Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2478032 A1 2003/09/18

(21) 2 478 032

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2003/03/06
(87) Date publication PCT/PCT Publication Date: 2003/09/18
(85) Entrée phase nationale/National Entry: 2004/09/02
(86) N° demande PCT/PCT Application No.: US 2003/006794
(87) N° publication PCT/PCT Publication No.: 2003/076578
(30) Priorités/Priorities: 2002/03/06 (60/361,974) US;
2002/03/08 (10/093,463) US;
2002/03/08 (PCT/US02/07288) US;
2002/03/15 (60/365,034) US; 2002/03/19 (60/365,477) US;
2002/03/20 (60/365,984) US; 2002/03/20 (60/365,884) US;
2002/03/20 (60/365,985) US; 2002/03/22 (60/366,928) US;
2002/04/12 (60/372,018) US; 2002/04/12 (60/372,022) US;
...

(51) Cl.Int.^{7}/Int.Cl.^{7} C12N 15/12, A61K 38/17, C12Q 1/02, C12P 21/02, C12Q 1/00, C12Q 1/68, C07K 14/47, C07K 16/18, G01N 33/53

(71) Demandeur/Applicant:
CURAGEN CORPORATION, US

(72) Inventeurs/Inventors:
ALSOBROOK, JOHN P., II, US;
ANDERSON, DAVID W., US;
BOLDOG, FERENC L., US;
BURGESS, CATHERINE E., US;
CHAUDHURI, AMITABHA, US;
...

(74) Agent: GOUDREAU GAGE DUBUC

(54) Titre : POLYPEPTIDES THERAPEUTIQUES, ACIDES NUCLEIQUES CODANT POUR EUX ET LEUR PROCEDE D'UTILISATION

(54) Title: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

(57) Abrégé/Abstract:
The present invention relates to novel polypeptides, and the nucleic acids encoding them, having 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.

(21) 2 478 032

(13) A1

(30) Priorités(suite)/Priorities(continued): 2002/04/23 (60/374,682) US; 2002/06/12 (60/388,096) US; 2002/06/14 (60/389,143) US; 2002/06/26 (60/391,779) US; 2002/08/15 (60/403,743) US; 2002/09/13 (60/410,755) US; 2002/09/23 (60/412,957) US; 2002/10/22 (60/420,382) US; 2003/03/05 () US

(72) Inventeurs(suite)/Inventors(continued): COLMAN, STEVEN D., US; EDINGER, SHLOMIT R., US; ETTENBERG, SETH, US; GANGOLLI, ESHA A., US; GERLACH, VALERIE L., US; GORMAN, LINDA, US; GUO, XIAOJIA, US; KEKUDA, RAMESH, US; LI, LI, US; MACLACHLAN, TIMOTHY, US; MALYANKAR, URIEL M., US; MEZES, PETER S., US; MILLER, CHARLES E., US; MILLET, ISABELLE, US; PADIGARU, MURALIDHARA, US; PATTURAJAN, MEERA, US; PEYMAN, JOHN, US; QIAN, XIAZHONG, US; RASTELLI, LUCA, US; RIEGER, DANIEL K., US; SMITHSON, GLENNDA, US; SPYTEK, KIMBERLY A., US; STONE, DAVID J., US; SUKUMARAN, SUJATHA, US; VERNET, CORINE A. M., US; VOSS, EDWARD Z., US; ZHONG, MEI, US

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 September 2003 (18.09.2003)

PCT

(10) International Publication Number
WO 03/076578 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US03/06794

(22) International Filing Date: 6 March 2003 (06.03.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/361,974	6 March 2002 (06.03.2002)	US
10/093,463	8 March 2002 (08.03.2002)	US
PCT/US02/07288	8 March 2002 (08.03.2002)	US
60/365,034	15 March 2002 (15.03.2002)	US
60/365,477	19 March 2002 (19.03.2002)	US
60/365,984	20 March 2002 (20.03.2002)	US
60/365,985	20 March 2002 (20.03.2002)	US
60/365,884	20 March 2002 (20.03.2002)	US
60/366,928	22 March 2002 (22.03.2002)	US
60/372,018	12 April 2002 (12.04.2002)	US
60/372,022	12 April 2002 (12.04.2002)	US
60/374,682	23 April 2002 (23.04.2002)	US
60/388,096	12 June 2002 (12.06.2002)	US
60/389,143	14 June 2002 (14.06.2002)	US
60/391,779	26 June 2002 (26.06.2002)	US
60/403,743	15 August 2002 (15.08.2002)	US
60/410,755	13 September 2002 (13.09.2002)	US
60/412,957	23 September 2002 (23.09.2002)	US
60/420,382	22 October 2002 (22.10.2002)	US
Not furnished	5 March 2003 (05.03.2003)	US

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

US	60/372,022 (CIP)
Filed on	12 April 2002 (12.04.2002)
US	60/365,984 (CIP)
Filed on	20 March 2002 (20.03.2002)
US	60/389,143 (CIP)
Filed on	14 June 2002 (14.06.2002)
US	60/391,779 (CIP)
Filed on	26 June 2002 (26.06.2002)
US	60/410,755 (CIP)
Filed on	13 September 2002 (13.09.2002)
US	60/412,957 (CIP)
Filed on	23 September 2002 (23.09.2002)
US	60/366,928 (CIP)
Filed on	22 March 2002 (22.03.2002)
US	60/388,096 (CIP)
Filed on	12 June 2002 (12.06.2002)
US	60/365,985 (CIP)
Filed on	20 March 2002 (20.03.2002)
US	60/365,884 (CIP)
Filed on	20 March 2002 (20.03.2002)
US	60/361,974 (CIP)
Filed on	6 March 2002 (06.03.2002)
US	10/093,463 (CIP)
Filed on	8 March 2002 (08.03.2002)
US	60/365,884 (CIP)
Filed on	20 March 2002 (20.03.2002)
US	60/372,018 (CIP)
Filed on	12 April 2002 (12.04.2002)
US	60/365,034 (CIP)
Filed on	15 March 2002 (15.03.2002)
US	60/403,743 (CIP)
Filed on	15 August 2002 (15.08.2002)
US	60/420,382 (CIP)
Filed on	22 October 2002 (22.10.2002)
US	60/365,477 (CIP)
Filed on	19 March 2002 (19.03.2002)
US	60/374,682 (CIP)
Filed on	23 April 2002 (23.04.2002)
US	Not furnished (CIP)
Filed on	5 March 2003 (05.03.2003)

(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). **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). **CHAUDHURI, Amitabha** [IN/US]; 99 Harbor Avenue, Madison, CT 06443 (US). **COLMAN, Steven, D.** [US/US]; 24 Andy Lane, Guilford, CT 06437 (US). **EDINGER, Shlomit, R.** [US/US]; 766 Edgewood Avenue, New Haven, CT 06515 (US). **ETTENBERG, Seth** [US/US]; 583 Orange Street, New Haven, CT 06551 (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). **GUO, Xiaojia** [CN/US]; 713 Roberst Frost Drive,

[Continued on next page]

(54) Title: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

(57) Abstract: The present invention relates to novel polypeptides, and the nucleic acids encoding them, having 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.

WO 03/076578 A2

WO 03/076578 A2

Branford, CT 06405 (US). **KEKUDA, Ramesh** [IN/US]; 71 Aiken Street, Unit R3, Norwalk, CT 06851 (US). **LI, Li** [CN/US]; 56 Jerimoth Drive, Branford, CT 06405 (US). **MACLACHLAN, Timothy** [US/US]; 8 Jefferson Street, Unionville, CT 06085 (US). **MALYANKAR, Uriel, M.** [IN/US]; 229 Brandford Road, Number 330, Branford, CT 06405 (US). **MEZES, Peter, S.** [CA/US]; 7 Clark's Lane, Old Lyme, CT 06371 (US). **MILLER, Charles, E.** [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US). **MILLET, Isabelle** [FR/US]; 74 Carrington Avenue, Mildford, CT 06460 (US). **PADIGARU, Muralidhara** [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apt. 1C, Branford, CT 06405 (US). **PEYMAN, John** [US/US]; 336 West Rock Avenue, New Haven, CT 06515 (US). **QIAN, Xiazhong** [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). **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, CN 06437 (US). **SUKUMARAN, Sujatha** [US/US]; 8B Stonegate Circle, Branford, CT 06405 (US). **VERNET, Corine, A., M.** [FR/US]; 1739 Foxon Road, Box L6, Branford, CT 06471 (US). **VOSS, Edward, Z.** [US/US]; 123 Knollwood Drive, Wallingford, CT 06492 (US). **ZHONG, Mei** [CA/US]; 152 Brushy Plain Road, Branford, CT 06405 (US).

(74) **Agent: ELRIFI, Ivro, 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, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, 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, 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.

