a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 74D.

Table 74D. Geneseq Results for NOV74a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV74a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU03554	Human protein kinase #54 - Homo sapiens, 909 aa. [WO200138503-A2, 31-MAY-2001]	1..845 1..909	835/909 (91%) 839/909 (91%)	0.0
ABU11054	Human protein NOV19b - Homo sapiens, 911 aa. [WO200281629-A2, 17-OCT-2002]	130..978 4..911	754/963 (78%) 765/963 (79%)	0.0
ABU11053	Human protein NOV19a - Homo sapiens, 915 aa. [WO200281629-A2, 17-OCT-2002]	130..978 4..915	754/967 (77%) 765/967 (78%)	0.0
AAE16259	Human kinase PKIN-5 protein - Homo sapiens, 656 aa. [WO200196547-A2, 20-DEC-2001]	401..978 15..656	567/642 (88%) 571/642 (88%)	0.0
AAU11287	Human transducin polypeptide 41 - Homo sapiens, 373 aa. [CN1306988-A, 08-AUG-2001]	670..978 1..373	298/373 (79%) 302/373 (80%)	e-162

In a BLAST search of public sequence databases, the NOV74a protein was found to have homology to the proteins shown in the BLASTP data in Table 74E.

Table 74E. Public BLASTP Results for NOV74a				
Protein Accession Number	**Protein/Organism/Length**	**NOV74a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8K062	Similar to RIKEN cDNA 4921513O20 gene - Mus musculus (Mouse), 840 aa (fragment).	208..978 1..840	642/884 (72%) 694/884 (77%)	0.0
Q9CQG8	4921513O20Rik protein - Mus musculus (Mouse), 561 aa.	499..978 18..561	401/545 (73%) 441/545 (80%)	0.0
Q8NCX9	Hypothetical protein - Homo sapiens (Human), 400 aa (fragment).	643..978 1..400	326/400 (81%) 330/400 (82%)	e-178
Q8CI84	Hypothetical protein - Mus musculus (Mouse), 327 aa (fragment).	669..978 1..327	239/329 (72%) 269/329 (81%)	e-133
Q8BZJ6	Hypothetical serine/threonine protein kinase containing protein - Mus musculus (Mouse), 546 aa.	14..224 345..543	183/211 (86%) 190/211 (89%)	1e-99

PFam analysis indicates that the NOV74a protein contains the domains shown in the Table 74F.

Table 74F. Domain Analysis of NOV74a			
Pfam Domain	**NOV74a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
pkinase	394..644	90/302 (30%) 169/302 (56%)	1.4e-33

Example 75.

The NOV75 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 75A.

Table 75A. NOV75 Sequence Analysis		
NOV75a, CG91708-02 DNA Sequence	SEQ ID NO: 1081	1580 bp
	ORF Start: ATG at 51	ORF Stop: TGA at 1482

CAAGACAGCAAGGCATAGAGACAACATAGAGCTAAGTAAAGCCAGTGGAAATGAAGAGTCTTCCAATC
CTACTGTTGCTGTGCGTGGCAGTTTGCTCAGCCTATCCATTGGATGGAGCTGCAAGGGGTGAGGACAC
CAGCATGAACCTTGTTCAGAAATATCTAGAAAACTACTACGACCTCGAAAAAGATGTGAAACAGTTTG
TTAGGAGAAAGGACAGTGGTCCTGTTGTTAAAAAAATCCGAGAAATGCAGAAGTTCCTTGGATTGGAG
GTGACGGGGAAGCTGGACTCCGACACTCTGGAGGTGATGCGCAAGCCCATGTGTGGAGTTCCTGACGT
TGGTCACTTCAGAACCTTTCCTGGCATCCCGAAGTGGAGGAAAACCCACCTTACATACAGGATTGTGA

ATTATACACCAGATTTGCCAAAAGATGCTGTTGATTCTGCTGTTGAGAAAGCTCTGAAAGTCTGGGAA
GAGGTGACTCCACTCACATTCTCCAGGCTGTATGAAGGAGAGACTGATATAATGATCTCTTTTGCAGT
TAGAGAACATGGAGACTTTTACCCTTTTGATGGACCTGGAAATGTTTTGGCCCATGCCTATGCCCCTG
GGCCAGGGATTAATGGAGATGCCCACTTTGATGATGATGAACAATGGACAAAGGATACAACAGGGACC
AATTTATTTCTCGTTGCTGCTCATGAAATTGGCCACTCCCTGGGTCTCTTTCACTCAGCCAACACTGA
AGCTTTGATGTACCCACTCTATCACTCACTCACAGACCTGACTCGGTTCCGCCTGTCTCAAGATGATA
TAAATGGCATTCAGTCCCTCTATGGACCTCCCCCTGACTCCCCTGAGACCCCCCTGGTACCCACGGAA
CCTGTCCCTCCAGAACCTGGGACGCCAGCCAACTGTGATCCTGCTTTGTCCTTTGATGCTGTCAGCAC
TCTGAGGGGAGAAATCCTGATCTTTAAAGACAGGCACTTTTGGCGCAAATCCCTCAGGAAGCTTGAAC
CTGAATTGCATTTGATCTCTTCATTTTGGCCATCTCTTCCTTCAGGCGTGGATGCCGCATATGAAGTT
ACTAGCAAGGACCTCGTTTTCATTTTTAAAGGAAATCAATTCTGGGCCATCAGAGGAAATGAGGTACG
AGCTGGATACCCAAGAGGCATCCACACCCTAGGTTTCCCTCCAACCGTGAGGAAAATCGATGCAGCCA
TTTCTGATAAGGAAAAGAACAAAACATATTTCTTTGTAGAGGACAAATACTGGAGATTTGATGAGAAG
AGAAATTCCATGGAGCCAGGCTTTCCCAAGCAAATAGCTGAAGACTTTCCAGGGATTGACTCAAAGAT
TGATGCTGTTTTTGAAGAATTTGGGTTCTTTTATTTCTTTACTGGATCTTCACAGTTGGAGTTTGACC
CAAATGCAAAGAAAGTGACACACACTTTGAAGAGTAACAGCTGGCTTAATTGTTGAAAGAGATATGTA
GAAGGCACAATATGGGCACTTTAAATGAAGCTAATAATTCTTCACCTAAGTCTCTGTGAATTGAAATG
TTCGTTTTCTCCTGCT

NOV75a, CG91708-02 Protein Sequence	SEQ ID NO: 1082	477 aa	MW at 53982.7kD

MKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLEKDVKQFVRRKDSGPVVKKIREMQ
KFLGLEVTGKLDSDTLEVMRKPMCGVPDVGHFRTFPGIPKWRKTHLTYRIVNYTPDLPKDAVDSAVEK
ALKVWEEVTPLTFSRLYEGETDIMISFAVREHGDFYPFDGPGNVLAHAYAPGPGINGDAHFDDDEQWT
KDTTGTNLFLVAAHEIGHSLGLFHSANTEALMYPLYHSLTDLTRFRLSQDDINGIQSLYGPPPDSPET
PLVPTEPVPPEPGTPANCDPALSFDAVSTLRGEILIFKDRHFWRKSLRKLEPELHLISSFWPSLPSGV
DAAYEVTSKDLVFIFKGNQFWAIRGNEVRAGYPRGIHTLGFPPTVRKIDAAISDKEKNKTYFFVEDKY
WRFDEKRNSMEPGFPKQIAEDFPGIDSKIDAVFEEFGFFYFFTGSSQLEFDPNAKKVTHTLKSNSWLN
C

NOV75b, 262751856 DNA Sequence	SEQ ID NO: 1083	1446 bp
	ORF Start: at 1	ORF Stop: end of sequence

GGATCCACCATGAAGAGTCTTCCAATCCTACTGTTGCTGTGCGTGGCAGTTTGCTCAGCCTATCCATT
GGATGGAGCTGCAAGGGGTGAGGACACCAGCATGAACCTTGTTCAGAAATATCTAGAAAACTACTACG
ACCTCGAAAAAGATGTGAAACAGTTTGTTAGGAGAAAGGACAGTGGTCCTGTTGTTAAAAAAATCCGA
GAAATGCAGAAGTTCCTTGGATTGGAGGTGACGGGGAAGCTGGACTCCGACACTCTGGAGGTGATGCG
CAAGCCCAGGTGTGGAGTTCCTGACGTTGGTCACTTCAGAACCTTTCCTGGCATCCCGAAGTGGAGGA
AAACCCACCTTACATACAGGATTGTGAATTATACACCAGATTTGCCAAAAGATGCTGTTGATTCTGCT
GTTGAGAAAGCTCTGAAAGTCTGGGAAGAGGTGACTCCACTCACATTCTCCAGGCTGTATGAAGGAGA
GGCTGATATAATGATCTCTTTTGCAGTTAGAGAACATGGAGACTTTTACCCTTTTGATGGACCTGGAA
ATGTTTTGGCCCATGCCTATGCCCCTGGGCCAGGGATTAATGGAGATGCCCACTTTGATGATGATGAA
CAATGGACAAAGGATACAACAGGGACCAATTTATTTCTCGTTGCTGCTCATGAAATTGGCCACTCCCT
GGGTCTCTTTCACTCAGCCAACACTGAAGCTTTGATGTACCCACTCTATCACTCACTCACAGACCTGA
CTCGGTTCCGCCTGTCTCAAGATGATATAAATGGCATTCAGTCCCTCTATGGACCTCCCCCTGACTCC
CCTGAGACCCCCCCTGGTACCCACGGAACCTGTCCCTCCAGAACCTGGGACGCCAGCCAACTGTGATCC
TGCTTTGTCCTTTGATGCTGTCAGCACTCTGAGGGGAGAAATCCTGATCTTTAAAGACAGGCACTTTT
GGCGCAAATCCCTCAGGAAGCTTGAACCTGAATTGCATTTGATCTCTTCATTTTGGCCATCTCTTCCT
TCAGGCGTGGATGCCGCATATGAAGTTACTAGCAAGGACCTCGTTTTCATTTTTAAAGGAAATCAATT
CTGGGCCATCAGAGGAAATGAGGTACGAGCTGGATACCCAAGAGGCATCCACACCCTAGGTTTCCCTC
CAACCGTGAGGAAAATCGATGCAGCCATTTCTGATAAGGAAAAGAACAAAACATATTTCTTTGTAGAG
GACAAATACTGGAGATTTGATGAGAAGAGAAATTCCATGGAGCCAGGCTTTCCCAAGCAAATAGCTGA
AGACTTTCCAGGGATTGACTCAAAGATTGATGCTGTTTTTGAAGAATTTGGGTTCTTTTATTTCTTTA
CTGGATCTTCACAGTTGGAGTTTGACCCAAATGCAAAGAAAGTGACACACACTTTGAAGAGTAACAGC
TGGCTTAATTGTCTCGAG

NOV75b, 262751856 Protein Sequence	SEQ ID NO: 1084	482 aa	MW at 54465.2kD

GSTMKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLEKDVKQFVRRKDSGPVVKKIR
EMQKFLGLEVTGKLDSDTLEVMRKPRCGVPDVGHFRTFPGIPKWRKTHLTYRIVNYTPDLPKDAVDSA
VEKALKVWEEVTPLTFSRLYEGEADIMISFAVREHGDFYPFDGPGNVLAHAYAPGPGINGDAHFDDDE
QWTKDTTGTNLFLVAAHEIGHSLGLFHSANTEALMYPLYHSLTDLTRFRLSQDDINGIQSLYGPPPDS
PETPLVPTEPVPPEPGTPANCDPALSFDAVSTLRGEILIFKDRHFWRKSLRKLEPELHLISSFWPSLP

SGVDAAYEVTSKDLVFIFKGNQFWAIRGNEVRAGYPRGIHTLGFPPTVRKIDAAISDKEKNKTYFFVE
DKYWRFDEKRNSMEPGFPKQIAEDFPGIDSKIDAVFEEFGFFYFFTGSSQLEFDPNAKKVTHTLKSNS
WLNCLE

NOV75c, CG91708-01 DNA Sequence	SEQ ID NO: 1085	1821 bp
	ORF Start: ATG at 64	ORF Stop: TGA at 1495

ACAAGGAGGCAGGCAAGACAGCAAGGCATAGAGACAACATAGAGCTAAGTAAAGCCAGTGGAAATGAA
GAGTCTTCCAATCCTACTGTTGCTGTGCGTGGCAGTTTGCTCAGCCTATCCATTGGATGGAGCTGCAA
GGGGTGAGGACACCAGCATGAACCTTGTTCAGAAATATCTAGAAAACTACTACGACCTCAAAAAAGAT
GTGAAACAGTTTGTTAGGAGAAAGGACAGTGGTCCTGTTGTTAAAAAAATCCGAGAAATGCAGAAGTT
CCTTGGATTGGAGGTGACGGGGAAGCTGGACTCCGACACTCTGGAGGTGATGCGCAAGCCCAGGTGTG
GAGTTCCTGATGTTGGTCACTTCAGAACCTTTCCTGGCATCCCGAAGTGGAGGAAAACCCACCTTACA
TACAGGATTGTGAATTATACACCAGATTTGCCAAAAGATGCTGTTGATTCTGCTGTTGAGAAAGCTCT
GAAAGTCTGGGAAGAGGTGACTCCACTCACATTCTCCAGGCTGTATGAAGGAGAGGCTGATATAATGA
TCTCTTTTGCAGTTAGAGAACATGGAGACTTTTACCCTTTTGATGGACCTGGAAATGTTTTGGCCCAT
GCCTATGCCCCTGGGCCAGGGATTAATGGAGATGCCCACTTTGATGATGATGAACAATGGACAAAGGA
TACAACAGGGACCAATTTATTTCTCGTTGCTGCTCATGAAATTGGCCACTCCCTGGGTCTCTTTCACT
CAGCCAACACTGAAGCTTTGATGTACCCACTCTATCACTCACTCACAGACCTGACTCGGTTCCGCCTG
TCTCAAGATGATATAAATGGCATTCAGTCCCTCTATGGACCTCCCCCTGACTCCCCTGAGACCCCCCT
GGTACCCACGGAACCTGTCCCTCCAGAACCTGGGACGCCAGCCAACTGTGATCCTGCTTTGTCCTTTG
ATGCTGTCAGCACTCTGAGGGGAGAAATCCTGATCTTTAAAGACAGGCACTTTTGGCGCAAATCCCTC
AGGAAGCTTGAACCTGAATTGCATTTGATCTCTTCATTTTGGCCATCTCTTCCTTCAGGCGTGGATGC
CGCATATGAAGTTACTAGCAAGGACCTCGTTTTCATTTTTAAAGGAAATCAATTCTGGGCCATCAGAG
GAAATGAGGTACGAGCTGGATACCCAAGAGGCATCCACACCCTAGGTTTCCCTCCAACCGTGAGGAAA
ATCGATGCAGCCATTTCTGATAAGGAAAAGAACAAAACATATTTCTTTGTAGAGGACAAATACTGGAG
ATTTGATGAGAAGAGAAATTCCATGGAGCCAGGCTTTCCCAAGCAAATAGCTGAAGACTTTCCAGGGA
TTGACTCAAAGATTGATGCTGTTTTTGAAGAATTTGGGTTCTTTTATTTCTTTACTGGATCTTCACAG
TTGGAGTTTGACCCAAATGCAAAGAAAGTGACACACACTTTGAAGAGTAACAGCTGGCTTAATTGTTG
AAAGAGATATGTAGAAGGCACAATATGGGCACTTTAAATGAAGCTAATAATTCTTCACCTAAGTCTCT
GTGAATTGAAATGTTCGTTTTCTCCTGCCTGTGCTGTGACTCGAGTCACACTCAAGGGAACTTGAGCG
TGAATCTGTATCTTGCCGGTCATTTTTATGTTATTACAGGGCATTCAAATGGGCTGCTGCTTAGCTTG
CACCTTGTCACATAGAGTGATCTTTCCCAAGAGAAGGGGAAGCACTCGTGTGCAACAGACAAGTGACT
GTATCTGTGTAGACTATTTGCTTATTTAATAAAGACGATTTGTCAGTTGTTTT