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 224

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 224

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

FIELD OF THE INVENTION

The present invention relates to novel polypeptides, and the nucleic acids encoding them, having 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.

BACKGROUND OF THE INVENTION

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.

Antibodies are multichain proteins that bind specifically to a given antigen, and bind poorly, or not at all, to substances deemed not to be cognate antigens. Antibodies are comprised of two short chains termed light chains and two long chains termed heavy chains. These chains are constituted of immunoglobulin domains, of which generally there are two classes: one variable domain per chain, one constant domain in light chains, and three or more constant domains in heavy chains. The antigen-specific portion of the immunoglobulin molecules resides in the variable domains; the variable domains of one light chain and one heavy chain associate with each other to generate the antigen-binding moiety. Antibodies that bind immunospecifically to a cognate or target antigen bind with high affinities. Accordingly, they are useful in assaying specifically for the presence of the antigen in a sample. In addition, they have the potential of inactivating the activity of the antigen.

Therefore there is a need to assay for the level of a protein effector of interest in a biological sample from such a subject, and to compare this level with that characteristic of a nonpathological condition. In particular, there is a need for such an assay based on the use of an antibody that binds immunospecifically to the antigen. There further is a need to inhibit the activity of the protein effector in cases where a pathological condition arises from elevated or excessive levels of the effector based on the use of an antibody that binds immunospecifically to the effector. Thus, there is a need for the antibody as a product of manufacture. There further is a need for a method of treatment of a pathological condition brought on by an elevated or excessive level of the protein effector of interest based on administering the antibody to the subject.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of isolated polypeptides including amino acid sequences selected from mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58. The novel nucleic acids and polypeptides are referred to herein as NOV1a, NOV1b, NOV1c, NOV1d, NOV2a, NOV2b, NOV2c, NOV2d, NOV3a, NOV3b, etc. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid or polypeptide sequences.

The invention also is based in part upon variants 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 58, 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. In another embodiment, the invention includes the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58. In another embodiment, the invention also comprises variants of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58 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 involves fragments of any of the mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58, or any other amino acid sequence selected from this group. The invention also comprises fragments from these groups in which up to 15% of the residues are changed.

In another embodiment, the invention encompasses polypeptides that are naturally occurring allelic variants of the sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58. These allelic variants include amino acid sequences that are the translations of nucleic acid sequences differing by a single nucleotide from nucleic acid sequences selected from the group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 58. The variant polypeptide where any amino acid changed in the chosen sequence is changed to provide a conservative substitution.

In another embodiment, the invention comprises a pharmaceutical composition involving a polypeptide with an amino acid sequence selected from the group consisting of

SEQ ID NO:2n, wherein n is an integer between 1 and 58 and a pharmaceutically acceptable carrier. In another embodiment, the invention involves a kit, including, in one or more containers, this pharmaceutical composition.

In another embodiment, the invention includes the use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease being selected from a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58 wherein said therapeutic is the polypeptide selected from this group.

In another embodiment, the invention comprises a method for determining the presence or amount of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58 in a sample, the method involving providing the sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the polypeptide, thereby determining the presence or amount of polypeptide in the sample.

In another embodiment, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58 in a first mammalian subject, the method involving measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in this sample 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 another embodiment, the invention involves a method of identifying an agent that binds to a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58, the method including introducing the polypeptide to the agent; and determining whether the agent binds to the polypeptide. The agent could be a cellular receptor or a downstream effector.

In another embodiment, the invention involves 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 polypeptide with an amino

acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58, the method including providing a cell expressing the polypeptide of the invention 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 embodiment, the invention involves a method for screening for a modulator of activity or of latency or predisposition to a pathology associated with 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 58, the method including administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of the invention, wherein the test animal recombinantly expresses the polypeptide of the invention; measuring the activity of the polypeptide in the test animal after administering the test compound; and comparing the activity of the protein in the test animal with the activity of the polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the 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 polypeptide of the invention. The recombinant test animal could express a test protein transgene or express the transgene under the control of a promoter at an increased level relative to a wild-type test animal The promoter may or may not b the native gene promoter of the transgene.

In another embodiment, the invention involves a method for modulating the activity of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58, the method including introducing a cell sample expressing the 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 involves a method of treating or preventing a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58, the method including administering the polypeptide to a subject in which such treatment or prevention

is desired in an amount sufficient to treat or prevent the pathology in the subject. The subject could be human.

In another embodiment, the invention involves a method of treating a pathological state in a mammal, the method including 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 having the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58 or a biologically active fragment thereof.

In another embodiment, the invention involves an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 58; 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 58 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; the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58; 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 58, 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; 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 58 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; and the complement of any of the nucleic acid molecules.

In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 58, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant.

In another embodiment, the invention involves an isolated nucleic acid molecule including a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 58 that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.

In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 58, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 58.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 58, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 58; 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 58 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; 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 58; and 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 58 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 an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 58, wherein the 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 58, or a complement of the nucleotide sequence.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 58, wherein the nucleic acid molecule has a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen 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 in the chosen coding sequence are so changed, an isolated second polynucleotide that is a complement of the first polynucleotide, or a fragment of any of them.

In another embodiment, the invention includes a vector involving the nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 58. This vector can have a promoter operably linked to the nucleic acid molecule. This vector can be located within a cell.

In another embodiment, the invention involves a method for determining the presence or amount of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 58 in a sample, the method including 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 nucleic acid molecule, thereby determining the presence or amount of the nucleic acid molecule in the sample. The presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type. The cell type can be cancerous.

In another embodiment, the invention involves a method for determining the presence of or predisposition for a disease associated with altered levels of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 58 in a first mammalian

subject, the method including measuring the amount of the 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 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 the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

The invention further provides an antibody that binds immunospecifically to a NOVX polypeptide. The NOVX antibody may be monoclonal, humanized, or a fully human antibody. Preferably, the antibody has a dissociation constant for the binding of the NOVX polypeptide to the antibody less than 1 x 10^{-9} M. More preferably, the NOVX antibody neutralizes the activity of the NOVX polypeptide.

In a further aspect, the invention provides for the use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, associated with a NOVX polypeptide. Preferably the therapeutic is a NOVX antibody.

In yet a further aspect, the invention provides a method of treating or preventing a NOVX-associated disorder, a method of treating a pathological state in a mammal, and a method of treating or preventing a pathology associated with a polypeptide by administering a NOVX antibody to a subject in an amount sufficient to treat or prevent the disorder.

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 are 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 DRAWINGS

FIG. 1 is a histogram illustrating the ability of the EGF repeat domain of the Ten-M4-like CG50301-02 polypeptide to inhibit 786-0 cell migration in a dose dependent manner. For further details see Example E.

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	CG110853-02	1	2	Testis-specific protein TPX-1 precursor (Cysteine-rich secretory protein-2) (CRISP-2) - Homo sapiens
NOV1b	CG110853-01	3	4	Testis-specific protein TPX-1 precursor (Cysteine-rich secretory protein-2) (CRISP-2) - Homo sapiens
NOV1c	209934326	5	6	Testis-specific protein TPX-1 precursor (Cysteine-rich secretory protein-2) (CRISP-2) - Homo sapiens
NOV1d	CG110853-03	7	8	Testis-specific protein TPX-1 precursor (Cysteine-rich secretory protein-2) (CRISP-2) - Homo sapiens
NOV2a	CG113367-02	9	10	Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin) (PRO1400) - Homo sapiens
NOV2b	CG113367-03	11	12	Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin) (PRO1400) - Homo sapiens

NOV2c	CG113367-01	13	14	Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin) (PRO1400) - Homo sapiens
NOV3a	CG132364-01	15	16	Programmed cell death 1 - Homo sapiens
NOV3b	CG132364-02	17	18	Programmed cell death 1 - Homo sapiens
NOV4a	CG161497-01	19	20	Cytotoxic T-lymphocyte-associated protein 4 - Homo sapiens
NOV4b	CG161497-02	21	22	Cytotoxic T-lymphocyte-associated protein 4 - Homo sapiens
NOV5a	CG169401-01	23	24	Activation-inducible lymphocyte IMMUNOMEDIATORY molecule AILIM precursor (Inducible T-cell co-stimulator) (Inducible COSTIMULATOR precursor) - Homo sapiens
NOV5b	CG169401-02	25	26	Activation-inducible lymphocyte IMMUNOMEDIATORY molecule AILIM precursor (Inducible T-cell co-stimulator) (Inducible COSTIMULATOR precursor) - Homo sapiens
NOV6a	CG176697-02	27	28	Hypothetical protein FLJ38482 - Homo sapiens
NOV6b	CG176697-01	29	30	Hypothetical protein FLJ38482 - Homo sapiens
NOV6c	316784436	31	32	Hypothetical protein FLJ38482 - Homo sapiens
NOV7a	CG178810-03	33	34	Poliovirus receptor - Homo sapiens
NOV7b	CG178810-01	35	36	Poliovirus receptor - Homo sapiens
NOV7c	CG178810-02	37	38	Poliovirus receptor - Homo sapiens
NOV8a	CG179299-01	39	40	Hypothetical protein - Homo sapiens (Human), 1500 aa
NOV9a	CG50301-07	41	42	Ten-m4 - Mus musculus
NOV9b	CG50301-01	43	44	Ten-m4 - Mus musculus
NOV9c	172885447	45	46	Ten-m4 - Mus musculus
NOV9d	312712785	47	48	Ten-m4 - Mus musculus
NOV9e	311748770	49	50	Ten-m4 - Mus musculus
NOV9f	311748779	51	52	Ten-m4 - Mus musculus
NOV9g	CG50301-02	53	54	Ten-m4 - Mus musculus
NOV9h	CG50301-03	55	56	Ten-m4 - Mus musculus
NOV9i	CG50301-04	57	58	Ten-m4 - Mus musculus
NOV9j	CG50301-05	59	60	Ten-m4 - Mus musculus
NOV9k	CG50301-06	61	62	Ten-m4 - Mus musculus
NOV10a	CG56201-03	63	64	Similar to fibromodulin - Homo sapiens
NOV10b	CG56201-04	65	66	Similar to fibromodulin - Homo sapiens
NOV10c	247618257	67	68	Similar to fibromodulin - Homo sapiens
NOV10d	CG56201-01	69	70	Similar to fibromodulin - Homo sapiens
NOV10e	CG56201-02	71	72	Similar to fibromodulin - Homo sapiens

NOV11a	CG56653-11	73	74	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11b	CG56653-01	75	76	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11c	CG56653-05	77	78	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11d	CG56653-07	79	80	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11e	CG56653-06	81	82	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11f	247675490	83	84	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11g	247675499	85	86	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11h	277765267	87	88	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11i	277580705	89	90	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11j	275623919	91	92	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11k	274056295	93	94	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11l	268951998	95	96	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11m	CG56653-02	97	98	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11n	CG56653-03	99	100	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11o	CG56653-04	101	102	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11p	CG56653-08	103	104	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens

NOV11q	CG56653-09	105	106	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11r	CG56653-10	107	108	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV11s	CG56653-12	109	110	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens
NOV12a	CG59713-03	111	112	Four-transmembrane protein associating with dishevelled (VANGL1) - Homo sapiens
NOV12b	CG59713-01	113	114	Four-transmembrane protein associating with dishevelled (VANGL1) - Homo sapiens
NOV12c	CG59713-02	115	116	Four-transmembrane protein associating with dishevelled (VANGL1) - Homo sapiens

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), vascular calcification, fibrosis, 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, osteoarthritis, rheumatoid arthritis, osteochondrodysplasia, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, glomerulonephritis, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, psoriasis, skin disorders, graft versus host disease, AIDS, bronchial asthma, lupus, Crohn's disease; inflammatory bowel disease, ulcerative colitis, 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, schizophrenia, depression, asthma, emphysema, allergies, the metabolic

syndrome X and wasting disorders associated with chronic diseases and various cancers, as well as conditions such as transplantation, neuroprotection, fertility, or regeneration (*in vitro* and *in vivo*).

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.

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 58; (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 58, 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 58; (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 58 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 58; (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 58 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 58; (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 58, 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 58 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 58; (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 58 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 58; 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 58 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 58, 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 58, 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 58, 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 58, 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 58, 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 58, 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 58, 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 58, 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 58; or an anti-sense strand nucleotide

sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 58; or of a naturally occurring mutant of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 58.

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 58, 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 58, 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 58. 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 58.

In addition to the human NOVX nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 58, 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 58, 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 58. 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:2*n*-1, wherein *n* is an integer between 1 and 58, 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 58, 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 58, 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 58, 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 58. 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 predicted to be particularly non-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 58, 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 58. 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 58; more preferably at least about 70% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 58; still more preferably at least about 80% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 58; even more preferably at least about 90% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 58; and most preferably at least about 95% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 58.

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 58, 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 58, 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 58, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, 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 predicted 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 58, 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 2 X 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 58, 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 58, 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 58, 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 which 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 58). 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 58. 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 58, 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 58) 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 58. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 58, and retains the functional activity of the protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 58, 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 58, and retains the functional activity of the NOVX proteins of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 58.

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 58.

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 58, 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 58, 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 ≤ 100 nM, more preferably ≤ 10 nM, and most preferably ≤ 100 pM to about 1 pM, as measured by assays including radioligand binding assays or similar assays known to skilled artisans.

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 which 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, *e.g.*, 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 steps. 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 which 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 which 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 which 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, expression vectors of utility 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 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 58, 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 58), 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 58, 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 which 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 which 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 which 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 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 58, 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 58, 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 58, 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. NOV 1: CG110853, Testis Specific Protein TPX-1 like.

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

Table 1A. NOV1 Sequence Analysis

	SEQ ID NO: 1	799 bp	
NOV1a, CG110853-02 DNA Sequence	AGATATTTCATCCTGCTCAGAAAACCAACATTTCCAGCA**ATG**GCTTTACTACCGGTGTTGTTTC TGGTTACTGTGCTGCCTCCATCTTTACCTGCAGAAGGAAAGGATCCCGCTTTTACTGCTTTGTT AACCACCCAGTTGCAAGTGCAAAGGGAGATTGTAAATAAACACAATGAACTAAGGAAAGCAGTC TCTCCACCTGCCAGCAACATGCTAAAGATGGAATGGAGCAGAGAGGTAACAACGAATGCCCAAA GGTGGGCAAACAAGTGCACTTTACAACATAGTGATCCAGAGGACCGCAAAACCAGTACAAGATG TGGTGAGAATCTCTATATGTCAAGTGACCCTACTTCCTGGTCTTCTGCAATCCAAAGCTGGTAT GACGAGATCCTAGATTTTGTCTATGGTGTAGGACCAAAGAGTCCCAATGCAGTTGTTGGACATT		