NOV75c, CG91708-01 Protein Sequence	SEQ ID NO: 1086	477 aa	MW at 53976.7kD

MKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLKKDVKQFVRRKDSGPVVKKIREMQ
KFLGLEVTGKLDSDTLEVMRKPRCGVPDVGHFRTFPGIPKWRKTHLTYRIVNYTPDLPKDAVDSAVEK
ALKVWEEVTPLTFSRLYEGEADIMISFAVREHGDFYPFDGPGNVLAHAYAPGPGINGDAHFDDDEQWT
KDTTGTNLFLVAAHEIGHSLGLFHSANTEALMYPLYHSLTDLTRFRLSQDDINGIQSLYGPPPDSPET
PLVPTEPVPPEPGTPANCDPALSFDAVSTLRGEILIFKDRHFWRKSLRKLEPELHLISSFWPSLPSGV
DAAYEVTSKDLVFIFKGNQFWAIRGNEVRAGYPRGIHTLGFPPTVRKIDAAISDKEKNKTYFFVEDKY
WRFDEKRNSMEPGFPKQIAEDFPGIDSKIDAVFEEFGFFYFFTGSSQLEFDPNAKKVTHTLKSNSWLN
C

NOV75d, CG91708-03 DNA Sequence	SEQ ID NO: 1087	1446 bp
	ORF Start: ATG at 10	ORF Stop: at 1441

GGATCCACCATGAAGAGTCTTCCAATCCTACTGTTGCTGTGCGTGGCAGTTTGCTCAGCCTATCCATT
GGATGGAGCTGCAAGGGGTGAGGACACCAGCATGAACCTTGTTCAGAAATATCTAGAAAACTACTACG
ACCTCGAAAAAGATGTGAAACAGTTTGTTAGGAGAAAGGACAGTGGTCCTGTTGTTAAAAAAATCCGA
GAAATGCAGAAGTTCCTTGGATTGGAGGTGACGGGGAAGCTGGACTCCGACACTCTGGAGGTGATGCG
CAAGCCCAGGTGTGGAGTTCCTGACGTTGGTCACTTCAGAACCTTTCCTGGCATCCCGAAGTGGAGGA
AAACCCACCTTACATACAGGATTGTGAATTATACACCAGATTTGCCAAAAGATGCTGTTGATTCTGCT
GTTGAGAAAGCTCTGAAAGTCTGGGAAGAGGTGACTCCACTCACATTCTCCAGGCTGTATGAAGGAGA
GGCTGATATAATGATCTCTTTTGCAGTTAGAGAACATGGAGACTTTTACCCTTTTGATGGACCTGGAA
ATGTTTTGGCCCATGCCTATGCCCCTGGGCCAGGGATTAATGGAGATGCCCACTTTGATGATGATGAA
CAATGGACAAAGGATACAACAGGGACCAATTTATTTCTCGTTGCTGCTCATGAAATTGGCCACTCCCT
GGGTCTCTTTCACTCAGCCAACACTGAAGCTTTGATGTACCCACTCTATCACTCACTCACAGACCTGA
CTCGGTTCCGCCTGTCTCAAGATGATATAAATGGCATTCAGTCCCTCTATGGACCTCCCCCTGACTCC
CCTGAGACCCCCCTGGTACCCACGGAACCTGTCCCTCCAGAACCTGGGACGCCAGCCAACTGTGATCC
TGCTTTGTCCTTTGATGCTGTCAGCACTCTGAGGGGAGAAATCCTGATCTTTAAAGACAGGCACTTTT
GGCGCAAATCCCTCAGGAAGCTTGAACCTGAATTGCATTTGATCTCTTCATTTTGGCCATCTCTTCCT

TCAGGCGTGGATGCCGCATATGAAGTTACTAGCAAGGACCTCGTTTTCATTTTAAAGGAAATCAATT
CTGGGCCATCAGAGGAAATGAGGTACGAGCTGGATACCCAAGAGGCATCCACACCCTAGGTTTCCCTC
CAACCGTGAGGAAAATCGATGCAGCCATTTCTGATAAGGAAAAGAACAAAACATATTTCTTTGTAGAG
GACAAATACTGGAGATTTGATGAGAAGAGAAATTCCATGGAGCCAGGCTTTCCCAAGCAAATAGCTGA
AGACTTTCCAGGGATTGACTCAAAGATTGATGCTGTTTTTGAAGAATTTGGGTTCTTTTATTTCTTTA
CTGGATCTTCACAGTTGGAGTTTGACCCAAATGCAAAGAAAGTGACACACACTTTGAAGAGTAACAGC
TGGCTTAATTGTCTCGAG

NOV75d, CG91708-03 Protein Sequence	SEQ ID NO: 1088	477 aa	MW at 53977.6kD

MKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLEKDVKQFVRRKDSGPVVKKIREMQ
KFLGLEVTGKLDSDTLEVMRKPRCGVPDVGHFRTFPGIPKWRKTHLTYRIVNYTPDLPKDAVDSAVEK
ALKVWEEVTPLTFSRLYEGEADIMISFAVREHGDFYPFDGPGNVLAHAYAPGPGINGDAHFDDDEQWT
KDTTGTNLFLVAAHEIGHSLGLFHSANTEALMYPLYHSLTDLTRFRLSQDDINGIQSLYGPPPDSPET
PLVPTEPVPPEPGTPANCDPALSFDAVSTLRGEILIFKDRHFWRKSLRKLEPELHLISSFWPSLPSGV
DAAYEVTSKDLVFIFKGNQFWAIRGNEVRAGYPRGIHTLGFPPTVRKIDAAISDKEKNKTYFFVEDKY
WRFDEKRNSMEPGFPKQIAEDFPGIDSKIDAVFEEFGFFYFFTGSSQLEFDPNAKKVTHTLKSNSWLN
C