	ATACTCAGCTTGTTTGGTACTCGACTTACCAGGTAGGCTGTGGAATTGCCTACTGTCCCAATCA AGATAGTCTAAAATACTACTATGTTTGCCAATATTGTCCTGCTGGTAATAATATGAATAGAAAG AATACCCCGTACCAACAAGGAACACCTTGTGCCGGTTGCCCTGATGACTGTGACAAAGGACTAT GCACCAATAGTTGCCAGTATCAAGATCTCCTAAGTAACTGTGATTCCTTGAAGAATACAGCTGG CTGTGAACATGAGTTACTCAAGGAAAAGTGCAAGGCTACTTGCCTATGTGAGAACAAAATTTAC **TGA**TTTACCTAGTGAGCATTGTGCAAGACTG		
	ORF Start: ATG at 40		ORF Stop: TGA at 769
	SEQ ID NO: 2	243 aa	MW at 27242.6kD
NOV1a, CG110853-02 Protein Sequence	MALLPVLFLVTVLPPSLPAEGKDPAFTALLTTQLQVQREIVNKHNELRKAVSPPASNMLKMEWS REVTTNAQRWANKCTLQHSDPEDRKTSTRCGENLYMSSDPTSWSSAIQSWYDEILDFVYGVGPK SPNAVVGHYTQLVWYSTYQVGCGIAYCPNQDSLKYYYVCQYCPAGNNMNRKNTPYQQGTPCAGC PDDCDKGLCTNSCQYQDLLSNCDSLKNTAGCEHELLKEKCKATCLCENKIY		
	SEQ ID NO: 3	690 bp	
NOV1b, CG110853-01 DNA Sequence	TTCCAGCA**ATG**GCTTTACTACCGGTGTTGTTTCTGGTTACTGTGCTGCCTCCATCTTTACCTGC AGAAGGAAAGGATCCCGCTTTTACTGCTTTGTTAACCACCCAGTTGCAAGTGCAAAGGGAGATT GTAAATAAACACAATGAACTAAGGAAAGCAGTCTCTCCACCTGCCAGTAACATGCTAAAGATGG AATGGAGCAGAGAGGTAACAACGAATGCCCAAAGGTGGGCAAACAAGTGCACTTTACAACATAG TGATCCAGAGGACCGCAAAACCAGTACAAGATGTGGTGAGAATCTCTATATGTCAAGTGACCCT ACTTCCTGGTCTTCTGCAATCCAAAGCTGGTATGACGAGATCCTAGATTTTGTCTATGGTGTAG GACCAAAGAGTCCCAATCAAGATAGTCTAAAATACTACTATGTTTGCCAATATTGTCCTGCTGG TAATAATATGAATAGAAAGAATACCCCGTACCAACAAGGAACACCTTGTGCCGGTTGCCCTGAT GACTGTGACAAAGGACTATGCACCAATAGTTGCCAGTATCAAGATCTCCTAAGTAACTGTGATT CCTTGAAGAATACAGCTGGCTGTGAACATGAGTTACTCAAGGAAAAGTGCAAGGCTACTTGCCT ATGTGAGAACAAAATTTAC**TGA**TTTACCTAGTGAGCATTGTGCAAGACTG		
	ORF Start: ATG at 9		ORF Stop: TGA at 660
	SEQ ID NO: 4	217 aa	MW at 24367.3kD
NOV1b, CG110853-01 Protein Sequence	MALLPVLFLVTVLPPSLPAEGKDPAFTALLTTQLQVQREIVNKHNELRKAVSPPASNMLKMEWS REVTTNAQRWANKCTLQHSDPEDRKTSTRCGENLYMSSDPTSWSSAIQSWYDEILDFVYGVGPK SPNQDSLKYYYVCQYCPAGNNMNRKNTPYQQGTPCAGCPDDCDKGLCTNSCQYQDLLSNCDSLK NTAGCEHELLKEKCKATCLCENKIY		
	SEQ ID NO: 5	609 bp	
NOV1c, 209934326 DNA Sequence	**AAG**CTTGCAGAAGGAAAGGATCCCGCTTTTACTGCTTTGTTAACCACCCAGTTGCAAGTGCAAAG GGAGATTGTAAATAAACACAATGAACTAAGGAAAGCAGTCTCTCCACCTGCCAGTAACATGCTAA AGATGGAATGGAGCAGAGAGGTAACAACGAATGCCCAAAGGTGGGCAAACAAGTGCACTTTACAA CATAGTGATCCAGAGGACCGCAAAACCAGTACAAGATGTGGTGAGAATCTCTATATGTCAAGTGA CCCTACTTCCTGGTCTTCTGCAATCCAAAGCTGGTATGACGAGATCCTAGATTTTGTCTATGGTG TAGGACCAAAGAGTCCCAATCAAGATAGTCTAAAATACTACTATGTTTGCCAATATTGTCCTGCT GGTAATAATATGAATAGAAAGAATACCCCGTACCAACAAGGAACACCTTGTGCCGGTTGCCCTGA TGACTGTGACAAAGGACTATGCACCAATAGTTGCCAGTATCAAGATCTCCTAAGTAACTGTGATT CCTTGAAGAATACAGCTGGCTGTGAACATGAGTTACTCAAGGAAAAGTGCAAGGCTACTTGCCTA TGTGAGAACAAAATTTACCTCGAG		
	ORF Start: at 1		ORF Stop: end of sequence
	SEQ ID NO: 6	203 aa	MW at 22948.5kD
NOV1c, 209934326 Protein Sequence	KLAEGKDPAFTALLTTQLQVQREIVNKHNELRKAVSPPASNMLKMEWSREVTTNAQRWANKCTL QHSDPEDRKTSTRCGENLYMSSDPTSWSSAIQSWYDEILDFVYGVGPKSPNQDSLKYYYVCQYC PAGNNMNRKNTPYQQGTPCAGCPDDCDKGLCTNSCQYQDLLSNCDSLKNTAGCEHELLKEKCKA TCLCENKIYLE		
	SEQ ID NO: 7	609 bp	
NOV1d, CG110853-03 DNA Sequence	AAGCTT**GCA**GAAGGAAAGGATCCCGCTTTTACTGCTTTGTTAACCACCCAGTTGCAAGTGCAAA GGGAGATTGTAAATAAACACAATGAACTAAGGAAAGCAGTCTCTCCACCTGCCAGTAACATGCT AAAGATGGAATGGAGCAGAGAGGTAACAACGAATGCCCAAAGGTGGGCAAACAAGTGCACTTTA CAACATAGTGATCCAGAGGACCGCAAAACCAGTACAAGATGTGGTGAGAATCTCTATATGTCAA GTGACCCTACTTCCTGGTCTTCTGCAATCCAAAGCTGGTATGACGAGATCCTAGATTTTGTCTA TGGTGTAGGACCAAAGAGTCCCAATCAAGATAGTCTAAAATACTACTATGTTTGCCAATATTGT CCTGCTGGTAATAATATGAATAGAAAGAATACCCCGTACCAACAAGGAACACCTTGTGCCGGTT GCCCTGATGACTGTGACAAAGGACTATGCACCAATAGTTGCCAGTATCAAGATCTCCTAAGTAA CTGTGATTCCTTGAAGAATACAGCTGGCTGTGAACATGAGTTACTCAAGGAAAAGTGCAAGGCT ACTTGCCTATGTGAGAACAAAATTTAC**CTC**GAG		
	ORF Start: at 7		ORF Stop: at 604

	SEQ ID NO: 8	199 aa	MW at 22464.9kD
NOV1d, CG110853-03 Protein Sequence	AEGKDPAFTALLTTQLQVQREIVNKHNELRKAVSPPASNMLKMEWSREVTTNAQRWANKCTLQH SDPEDRKTSTRCGENLYMSSDPTSWSSAIQSWYDEILDFVYGVGPKSPNQDSLKYYYVCQYCPA GNNMNRKNTPYQQGTPCAGCPDDCDKGLCTNSCQYQDLLSNCDSLKNTAGCEHELLKEKCKATC LCENKIY		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 1B.

Table 1B. Comparison of NOV1a against NOV1b through NOV1d.

Protein Sequence	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV1b	1..243 1..217	217/243 (89%) 217/243 (89%)
NOV1c	19..243 3..201	199/225 (88%) 199/225 (88%)
NOV1d	19..243 1..199	199/225 (88%) 199/225 (88%)

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 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 19; peak value 11.04 PSG score: 6.64 GvH: von Heijne's method for signal seq. recognition GvH score (threshold: -2.1): 1.71 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 = 7.53 (at 118) ALOM score: 7.53 (number of TMSs: 0) MTOP: Prediction of membrane topology (Hartmann et al.) Center position for calculation: 9 Charge difference: -2.0 C(-1.0) - N(1.0) N >= C: N-terminal side will be inside MITDISC: discrimination of mitochondrial targeting seq R content: 0 Hyd Moment(75): 3.05 Hyd Moment(95): 1.96 G content: 0 D/E content: 1 S/T content: 2 Score: -5.30

```
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:
      KKXX-like motif in the C-terminus: ENKI

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: 94.1

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

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

Final Results (k = 9/23):

       66.7 %: extracellular, including cell wall
       11.1 %: cytoplasmic
       11.1 %: mitochondrial
       11.1 %: nuclear

>> prediction for CG110853-02 is exc (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
ABG06656	Novel human diagnostic protein #6647 - Homo sapiens, 257 aa. [WO200175067-A2, 11-OCT-2001]	1..243 15..257	242/243 (99%) 242/243 (99%)	e-149
AAE13072	Homo sapiens (Hs)-Tpx protein - Homo sapiens, 243 aa. [WO200174385-A1, 11-OCT-2001]	1..243 1..243	241/243 (99%) 242/243 (99%)	e-148
AAE02211	Human full-length 36P1G3/SGP28 protein - Homo sapiens, 258 aa. [WO200131343-A2, 03-MAY-2001]	1..243 14..258	175/245 (71%) 200/245 (81%)	e-106
AAM24000	Human EST encoded protein SEQ ID NO: 1525 - Homo sapiens, 245 aa. [WO200154477-A2, 02-AUG-2001]	1..243 1..245	174/245 (71%) 199/245 (81%)	e-105
AAM23992	Human EST encoded protein SEQ ID NO: 1517 - Homo sapiens, 245 aa. [WO200154477-A2, 02-AUG-2001]	1..243 1..245	174/245 (71%) 199/245 (81%)	e-105

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
P16562	Testis-specific protein TPX-1 precursor (Cysteine-rich secretory protein-2) (CRISP-2) - Homo sapiens (Human), 243 aa.	1..243 1..243	242/243 (99%) 242/243 (99%)	e-148
CAD31226	Testis specific protein 1 - Equus caballus (Horse), 244 aa.	1..243 1..244	204/244 (83%) 220/244 (89%)	e-127
Q60477	Testis-specific protein TPX-1 precursor (Autoantigen 1) (25 kDa acrosomal autoantigen) (AA1) - Cavia porcellus (Guinea pig), 244 aa.	1..243 1..244	191/244 (78%) 217/244 (88%)	e-119

CAD31227	Acidic epididymal glycoprotein 2 - Equus caballus (Horse), 245 aa.	1..243 1..245	178/245 (72%) 203/245 (82%)	e-109
O19010	Cysteine-rich secretory protein-3 precursor (CRISP-3) - Equus caballus (Horse), 245 aa.	1..243 1..245	177/245 (72%) 202/245 (82%)	e-108

PFam analysis predicts 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
SCP	41..168	49/163 (30%) 107/163 (66%)	7.9e-52
ShTK	204..239	12/44 (27%) 25/44 (57%)	0.75

Example 2. NOV 2: CG113367: Serotransferrin precursor.

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

Table 2A. NOV2 Sequence Analysis

	SEQ ID NO: 9	2185 bp	
NOV2a, CG113367-02 DNA Sequence	CGACTGTGCTCGCTGCTCAGCGCCGCACCCGGAGG**ATG**AGGCTCGCCGTGGGAGCCCTGCTGGT CTGCGCCGTCCTGGGGCTGTGTCTGGCTGTCCCTGATAAAACTGTGAGATGGTGTGCAGTGTCG GAGCATGAGGCCACTAAGTGCCAGAGTTTCCGCGACCATATGAAAAGCGTCATTCCATCCGATG GTCCCAGTGTTGCTTGTGTGAAGAAAGCCTCCTACCTTGATTGCATCAGGGCCATTGCGGCAAA CGAAGCGGATGCTGTGACACTGGATGCAGGTTTGGTGTATGATGCTTACCTGGCTCCCAATAAC CTGAAGCCTGTGGTGGCAGAGTTCTATGGGTCAAAAGAGGATCCACAGACTTTCTATTATGCTG TTGCTGTGGTGAAGAAGGATAGTGGCTTCCAGATGAACCAGCTTCGAGGCAAGAAGTCCTGCCA CACGGGTCTAGGCAGGTCCGCTGGGTGGAACATCCCCATAGGCTTACTTTACTGTGACTTACCT GAGCCACGTAAACCTCTTGAGAAAGCAGTGGCCAATTTCTTCTCGGGCAGCTGTGCCCCTTGTG CGGATGGGACGGACTTCCCCCAGCTGTGTCAACTGTGTCCAGGGTGTGGCTGCTCCACCCTTAA CCAATACTTCGGCTACTCGGGAGCCTTCAAGTGTCTGAAGGATGGTGCTGGGGATGTGGCCTTT GTCAAGCACTCGACTATATTTGAGAACTTGGCAAACAAGGCTGACAGGGACCAGTATGAGCTGC TTTGCCTGGACAACACCCGGAAGCCGGTAGATGAATACAAGGACTGCCACTTGGCCCAGGTCCC TTCTCATACCGTCGTGGCCCGAAGTATGGGCGGCAAGGAGGACTTGATCTGGGAGCTTCTCAAC CAGGCCCAGGAACATTTTGGCAAAGACAAATCAAAAGAATTCCAACTATTCAGCTCTCCTCATG GGAAGGACCTGCTGTTTAAGGACTCTGCCCACGGGTTTTTAAAAGTCCCCCCCAGGATGGATGC CAAGATGTACCTGGGCTATGAGTATGTCACTGCCATCCGGAATCTACGGGAAGGCACATGCCCA GAAGCCCCAACAGATGAATGCAAGCCTGTGAAGTGGTGTGCGCTGAGCCACCACGAGAGGCTCA AGTGTGATGAGTGGAGTGTTAACAGTGTAGGGAAAATAGAGTGTGTATCAGCAGAGACCACCGA AGACTGCATCGCCAAGATCATGAATGGAGAAGCTGATGCCATGAGCTTGGATGGAGGGTTTGTC TACATAGCGGGCAAGTGTGGTCTGGTGCCTGTCTTGGCAGAAAACTACAATAAGAGCGATAATT GTGAGGATACACCAGAGGCAGGGTATTTTGCTGTAGCAGTGGTGAAGAAATCAGCTTCTGACCT CACCTGGGACAATCTGAAAGGCAAGAAGTCCTGCCATACGGCAGTTGGCAGAACCGCTGGCTGG AACATCCCCATGGGCCTGCTCTACAATAAGATCAACCACTGCAGATTTGATGAATTTTTCAGTG		

	AAGGTTGTGCCCCTGGGTCTAAGAAAGACTCCAGTCTCTGTAAGCTGTGTATGGGCTCAGGCCT AAACCTGTGTGAACCCAACAACAAAGAGGGATACTACGGCTACACAGGCGCTTTCAGGTGTCTG GTTGAGAAGGGAGATGTGGCCTTTGTGAAACACCAGACTGTCCCACAGAACACTGGGGGAAAAA ACCCTGATCCATGGGCTAAGAATCTGAATGAAAAAGACTATGAGTTGCTGTGCCTTGATGGTAC CAGGAAACCTGTGGAGGAGTATGCGAACTGCCACCTGGCCAGAGCCCCGAATCACGCTGTGGTC ACACGGAAAGATAAGGAAGCTTGCGTCCACAAGATATTACGTCAACAGCAGCACCTATTTGGAA GCAACGTAACTGACTGCTCGGGCAACTTTTGTTTGTTCCGGTCGGAAACCAAGGACCTTCTGTT CAGAGATGACACAGTATGTTTGGCCAAACTTCATGACAGAAACACATATGAAAAATACTTAGGA GAAGAATATGTCAAGGCTGTTGGTAACCTGAGAAAATGCTCCACCTCATCACTCCTGGAAGCCT GCACTTTCCGTAGACCT**TAA**AATCTCAGAGGTAGGGCTCCCACCTCTCCATCATAAGGGCGAAT TCCAGCACA
	ORF Start: ATG at 36 | | ORF Stop: TAA at 2130
	SEQ ID NO: 10 | 698 aa | MW at 77049.0kD