NOV75e, CG91708-04 DNA Sequence	SEQ ID NO: 1089	1446 bp
	ORF Start: ATG at 10	ORF Stop: at 1441

GGATCCACC**ATG**AAGAGTCTTCCAATCCTACTGTTGCTGTGCGTGGCAGTTTGCTCAGCCTATCCATT
GGATGGAGCTGCAAGGGGTGAGGACACCAGCATGAACCTTGTTCAGAAATATCTAGAAAACTACTACG
ACCTCAAAAAAGATGTGAAACAGTTTGTTAGGAGAAAGGACAGTGGTCCTGTTGTTAAAAAAATCCGA
GAAATGCAGAAGTTCCTTGGATTGGAGGTGACGGGGAAGCTGGACTCCGACACTCTGGAGGTGATGCG
CAAGCCCAGGTGTGGAGTTCCTGACGTTGGTCACTTCAGAACCTTTCCTGGCATCCCGAAGTGGAGGA
AAACCCACCTTACATACAGGATTGTGAATTATACACCAGATTTGCCAAAAGATGCTGTTGATTCTGCT
GTTGAGAAAGCTCTGAAAGTCTGGGAAGAGGTGACTCCACTCACATTCTCCAGGCTGTATGAAGGAGA
GGCTGATATAATGATCTCTTTTGCAGTTAGAGAACATGGAGACTTTTACCCTTTTGATGGACCTGGAA
ATGTTTTGGCCCATGCCTATGCCCCTGGGCCAGGGATTAATGGAGATGCCCACTTTGATGATGATGAA
CAATGGACAAAGGATACAACAGGGACCAATTTATTTCTCGTTGCTGCTCATGAAATTGGCCACTCCCT
GGGTCTCTTTCACTCAGCCAACACTGAAGCTTTGATGTACCCACTCTATCACTCACTCACAGACCTGA
CTCGGTTCCGCCTGTCTCAAGATGATATAAATGGCATTCAGTCCCTCTATGGACCTCCCCCTGACTCC
CCTGAGACCCCCCTGGTACCCACGGAACCTGTCCCTCCAGAACCTGGGACGCCAGCCAACTGTGATCC
TGCTTTGTCCTTTGATGCTGTCAGCACTCTGAGGGGAGAAATCCTGATCTTTAAAGACAGGCACTTTT
GGCGCAAATCCCTCAGGAAGCTTGAACCTGAATTGCATTTGATCTCTTCATTCTGGCCATCTCTTCCT
TCAGGCGTGGATGCCGCATATGAAGTTACTAGCAAGGACCTCGTTTTCATTTTAAAGGAAATCAATT
CTGGGCCATCAGAGGAAATGAGGTACGAGCTGGATACCCAAGAGGCATCCACACCCTAGGTTTCCCTC
CAACCGTGAGGAAAATCGATGCAGCCATTTCTGATAAGGAAAAGAACAAAACATATTTCTTTGTAGAG
GACAAATACTGGAGATTTGATGAGAAGAGAAATTCCATGGAGCCAGGCTTTCCCAAGCAAATAGCTGA
AGACTTTCCAGGGATTGACTCAAAGATTGATGCTGTTTTTGAAGAATTTGGGTTCTTTTATTTCTTTA
CTGGATCTTCACAGTTGGAGTTTGACCCAAATGCAAAGAAAGTGACACACACTTTGAAGAGTAACAGC
TGGCTTAATTGTCTCGAG

NOV75e, CG91708-04 Protein Sequence	SEQ ID NO: 1090	477 aa	MW at 53976.7kD

MKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLKKDVKQFVRRKDSGPVVKKIREMQ
KFLGLEVTGKLDSDTLEVMRKPRCGVPDVGHFRTFPGIPKWRKTHLTYRIVNYTPDLPKDAVDSAVEK
ALKVWEEVTPLTFSRLYEGEADIMISFAVREHGDFYPFDGPGNVLAHAYAPGPGINGDAHFDDDEQWT
KDTTGTNLFLVAAHEIGHSLGLFHSANTEALMYPLYHSLTDLTRFRLSQDDINGIQSLYGPPPDSPET
PLVPTEPVPPEPGTPANCDPALSFDAVSTLRGEILIFKDRHFWRKSLRKLEPELHLISSFWPSLPSGV
DAAYEVTSKDLVFIFKGNQFWAIRGNEVRAGYPRGIHTLGFPPTVRKIDAAISDKEKNKTYFFVEDKY
WRFDEKRNSMEPGFPKQIAEDFPGIDSKIDAVFEEFGFFYFFTGSSQLEFDPNAKKVTHTLKSNSWLN
C

NOV75f, SNP13380740 of CG91708-02, DNA Sequence	SEQ ID NO: 1091	1580 bp
	ORF Start: ATG at 51	ORF Stop: TGA at 1482
	SNP Pos: 183	SNP Change: G to A

CAAGACAGCAAGGCATAGAGACAACATAGAGCTAAGTAAAGCCAGTGGAAA**TG**AAGAGTCTTCCAATC
CTACTGTTGCTGTGCGTGGCAGTTTGCTCAGCCTATCCATTGGATGGAGCTGCAAGGGGTGAGGACAC
CAGCATGAACCTTGTTCAGAAATATCTAGAAAACTACTACGACCTCAAAAAAGATGTGAAACAGTTTG

TTAGGAGAAAGGACAGTGGTCCTGTTGTTAAAAAAATCCGAGAAATGCAGAAGTTCCTTGGATTGGAG
GTGACGGGGAAGCTGGACTCCGACACTCTGGAGGTGATGCGCAAGCCCATGTGTGGAGTTCCTGACGT
TGGTCACTTCAGAACCTTTCCTGGCATCCCGAAGTGGAGGAAAACCCACCTTACATACAGGATTGTGA
ATTATACACCAGATTTGCCAAAAGATGCTGTTGATTCTGCTGTTGAGAAAGCTCTGAAAGTCTGGGAA
GAGGTGACTCCACTCACATTCTCCAGGCTGTATGAAGGAGAGACTGATATAATGATCTCTTTTGCAGT
TAGAGAACATGGAGACTTTTACCCTTTTGATGGACCTGGAAATGTTTTGGCCCATGCCTATGCCCCTG
GGCCAGGGATTAATGGAGATGCCCACTTTGATGATGATGAACAATGGACAAAGGATACAACAGGGACC
AATTTATTTCTCGTTGCTGCTCATGAAATTGGCCACTCCCTGGGTCTCTTTCACTCAGCCAACACTGA
AGCTTTGATGTACCCACTCTATCACTCACTCACAGACCTGACTCGGTTCCGCCTGTCTCAAGATGATA
TAAATGGCATTCAGTCCCTCTATGGACCTCCCCCTGACTCCCCTGAGACCCCCCTGGTACCCACGGAA
CCTGTCCCTCCAGAACCTGGGACGCCAGCCAACTGTGATCCTGCTTTGTCCTTTGATGCTGTCAGCAC
TCTGAGGGGAGAAATCCTGATCTTTAAAGACAGGCACTTTTGGCGCAAATCCCTCAGGAAGCTTGAAC
CTGAATTGCATTTGATCTCTTCATTTTGGCCATCTCTTCCTTCAGGCGTGGATGCCGCATATGAAGTT
ACTAGCAAGGACCTCGTTTTCATTTTTAAAGGAAATCAATTCTGGGCCATCAGAGGAAATGAGGTACG
AGCTGGATACCCAAGAGGCATCCACACCCTAGGTTTCCCTCCAACCGTGAGGAAAATCGATGCAGCCA
TTTCTGATAAGGAAAAGAACAAAACATATTTCTTTGTAGAGGACAAATACTGGAGATTTGATGAGAAG
AGAAATTCCATGGAGCCAGGCTTTCCCAAGCAAATAGCTGAAGACTTTCCAGGGATTGACTCAAAGAT
TGATGCTGTTTTTGAAGAATTTGGGTTCTTTTATTTCTTTACTGGATCTTCACAGTTGGAGTTTGACC
CAAATGCAAAGAAAGTGACACACACTTTGAAGAGTAACAGCTGGCTTAATTGTTGAAAGAGATATGTA
GAAGGCACAATATGGGCACTTTAAATGAAGCTAATAATTCTTCACCTAAGTCTCTGTGAATTGAAATG
TTCGTTTTCTCCTGCT