NOV2a, CG113367-02 Protein Sequence	MRLAVGALLVCAVLGLCLAVPDKTVRWCAVSEHEATKCQSFRDHMKSVIPSDGPSVACVKKASY LDCIRAIAANEADAVTLDAGLVYDAYLAPNNLKPVVAEFYGSKEDPQTFYYAVAVVKKDSGFQM NQLRGKKSCHTGLGRSAGWNIPIGLLYCDLPEPRKPLEKAVANFFSGSCAPCADGTDFPQLCQL CPGCGCSTLNQYFGYSGAFKCLKDGAGDVAFVKHSTIFENLANKADRDQYELLCLDNTRKPVDE YKDCHLAQVPSHTVVARSMGGKEDLIWELLNQAQEHFGKDKSKEFQLFSSPHGKDLLFKDSAHG FLKVPPRMDAKMYLGYEYVTAIRNLREGTCPEAPTDECKPVKWCALSHHERLKCDEWSVNSVGK IECVSAETTEDCIAKIMNGEADAMSLDGGFVYIAGKCGLVPVLAENYNKSDNCEDTPEAGYFAV AVVKKSASDLTWDNLKGKKSCHTAVGRTAGWNIPMGLLYNKINHCRFDEFFSEGCAPGSKKDSS LCKLCMGSGLNLCEPNNKEGYYGYTGAFRCLVEKGDVAFVKHQTVPQNTGGKNPDPWAKNLNEK DYELLCLDGTRKPVEEYANCHLARAPNHAVVTRKDKEACVHKILRQQQHLFGSNVTDCSGNFCL FRSETKDLLFRDDTVCLAKLHDRNTYEKYLGEEYVKAVGNLRKCSTSSLLEACTFRRP
	SEQ ID NO: 11 | 773 bp |
NOV2b, CG113367-03 DNA Sequence	GAGCGGAACAGTGGGACGAGTAAGGAAGGGGGGTTGGGAGAGGGGCGATTGGGCAACCCGGCTG CACAAACACGGGAGGTCAAAGATTGCGCCCAGCCCGCCCAGGCCGGGAATGGAATAAAGGGACG CGGGGCGCCGGAGGCTGCACAGAAGCGAGTCCGACTGTGCTCGCTGCTCAGCGCCGCACCCGGA AG**ATG**AGGCTCGCCGTGGGAGCCCTGCTGGTCTGCGCCGTCCTGGGGCTGTGTCTGGCTGTCCC TGATAAAACTGTGAGATGGTGTGCAGTGTCGGAGCATGAGGCCACTAAGTGCCAGACCCCGAAT CACGCTGTGGTCACACGGAAAGATAAGGAAGCTTGCGTCCACAAGATATTACGTCAACAGCAGC ACCTATTTGGAAGCAACGTAACTGACTGCTCGGGCAACTTTTGTTTGTTCCGGTCGGAAACCAA GGACCTTCTGTTCAGAGATGACACAGTATGTTTGGCCAAACTTCATGACAGAAACACATATGAA AAATACTTAGGAGAAGAATATGTCAAGGCTGTTGGTAACCTGAGAAAATGCTCCACCTCATCAC TCCTGGAAGCCTGCACTTTCCGTAGACCT**TAA**AATCTCAGAGGTAGGGCTGCCACCAAGGTGAA GATGGGAACGCAGATGATCCATGAGTTTGCCCTGGTTTCACTGGCCCAAGTGGTTTGTGCTAAC CACGTCTGTCTTCACAGCTCTGTGTTGCCATGTGTGCTGAACAAAAAATTAAAATTATTATTGA TTCAA
	ORF Start: ATG at 195 | | ORF Stop: TAA at 606
	SEQ ID NO: 12 | 137 aa | MW at 15404.7kD
NOV2b, CG113367-03 Protein Sequence	MRLAVGALLVCAVLGLCLAVPDKTVRWCAVSEHEATKCQTPNHAVVTRKDKEACVHKILRQQQH LFGSNVTDCSGNFCLFRSETKDLLFRDDTVCLAKLHDRNTYEKYLGEEYVKAVGNLRKCSTSSL LEACTFRRP
	SEQ ID NO: 13 | 2347 bp |
NOV2c, CG113367-01 DNA Sequence	ACGGCGAGCCTCATCCTCCGGGTGCGGCGCTGAGCAGCGAGTCNCGACTGTGCTCGCTGCTCAG CGCCGCACCCGGAGG**ATG**AGGCTCGCCGTGGGAGCCCTGCTGGTCTGCGCCGTCCTGGGGCTGT GTCTGGCTGTCCCTGATAAAACTGTGAGATGGTGTGCAGTGTCGGAGCATGAGGCCACTAAGTG CCAGAGTTTCCGCGACCATATGAAAAGCGTCATTCCATCCGATGGTCCCAGTGTTGCTTGTGTG AAGAAAGCCTCCTACCTTGATTGCATCAGGGCCATTGCGGCAAACGAAGCGGATGCTGTGACAC TGGATGCAGGTTTGGTGTATGATGCTTACCTGGCTCCCAATAACCTGAAGCCTGTGGTGGCAGA GTTCTATGGGTCAAAAGAGGATCCACAGACTTTCTATTATGCTGTTGCTGTGGTGAAGAAGGAT AGTGGCTTCCAGATGAACCAGCTTCGAGGCAAGAAGTCCTGCCACACGGGTCTAGGCAGGTCCG CTGGGTGGAACATCCCCATAGGCTTACTTTACTGTGACTTACCTGAGCCACGTAAACCTCTTGA GAAAGCAGTGGCCAATTTCTTCTCGGGCAGCTGTGCCCCTTGTGCGGATGGGACGGACTTCCCC CAGCTGTGTCAACTGTGTCCAGGGTGTGGCTGCTCCACCCTTAACCAATACTTCGGCTACTCAG GAGCCTTCAAGTGTCTGAAGGATGGTGCTGGGGATGTGGCCTTTGTCAAGCACTCGACTATATT TGAGAACTTGGCAAACAAGGCTGACAGGGACCAGTATGAGCTGCTTTGCCTGGACAACACCCGG AAGCCGGTAGATGAATACAAGGACTGCCACTTGGCCCAGGTCCCTTCTCATACCGTCGTGGCCC GAAGTATGGGCGGCAAGGAGGACTTGATCTGGGAGCTTCTCAACCAGGCCCAGGAACATTTTGG CAAAGACAAATCAAAAGAATTCCAACTATTCAGCTCTCCTCATGGGAAGGACCTGCTGTTTAAG GACTCTGCCCACGGGTTTTTAAAAGTCCCCCCCAGGATGGATGCCAAGATGTACCTGGGCTATG AGTATGTCACTGCCATCCGGAATCTACGGGAAGGCACATGCCCAGAAGCCCCAACAGATGAATG

	CAAGCCTGTGAAGTGGTGTGCGCTGAGCCACCACGAGAGGCTCAAGTGTGATGAGTGGAGTGTT AACAGTGTAGGGAAAATAGAGTGTGTATCAGCAGAGACCACCGAAGACTGCATCGCCAAGATCA TGAATGGAGAAGCTGATGCCATGAGCTTGGATGGAGGGTTTGTCTACATAGCGGGCAAGTGTGG TCTGGTGCCTGTCTTGGCAGAAAACTACAATAAGAGCGATAATTGTGAGGATACACCAGAGGCA GGGTATTTTGCTGTAGCAGTGGTGAAGAAATCAGCTTCTGACCTCACCTGGGACAATCTGAAAG GCAAGAAGTCCTGCCATACGGCAGTTGGCAGAACCGCTGGCTGGAACATCCCCATGGGCCTGCT CTACAATAAGATCAACCACTGCAGATTTGATGAATTTTTCAGTGAAGGTTGTGCCCCTGGGTCT AAGAAAGACTCCAGTCTCTGTAAGCTGTGTATGGGCTCAGGCCTAAACCTGTGTGAACCCAACA ACAAAGAGGGATACTACGGCTACACAGGCGCTTTCAGGTGTCTGGTTGAGAAGGGAGATGTGGC CTTTGTGAAACACCAGACTGTCCCACAGAACACTGGGGGAAAAAACCCTGATCCATGGGCTAAG AATCTGAATGAAAAAGACTATGAGTTGCTGTGCCTTGATGGTACCAGGAAACCTGTGGAGGAGT ATGCGAACTGCCACCTGGCCAGAGCCCCGAATCACGCTGTGGTCACACGGAAAGATAAGGAAGC TTGCGTCCACAAGATATTACGTCAACAGCAGCACCTATTTGGAAGCAACGTAACTGACTGCTCG GGCAACTTTTGTTTGTTCCGGTCGGAAACCAAGGACCTTCTGTTCAGAGATGACACAGTATGTT TGGCCAAACTTCATGACAGAAACACATATGAAAAATACTTAGGAGAAGAATATGTCAAGGCTGT TGGTAACCTGAGAAAATGCTCCACCTCATCACTCCTGGAAGCCTGCACTTTCCGTAGACCT**TAA** AATCTCAGAGGTAGGGCTGCCACCAAGGTGAAGATGGGAACGCAGATGATCCATGAGTTTGCCC TGGTTTCACTGGCCCAAGTGGTTTGTGCTAACCACGTCTGTCTTCACAGCTCTGTGTTGCCATG TGTGCTGAACAAAAAATAAAAATTATTATTGATTTTAAAAAAA
	ORF Start: ATG at 80 | | ORF Stop: TAA at 2174
	SEQ ID NO: 14 | 698 aa | MW at 77049.0kD
NOV2c, CG113367-01 Protein Sequence	MRLAVGALLVCAVLGLCLAVPDKTVRWCAVSEHEATKCQSFRDHMKSVIPSDGPSVACVKKASY LDCIRAIAANEADAVTLDAGLVYDAYLAPNNLKPVVAEFYGSKEDPQTFYYAVAVVKKDSGFQM NQLRGKKSCHTGLGRSAGWNIPIGLLYCDLPEPRKPLEKAVANFFSGSCAPCADGTDFPQLCQL CPGCGCSTLNQYFGYSGAFKCLKDGAGDVAFVKHSTIFENLANKADRDQYELLCLDNTRKPVDE YKDCHLAQVPSHTVVARSMGGKEDLIWELLNQAQEHFGKDKSKEFQLFSSPHGKDLLFKDSAHG FLKVPPRMDAKMYLGYEYVTAIRNLREGTCPEAPTDECKPVKWCALSHHERLKCDEWSVNSVGK IECVSAETTEDCIAKIMNGEADAMSLDGGFVYIAGKCGLVPVLAENYNKSDNCEDTPEAGYFAV AVVKKSASDLTWDNLKGKKSCHTAVGRTAGWNIPMGLLYNKINHCRFDEFFSEGCAPGSKKDSS LCKLCMGSGLNLCEPNNKEGYYGYTGAFRCLVEKGDVAFVKHQTVPQNTGGKNPDPWAKNLNEK DYELLCLDGTRKPVEEYANCHLARAPNHAVVTRKDKEACVHKILRQQQHLFGSNVTDCSGNFCL FRSETKDLLFRDDTVCLAKLHDRNTYEKYLGEEYVKAVGNLRKCSTSSLLEACTFRRP

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 2B.

Table 2B. Comparison of NOV2a against NOV2b and NOV2c.

Protein Sequence	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV2b	600..698 39..137	97/99 (97%) 98/99 (98%)
NOV2c	1..698 1..698	698/698 (100%) 698/698 (100%)

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 20 and 21
PSORT II analysis:	PSG: a new signal peptide prediction method N-region: length 2; pos.chg 1; neg.chg 0 H-region: length 19; peak value 9.84 PSG score: 5.44

```
GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):   4.39
      possible cleavage site: between 19 and 20

>>> Seems to have a cleavable signal peptide (1 to 19)

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 20
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  0.90 (at 412)
      ALOM score:   0.90  (number of TMSs: 0)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 9
      Charge difference: -2.5   C(-0.5) - N( 2.0)
      N >= C: N-terminal side will be inside

MITDISC: discrimination of mitochondrial targeting seq
      R content:        1       Hyd Moment(75):  6.23
      Hyd Moment(95):  8.46     G content:       2
      D/E content:      1       S/T content:     0
      Score: -4.74

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 12  MRL|AV

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: KKSASDLTWDNLKGKKS at  452
      content of basic residues:  12.2%
      NLS Score:  0.02

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

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

none

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

PTS2: 2nd peroxisomal targeting signal:  found
      KILRQQQHL at 618

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):

        21.7 %: nuclear
        17.4 %: mitochondrial
        17.4 %: cytoplasmic
        17.4 %: endoplasmic reticulum
        13.0 %: extracellular, including cell wall
         8.7 %: vacuolar
         4.3 %: Golgi

>> prediction for CG113367-02 is nuc (k=23)
```

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
AAY50717	Human serum transferrin protein - Homo sapiens, 698 aa. [US5986067-A, 16-NOV-1999]	1..698 1..698	698/698 (100%) 698/698 (100%)	0.0
AAW54354	Serotransferrin precursor - Homo sapiens, 698 aa. [WO9810291-A1, 12-MAR-1998]	1..698 1..698	698/698 (100%) 698/698 (100%)	0.0
AAR66492	Human transferrin precursor - Homo sapiens, 698 aa. [FR2705194-A, 25-NOV-1994]	1..698 1..698	698/698 (100%) 698/698 (100%)	0.0
AAR12499	Human transferrin - Homo sapiens, 698 aa. [US5026651-A, 25-JUN-1991]	1..698 1..698	697/698 (99%) 697/698 (99%)	0.0
AAP70384	Sequence of human serotransferrin (Tf) - Homo sapiens, 696 aa. [NL8700285-A, 01-SEP-1987]	1..698 1..696	692/698 (99%) 693/698 (99%)	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
P02787	Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin) (PRO1400) - Homo sapiens (Human), 698 aa.	1..698 1..698	698/698 (100%) 698/698 (100%)	0.0
P19134	Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin) - Oryctolagus cuniculus (Rabbit), 695 aa.	1..697 1..694	549/697 (78%) 613/697 (87%)	0.0
TFRBP	transferrin precursor - rabbit, 694 aa.	1..697 1..693	548/697 (78%) 612/697 (87%)	0.0
P12346	Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin) - Rattus norvegicus (Rat), 698 aa.	1..697 1..697	512/701 (73%) 580/701 (82%)	0.0
Q921I1	Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin) - Mus musculus (Mouse), 697 aa.	1..697 1..696	512/702 (72%) 582/702 (81%)	0.0

PFam analysis predicts 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
transferrin	25..347	242/339 (71%) 319/339 (94%)	2.7e-256
transferrin	361..683	154/353 (44%) 234/353 (66%)	3.3e-109

Example 3. NOV 3: CG132364: Programed cell death protein 1 precursor (Protein PD-1).

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

Table 3A. NOV3 Sequence Analysis

	SEQ ID NO: 15	918 bp	
NOV3a, CG132364-01 DNA Sequence	CACTCTAGGTGGGGCTGCTCCAGGC**ATG**CAGATCCCACAGGCGCCCTGGCCAGTCGTCTGGGCG GTGCTACAACTGGGCTGGCGGCCAGGATGGTTCTTAGACTCCCCAGACAGGCCCTGGAACCCCC CCACCTTCTCCCCAGCCCTGCTCGTGGTGACCGAAGGGGACAACGCCACCTTCACCTGCAGCTT CTCCAACACATCGGAGAGCTTCGTGCTAAACTGGTACCGCATGAGCCCCAGCAACCAGACGGAC AAGCTGGCCGCCTTCCCCGAGGACCGCAGCCAGCCCGGCCAGGACTGCCGCTTCCGTGTCACAC AACTGCCCAACGGGCGTGACTTCCACATGAGCGTGGTCAGGGCCCGGCGCAATGACAGCGGCAC CTACCTCTGTGGGGCCATCTCCCTGGCCCCCAAGGCGCAGATCAAAGAGAGCCTGCGGGCAGAG CTCAGGGTGACAGAGAGAAGGGCAGAAGTGCCCACAGCCCACCCCAGCCCCTCACCCAGGCCAG CCGGCCAGTTCCAAACCCTGGTGGTTGGTGTCGTGGGCGGCCTGCTGGGCAGCCTGGTGCTGCT AGTCTGGGTCCTGGCCGTCATCTGCTCCCGGGCCGCACGAGGGACAATAGGAGCCAGGCGCACC GGCCAGCCCCTGAAGGAGGACCCCTCAGCCGTGCCGGTGTTCTCTGTGGACTATGGGGAGCTGG ATTTCCAGTGGCGAGAGAAGACCCCGGAGCCCCCCGTGCCCTGTGTCCCTGAGCAGACGGAGTA TGCCACCATTGTCTTTCCTAGCGGAATGGGCACCTCATCCCCCGCCCGCAGGGGCTCAGCTGAC GGCCCTCGGAGTGCCCAGCCACTGAGGCCTGAGGATGGACACTGCTCTTGGCCCCTC**TGA**CCGG CTTCCTTGGCCACCAGTGTTCT		
	ORF Start: ATG at 26		ORF Stop: TGA at 890
	SEQ ID NO: 16	288 aa	MW at 31646.6kD
NOV3a, CG132364-01 Protein Sequence	MQIPQAPWPVVWAVLQLGWRPGWFLDSPDRPWNPPTFSPALLVVTEGDNATFTCSFSNTSESFV LNWYRMSPSNQTDKLAAFPEDRSQPGQDCRFRVTQLPNGRDFHMSVVRARRNDSGTYLCGAISL APKAQIKESLRAELRVTERRAEVPTAHPSPSPRPAGQFQTLVVGVVGGLLGSLVLLVWVLAVIC SRAARGTIGARRTGQPLKEDPSAVPVFSVDYGELDFQWREKTPEPPVPCVPEQTEYATIVFPSG MGTSSPARRGSADGPRSAQPLRPEDGHCSWPL		
	SEQ ID NO: 17	772 bp	
NOV3b, CG132364-02 DNA Sequence	CACTCTAGGTGGGGCTGCTCCAGGC**ATG**CAGATCCCACAGGCGCCCTGGCCAGTCGTCTGGGCG GTGCTACAACTGGGCTGGCGGCCAGGATGGTTCTTAGACTCCCCAGACAGGCCCTGGAACCCCC CCACCTTCTCCCCAGCCCTGCTCGTGGTGACCGAAGGGGACAACGCCACCTTCACCTGCAGCTT CTCCAACACATCGGAGAGCTTCGTGCTAAACTGGTACCGCATGAGCCCCAGCAACCAGACGGAC AAGCTGGCCGCCTTCCCCGAGGACCGCAGCCAGCCCGGCCAGGACTGCCGCTTCCGTGTCACAC AACTGCCCAACGGGCGTGACTTCCACATGAGCGTGGTCAGGGCCCGGCGCAATGACAGCGGCAC CTACCTCTGTGGGGCCATCTCCCTGGCCCCCAAGGCGCAGATCAAAGAGAGCCTGCGGGCAGAG CTCAGGGTGACAGGGACAATAGGAGCCAGGCGCACCGGCCAGCCCCTGAAGGAGGACCCCTCAG CCGTGCCTGTGTTCTCTGTGGACTATGGGGAGCTGGATTTCCAGTGGCGAGAGAAGACCCCGGA GCCCCCCGTGCCCTGTGTCCCTGAGCAGACGGAGTATGCCACCATTGTCTTTCCTAGCGGAATG GGCACCTCATCCCCCGCCCGCAGGGGCTCAGCCGACGGCCCTCGGAGTGCCCAGCCACTGAGGC CTGAGGATGGACACTGCTCTTGGCCCCTC**TGA**CCGGCTTCCTTGGCCACCAGTGTTCTGCAGAC CCTC		
	ORF Start: ATG at 26		ORF Stop: TGA at 734
	SEQ ID NO: 18	236 aa	MW at 26232.2kD
NOV3b, CG132364-02 Protein Sequence	MQIPQAPWPVVWAVLQLGWRPGWFLDSPDRPWNPPTFSPALLVVTEGDNATFTCSFSNTSESFV LNWYRMSPSNQTDKLAAFPEDRSQPGQDCRFRVTQLPNGRDFHMSVVRARRNDSGTYLCGAISL APKAQIKESLRAELRVTGTIGARRTGQPLKEDPSAVPVFSVDYGELDFQWREKTPEPPVPCVPE QTEYATIVFPSGMGTSSPARRGSADGPRSAQPLRPEDGHCSWPL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 3B.