NOV75f, SNP13380740 of CG91708-02, Protein Sequence	SEQ ID NO: 1092	477 aa	MW at 53981.8kD
	SNP Pos: 45		SNP Change: Glu to Lys

MKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLKKDVKQFVRRKDSGPVVKKIREMQ
KFLGLEVTGKLDSDTLEVMRKPMCGVPDVGHFRTFPGIPKWRKTHLTYRIVNYTPDLPKDAVDSAVEK
ALKVWEEVTPLTFSRLYEGETDIMISFAVREHGDFYPFDGPGNVLAHAYAPGPGINGDAHFDDDEQWT
KDTTGTNLFLVAAHEIGHSLGLFHSANTEALMYPLYHSLTDLTRFRLSQDDINGIQSLYGPPPDSPET
PLVPTEPVPPEPGTPANCDPALSFDAVSTLRGEILIFKDRHFWRKSLRKLEPELHLISSFWPSLPSGV
DAAYEVTSKDLVFIFKGNQFWAIRGNEVRAGYPRGIHTLGFPPTVRKIDAAISDKEKNKTYFFVEDKY
WRFDEKRNSMEPGFPKQIAEDFPGIDSKIDAVFEEFGFFYFFTGSSQLEFDPNAKKVTHTLKSNSWLN
C

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table 75B.

Table 75B. Comparison of the NOV75 protein sequences.

```
NOV75a  ---MKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLEKDVKQFVRRKDS
NOV75b  GSTMKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLEKDVKQFVRRKDS
NOV75c  ---MKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLKKDVKQFVRRKDS
NOV75d  ---MKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLEKDVKQFVRRKDS
NOV75e  ---MKSLPILLLLCVAVCSAYPLDGAARGEDTSMNLVQKYLENYYDLKKDVKQFVRRKDS

NOV75a  GPVVKKIREMQKFLGLEVTGKLDSDTLEVMRKPMCGVPDVGHFRTFPGIPKWRKTHLTYR
NOV75b  GPVVKKIREMQKFLGLEVTGKLDSDTLEVMRKPRCGVPDVGHFRTFPGIPKWRKTHLTYR
NOV75c  GPVVKKIREMQKFLGLEVTGKLDSDTLEVMRKPRCGVPDVGHFRTFPGIPKWRKTHLTYR
NOV75d  GPVVKKIREMQKFLGLEVTGKLDSDTLEVMRKPRCGVPDVGHFRTFPGIPKWRKTHLTYR
NOV75e  GPVVKKIREMQKFLGLEVTGKLDSDTLEVMRKPRCGVPDVGHFRTFPGIPKWRKTHLTYR

NOV75a  IVNYTPDLPKDAVDSAVEKALKVWEEVTPLTFSRLYEGETDIMISFAVREHGDFYPFDGP
NOV75b  IVNYTPDLPKDAVDSAVEKALKVWEEVTPLTFSRLYEGEADIMISFAVREHGDFYPFDGP
NOV75c  IVNYTPDLPKDAVDSAVEKALKVWEEVTPLTFSRLYEGEADIMISFAVREHGDFYPFDGP
NOV75d  IVNYTPDLPKDAVDSAVEKALKVWEEVTPLTFSRLYEGEADIMISFAVREHGDFYPFDGP
NOV75e  IVNYTPDLPKDAVDSAVEKALKVWEEVTPLTFSRLYEGEADIMISFAVREHGDFYPFDGP

NOV75a  GNVLAHAYAPGPGINGDAHFDDDEQWTKDTTGTNLFLVAAHEIGHSLGLFHSANTEALMY
NOV75b  GNVLAHAYAPGPGINGDAHFDDDEQWTKDTTGTNLFLVAAHEIGHSLGLFHSANTEALMY
NOV75c  GNVLAHAYAPGPGINGDAHFDDDEQWTKDTTGTNLFLVAAHEIGHSLGLFHSANTEALMY
```

```
NOV75d  GNVLAHAYAPGPGINGDAHFDDDEQWTKDTTGTNLFLVAAHEIGHSLGLFHSANTEALMY
NOV75e  GNVLAHAYAPGPGINGDAHFDDDEQWTKDTTGTNLFLVAAHEIGHSLGLFHSANTEALMY

NOV75a  PLYHSLTDLTRFRLSQDDINGIQSLYGPPPDSPETPLVPTEPVPPEPGTPANCDPALSFD
NOV75b  PLYHSLTDLTRFRLSQDDINGIQSLYGPPPDSPETPLVPTEPVPPEPGTPANCDPALSFD
NOV75c  PLYHSLTDLTRFRLSQDDINGIQSLYGPPPDSPETPLVPTEPVPPEPGTPANCDPALSFD
NOV75d  PLYHSLTDLTRFRLSQDDINGIQSLYGPPPDSPETPLVPTEPVPPEPGTPANCDPALSFD
NOV75e  PLYHSLTDLTRFRLSQDDINGIQSLYGPPPDSPETPLVPTEPVPPEPGTPANCDPALSFD

NOV75a  AVSTLRGEILIFKDRHFWRKSLRKLEPELHLISSFWPSLPSGVDAAYEVTSKDLVFIFKG
NOV75b  AVSTLRGEILIFKDRHFWRKSLRKLEPELHLISSFWPSLPSGVDAAYEVTSKDLVFIFKG
NOV75c  AVSTLRGEILIFKDRHFWRKSLRKLEPELHLISSFWPSLPSGVDAAYEVTSKDLVFIFKG
NOV75d  AVSTLRGEILIFKDRHFWRKSLRKLEPELHLISSFWPSLPSGVDAAYEVTSKDLVFIFKG
NOV75e  AVSTLRGEILIFKDRHFWRKSLRKLEPELHLISSFWPSLPSGVDAAYEVTSKDLVFIFKG

NOV75a  NQFWAIRGNEVRAGYPRGIHTLGFPPTVRKIDAAISDKEKNKTYFFVEDKYWRFDEKRNS
NOV75b  NQFWAIRGNEVRAGYPRGIHTLGFPPTVRKIDAAISDKEKNKTYFFVEDKYWRFDEKRNS
NOV75c  NQFWAIRGNEVRAGYPRGIHTLGFPPTVRKIDAAISDKEKNKTYFFVEDKYWRFDEKRNS
NOV75d  NQFWAIRGNEVRAGYPRGIHTLGFPPTVRKIDAAISDKEKNKTYFFVEDKYWRFDEKRNS
NOV75e  NQFWAIRGNEVRAGYPRGIHTLGFPPTVRKIDAAISDKEKNKTYFFVEDKYWRFDEKRNS

NOV75a  MEPGFPKQIAEDFPGIDSKIDAVFEEFGFFYFFTGSSQLEFDPNAKKVTHTLKSNSWLNC
NOV75b  MEPGFPKQIAEDFPGIDSKIDAVFEEFGFFYFFTGSSQLEFDPNAKKVTHTLKSNSWLNC
NOV75c  MEPGFPKQIAEDFPGIDSKIDAVFEEFGFFYFFTGSSQLEFDPNAKKVTHTLKSNSWLNC
NOV75d  MEPGFPKQIAEDFPGIDSKIDAVFEEFGFFYFFTGSSQLEFDPNAKKVTHTLKSNSWLNC
NOV75e  MEPGFPKQIAEDFPGIDSKIDAVFEEFGFFYFFTGSSQLEFDPNAKKVTHTLKSNSWLNC

NOV75a  --
NOV75b  LE
NOV75c  --
NOV75d  --
NOV75e  --

NOV75a  (SEQ ID NO:  1082)
NOV75b  (SEQ ID NO:  1084)
NOV75c  (SEQ ID NO:  1086)
NOV75d  (SEQ ID NO:  1088)
NOV75e  (SEQ ID NO:  1090)
```

Further analysis of the NOV75a protein yielded the following properties shown in Table 75C.

Table 75C. Protein Sequence Properties NOV75a

SignalP analysis:	Cleavage site between residues 18 and 19
PSORT II analysis:	

```
PSG:  a new signal peptide prediction method
      N-region:  length 2;  pos.chg 1;  neg.chg 0
      H-region:  length 18;  peak value  10.73
      PSG score:   6.33

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   3.45
      possible cleavage site: between 17 and 18
```

```
>>> Seems to have a cleavable signal peptide (1 to 17)

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 18
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  3.87 (at 293)
      ALOM score:   3.87  (number of TMSs: 0)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 8
      Charge difference: -4.0   C(-2.0) - N( 2.0)
      N >= C: N-terminal side will be inside

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  3.04
      Hyd Moment(95):  6.62    G content:       0
      D/E content:     1       S/T content:     2
      Score: -4.76

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: PVVKKIR (3) at   59
      pat7: PKWRKTH (5) at  107
      bipartite: none
      content of basic residues:  11.3%
      NLS Score:  0.22

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none
```

```
checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination
      Prediction: cytoplasmic
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues

--------------------------

Final Results (k = 9/23):

       33.3 %: extracellular, including cell wall
       22.2 %: mitochondrial
       22.2 %: endoplasmic reticulum
       11.1 %: Golgi
       11.1 %: vacuolar

>> prediction for CG91708-02 is exc (k=9)
```

A search of the NOV75a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 75D.

Table 75D. Geneseq Results for NOV75a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV75a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABB04752	Human MMP3 protein SEQ ID NO:3 - Homo sapiens, 477 aa. [WO200179238-A2, 25-OCT-2001]	1..477 1..477	475/477 (99%) 475/477 (99%)	0.0
AAB84608	Amino acid sequence of matrix metalloproteinase-3 stromelysin 1 - Homo sapiens, 477 aa. [WO200149309-A2, 12-JUL-2001]	1..477 1..477	475/477 (99%) 475/477 (99%)	0.0
AAO20482	Prostromelysin protein - Homo sapiens, 477 aa. [US6284513-B1, 04-SEP-2001]	1..477 1..477	475/477 (99%) 475/477 (99%)	0.0
ABP54461	Matrix metalloproteinase 3 amino acid sequence SEQ ID NO:2919 - Homo sapiens, 477 aa. [WO200278516-A2, 10-OCT-2002]	1..477 1..477	474/477 (99%) 475/477 (99%)	0.0
ABU03473	Angiogenesis-associated human protein sequence #18 - Homo sapiens, 477 aa. [WO200279492-A2, 10-OCT-2002]	1..477 1..477	474/477 (99%) 475/477 (99%)	0.0

In a BLAST search of public sequence databases, the NOV75a protein was found to have homology to the proteins shown in the BLASTP data in Table 75E.

Table 75E. Public BLASTP Results for NOV75a				
Protein Accession Number	**Protein/Organism/Length**	**NOV75a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
P08254	Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3) (MMP-3) (Transin-1) (SL-1) - Homo sapiens (Human), 477 aa.	1..477 1..477	474/477 (99%) 475/477 (99%)	0.0
P28863	Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3) (MMP-3) (Transin-1) (SL-1) - Oryctolagus cuniculus (Rabbit), 478 aa.	1..477 1..478	401/478 (83%) 433/478 (89%)	0.0
Q28397	Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3) (MMP-3) - Equus caballus (Horse), 477 aa.	1..477 1..477	386/477 (80%) 427/477 (88%)	0.0
P09238	Stromelysin-2 precursor (EC 3.4.24.22) (Matrix metalloproteinase-10) (MMP-10) (Transin-2) (SL-2) - Homo sapiens (Human), 476 aa.	1..477 1..476	372/477 (77%) 418/477 (86%)	0.0
Q922W6	Matrix metalloproteinase 3 - Mus musculus (Mouse), 479 aa.	1..477 3..479	366/477 (76%) 413/477 (85%)	0.0

PFam analysis indicates that the NOV75a protein contains the domains shown in the Table 75F.

CLAIMS

What is claimed is:

1. An isolated polypeptide comprising the mature form of an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 566.

2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 566.

3. An isolated polypeptide comprising an amino acid sequence which is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 566.

4. An isolated polypeptide, wherein the polypeptide comprises an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 566.

5. The polypeptide of claim 1 wherein said polypeptide is naturally occurring.

6. A composition comprising the polypeptide of claim 1 and a carrier.

7. A kit comprising, in one or more containers, the composition of claim 6.

8. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein the therapeutic comprises the polypeptide of claim 1.

9. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

(a) providing said sample;

(b) introducing said sample to an antibody that binds immunospecifically to the polypeptide; and

(c) determining the presence or amount of antibody bound to said polypeptide,

thereby determining the presence or amount of polypeptide in said sample.

10. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and

b) comparing the expression of said polypeptide in the sample of step (a) to the expression of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,

wherein an alteration in the level of expression of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

11. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:

(a) introducing said polypeptide to said agent; and

(b) determining whether said agent binds to said polypeptide.

12. The method of claim 11 wherein the agent is a cellular receptor or a downstream effector.

13. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:

(a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;

(b) contacting the cell with a composition comprising a candidate substance; and

(c) determining whether the substance alters the property or function ascribable to the polypeptide;

whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition in the absence of the substance, the substance is identified as a potential therapeutic agent.

14. A method for screening for a modulator of activity of or of latency or predisposition to a pathology associated with the polypeptide of claim 1, said method comprising:

(a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein said test animal recombinantly expresses the polypeptide of claim 1;

(b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a); and

(c) comparing the activity of said polypeptide in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator of activity or of latency or of predisposition to, a pathology associated with the polypeptide of claim 1.

15. The method of claim 14, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a

promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

16. A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of claim 1 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

17. A method of treating or preventing a pathology associated with the polypeptide of claim 1, the method comprising administering the polypeptide of claim 1 to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

18. The method of claim 17, wherein the subject is a human.

19. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 566 or a biologically active fragment thereof.

20. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 566.

21. The nucleic acid molecule of claim 20, wherein the nucleic acid molecule is naturally occurring.

22. A nucleic acid molecule, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 566.

23. An isolated nucleic acid molecule encoding the mature form of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 566.

24. An isolated nucleic acid molecule comprising a nucleic acid selected from the group consisting of 2n-1, wherein n is an integer between 1 and 566.

25. The nucleic acid molecule of claim 20, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 566, or a complement of said nucleotide sequence.

26. A vector comprising the nucleic acid molecule of claim 20.

27. The vector of claim 26, further comprising a promoter operably linked to said nucleic acid molecule.

28. A cell comprising the vector of claim 26.

29. An antibody that immunospecifically binds to the polypeptide of claim 1.

30. The antibody of claim 29, wherein the antibody is a monoclonal antibody.

31. The antibody of claim 29, wherein the antibody is a humanized antibody.

32. A method for determining the presence or amount of the nucleic acid molecule of claim 20 in a sample, the method comprising:

(a) providing said sample;

(b) introducing said sample to a probe that binds to said nucleic acid molecule; and

(c) determining the presence or amount of said probe bound to said nucleic acid molecule,

thereby determining the presence or amount of the nucleic acid molecule in said sample.

33. The method of claim 32 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

34. The method of claim 33 wherein the cell or tissue type is cancerous.

35. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the nucleic acid molecule of claim 20 in a first mammalian subject, the method comprising:

a) measuring the level of expression of the nucleic acid in a sample from the first mammalian subject; and

b) comparing the level of expression of said nucleic acid in the sample of step (a) to the level of expression of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of expression of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

36. A method of producing the polypeptide of claim 1, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 566.

37. The method of claim 36 wherein the cell is a bacterial cell.

38. The method of claim 36 wherein the cell is an insect cell.

39. The method of claim 36 wherein the cell is a yeast cell.

40. The method of claim 36 wherein the cell is a mammalian cell.

41. A method of producing the polypeptide of claim 2, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 566.

42. The method of claim 41 wherein the cell is a bacterial cell.

43. The method of claim 41 wherein the cell is an insect cell.

44. The method of claim 41 wherein the cell is a yeast cell.

45. The method of claim 41 wherein the cell is a mammalian cell.

Figure D1. Alignment of CG55806-04 (SEQ ID NO:748), CG55806-02 (SEQ ID NO:752), and 1PFX (SEQ ID NO: 1476).

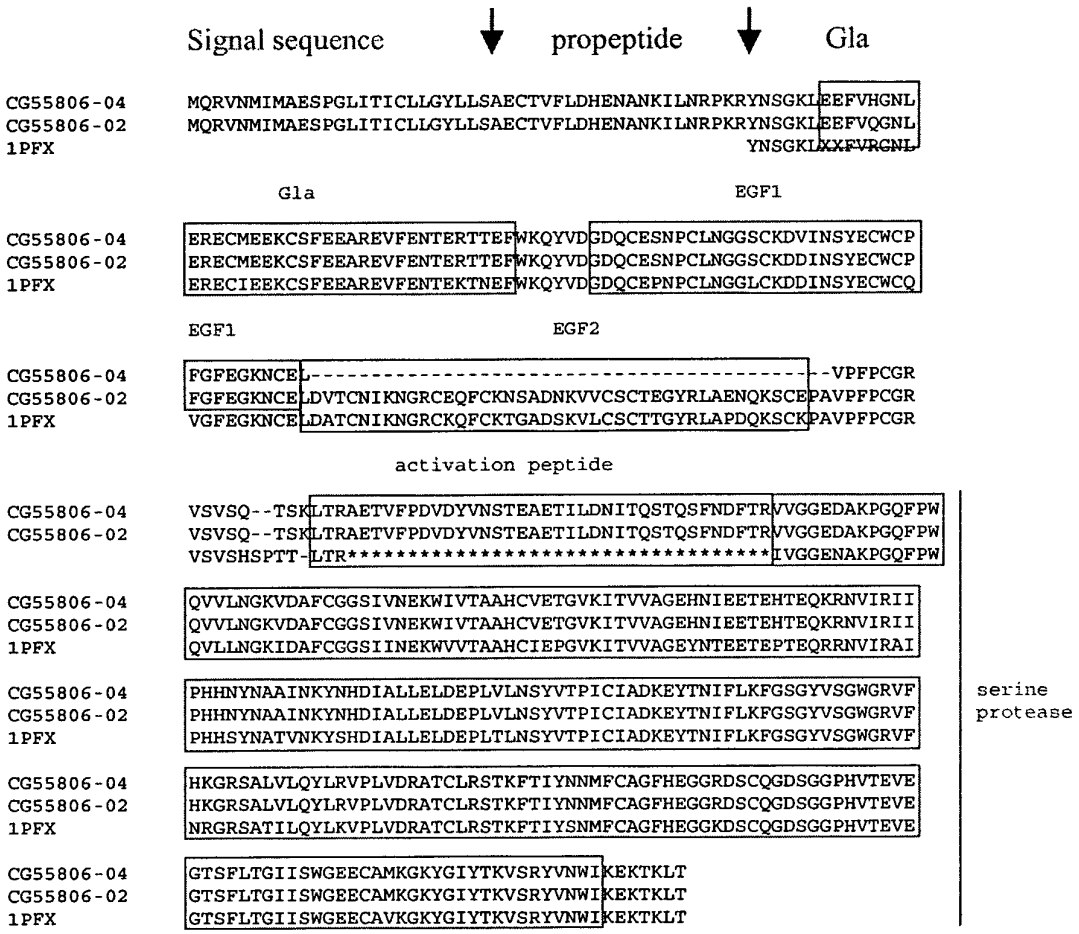
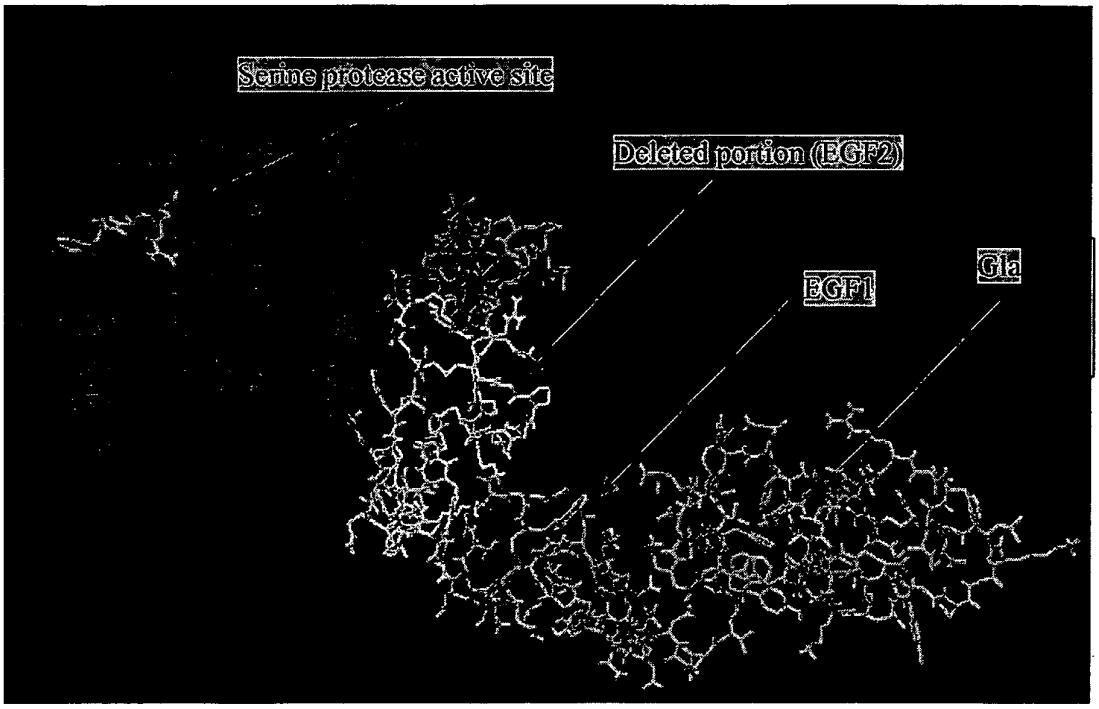


Figure D2



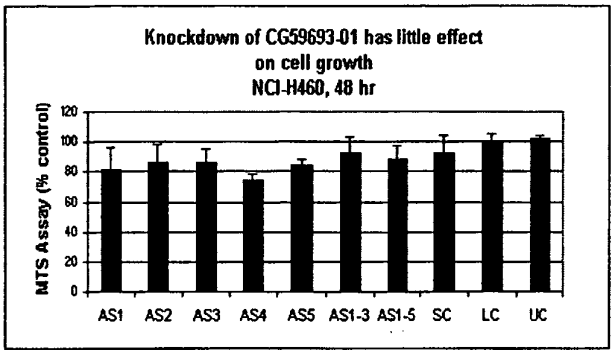

Knockdown of CG59693-01 has little effect
on cell growth
NCI-H460, 48 hr
MTS Assay (% control)
120
100
80
60
40
20
0
AS1
AS2
AS3
AS4
AS5
AS1-3
AS1-5
SC
LC
UC

Figure E1.

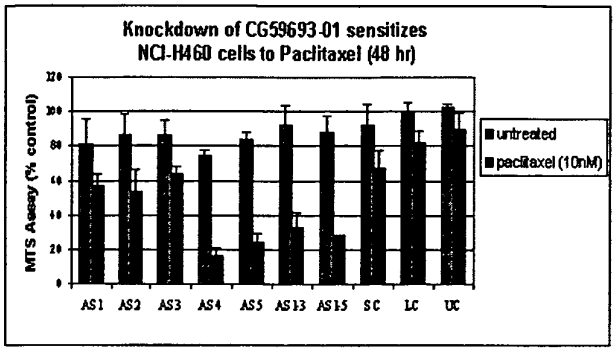


Figure E2

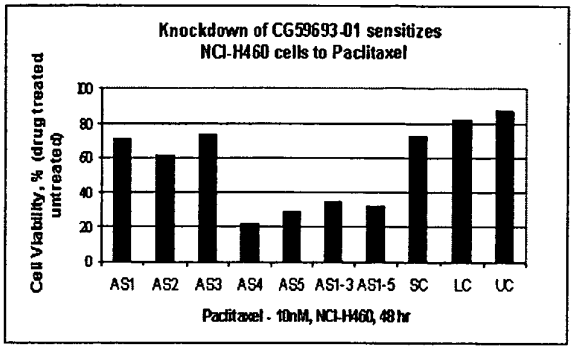


Figure E3

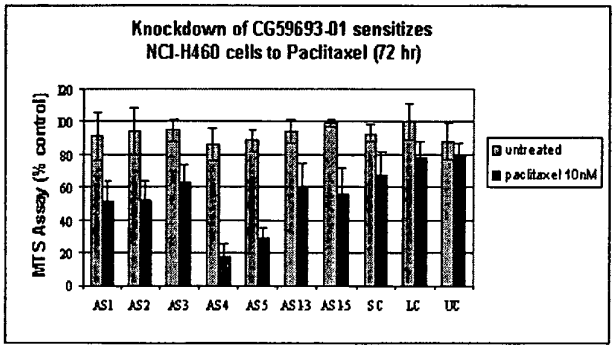


Figure E4

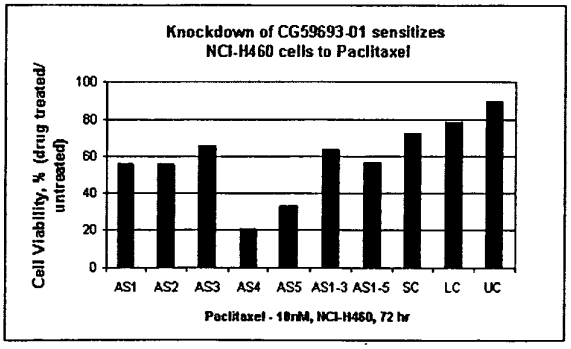


Figure E5

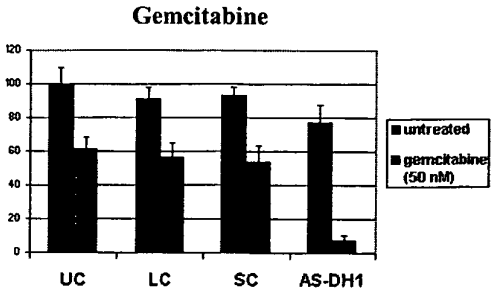


Figure E6

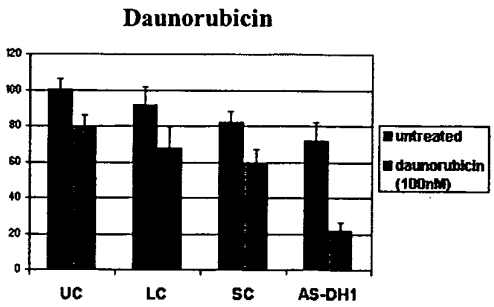


Figure E7

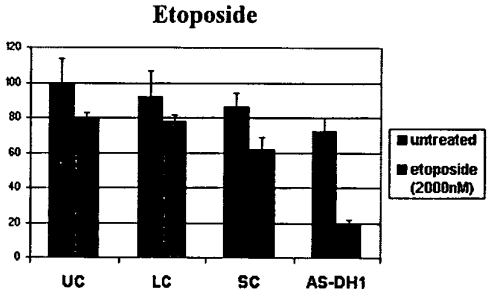


Figure E8

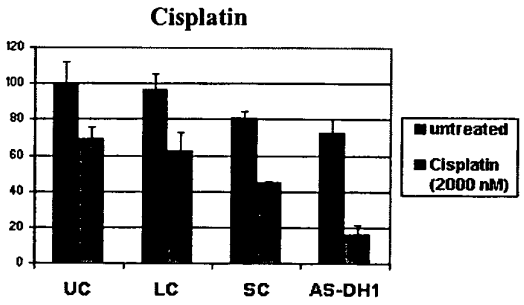


Figure E9