Table 3B. Comparison of NOV3a against NOV3b.		
Protein Sequence	**NOV3a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV3b	1..288 1..236	236/288 (81%) 236/288 (81%)

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 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 19; peak value 8.26
	PSG score: 3.86
	
	GvH: von Heijne's method for signal seq. recognition
	GvH score (threshold: -2.1): -8.30
	possible cleavage site: between 47 and 48
	
	>>> 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 =-10.08 Transmembrane 176 - 192
	PERIPHERAL Likelihood = 6.89 (at 9)
	ALOM score: -10.08 (number of TMSs: 1)
	
	MTOP: Prediction of membrane topology (Hartmann et al.)
	Center position for calculation: 183
	Charge difference: 1.5 C(4.0) - N(2.5)
	C > N: C-terminal side will be inside
	
	>>>Caution: Inconsistent mtop result with signal peptide
	>>> membrane topology: type 1b (cytoplasmic tail 176 to 288)
	
	MITDISC: discrimination of mitochondrial targeting seq
	R content: 1 Hyd Moment(75): 1.70
	Hyd Moment(95): 3.36 G content: 2
	D/E content: 1 S/T content: 0
	Score: -5.97
	
	Gavel: prediction of cleavage sites for mitochondrial preseq
	R-2 motif at 30 WRP|GW
	
	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: 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 177
      LL at 183

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 %: nuclear
        21.7 %: cytoplasmic
        13.0 %: mitochondrial
         8.7 %: vesicles of secretory system
         4.3 %: vacuolar
         4.3 %: peroxisomal
         4.3 %: endoplasmic reticulum

>> prediction for CG132364-01 is nuc (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**
AAR80443	Human programmed cell death protein-1 (huPD-1) - Homo sapiens, 288 aa. [EP670369-A2, 06-SEP-1995]	1..288 1..288	287/288 (99%) 287/288 (99%)	e-172
ABG30844	Mouse PD-1 protein - Mus musculus, 288 aa. [JP2002165592-A, 11-JUN-2002]	1..288 1..288	172/290 (59%) 202/290 (69%)	1e-93
AAR78614	mPD-1 - Mus musculus, 288 aa. [JP07115988-A, 09-MAY-1995]	1..288 1..288	172/290 (59%) 202/290 (69%)	1e-93
AAR47901	Murine programmed cell death polypeptide PD-1 - Mus musculus, 288 aa. [JP05336973-A, 21-DEC-1993]	1..288 1..288	172/290 (59%) 202/290 (69%)	1e-93
AAR78616	Expression vector pME18S/mPD-1.EXT-AIC2A protein prod - Mus musculus, 596 aa. [JP07115988-A, 09-MAY-1995]	1..167 1..167	103/167 (61%) 121/167 (71%)	1e-55

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**
AAN64003	Programmed cell death 1 - Homo sapiens (Human), 288 aa.	1..288 1..288	288/288 (100%) 288/288 (100%)	e-173
Q15116	Programmed cell death protein 1 precursor (Protein PD-1) (hPD-1) - Homo sapiens (Human), 288 aa.	1..288 1..288	287/288 (99%) 287/288 (99%)	e-172
CAC34698	Sequence 11 from Patent WO0114557 - Homo sapiens (Human), 288 aa.	1..288 1..288	287/288 (99%) 287/288 (99%)	e-172
Q02242	Programmed cell death protein 1 precursor (Protein PD-1) (mPD-1) - Mus musculus (Mouse), 288 aa.	1..288 1..288	172/290 (59%) 202/290 (69%)	4e-93
AAO25116	Programmed cell death 1 - Homo sapiens (Human), 122 aa (fragment).	58..231 1..122	122/174 (70%) 122/174 (70%)	4e-59

PFam analysis predicts 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**
ig	47..125	14/81 (17%) 58/81 (72%)	8.6e-08

Example 4.

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

Table 4A. NOV4 Sequence Analysis			
	SEQ ID NO: 19	677 bp	
NOV4a, CG161497-01 DNA Sequence	**ATG**GCTTGCCTTGGATTTCAGCGGCACAAGGCTCAGCTGAACCTGGCTGCCAGGACCTGGCCCT GCACTCTCCTGTTTTTTCTTCTCTTCATCCCTGTCTTCTGCAAAGCAATGCACGTGGCCCAGCC TGCTTTGGTACTGGCCAGCAGCCGAGGCATCGCCAGCTTTGTGTGTGAGTATGCATCTCCAGGC AAAGCCACTGAGGTCCGGGTGACAGTGCTTCGGCAGGCTGACAGCCAGGTGACTGAAGTCTGTG CGGCAACCTACATGATGGGGAATGAGTTGACCTTCCTAGATGATTCCATCTGCACGGGCACCTC CAGTGGAAATCAAGTGAACCTCACTATCCAAGGACTGAGGGCCATGGACACGGGACTCTACATC TGCAAGGTGGAGCTCATGTACCCACCGCCATACTACCTGGGCATAGGCAACGGAACCCAGATTT ATGTAATTGATCCAGAACCGTGCCCAGATTCTGACTTCCTCCTCTGGATCCTTGCAGCAGTTAG TTCGGGGTTGTTTTTTTATAGCTTTCTCCTCACAGCTGTTTCTTTGAGCAAAATGCTAAAGAAA AGAAGCCCTCTTACAACAGGGGTCTATGTGAAAATGCCCCCAACAGAGCCAGAATGTGAAAAGC AATTTCAGCCTTATTTTATTCCCATCAAT**TGA**GAATT		
	ORF Start: ATG at 1		ORF Stop: TGA at 670
	SEQ ID NO: 20	223 aa	MW at 24639.6kD
NOV4a, CG161497-01 Protein Sequence	MACLGFQRHKAQLNLAARTWPCTLLFFLLFIPVFCKAMHVAQPALVLASSRGIASFVCEYASPG KATEVRVTVLRQADSQVTEVCAATYMMGNELTFLDDSICTGTSSGNQVNLTIQGLRAMDTGLYI CKVELMYPPPYYLGIGNGTQIYVIDPEPCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKK RSPLTTGVYVKMPPTEPECEKQFQPYFIPIN		
	SEQ ID NO: 21	566 bp	
NOV4b, CG161497-02 DNA Sequence	**ATG**GCTTGCCTTGGATTTCAGCGGCACAAGGCTCAGCTGAACCTGGCTGCCAGGACCTGGCCCT GCACTCTCCTGTTTTTTCTTCTCTTCATCCCTGTCTTCTGCAAAGCAATGCACGTGGCCCAGCC TGCTTTGGTACTGGCCAGCAGCCGAGGCATCGCCAGCTTTGTGTGTGAGTATGCATCTCCAGGC AAAGCCACTGAGGTCCGGGTGACAGTGCTTCGGCAGGCTGACAGCCAGGTGACTGAAGTCTGTG CGGCAACCTACATGATGGGGAATGAGTTGACCTTCCTAGATGATTCCATCTGCACGGGCACCTC CAGTGGAAATCAAGTGAACCTCACTATCCAAGGACTGAGGGCCATGGACACGGGACTCTACATC TGCAAGGTGGAGCTCATGTACCCACCGCCATACTACCTGGGCATAGGCAACGGAACCCAGATTT ATGTAATTCTAAAGAAAAGAAGCCCTCTTACAACAGGGGTCTATGTGAAAATGCCCCCAACAGA GCCAGAATGTGAAAAGCAATTTCAGCCTTATTTTATTCCCATCAAT**TGA**GAATT		
	ORF Start: ATG at 1		ORF Stop: TGA at 559
	SEQ ID NO: 22	186 aa	MW at 20575.9kD
NOV4b, CG161497-02 Protein Sequence	MACLGFQRHKAQLNLAARTWPCTLLFFLLFIPVFCKAMHVAQPALVLASSRGIASFVCEYASPG KATEVRVTVLRQADSQVTEVCAATYMMGNELTFLDDSICTGTSSGNQVNLTIQGLRAMDTGLYI CKVELMYPPPYYLGIGNGTQIYVILKKRSPLTTGVYVKMPPTEPECEKQFQPYFIPIN		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 4B.

Table 4B. Comparison of NOV4a against NOV4b.		
Protein Sequence	**NOV4a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV4b	1..223 1..186	186/223 (83%) 186/223 (83%)

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

Table 4C. Protein Sequence Properties NOV4a	
SignalP analysis:	Cleavage site between residues 38 and 39

PSORT II analysis:	PSG: a new signal peptide prediction method
	N-region: length 10; pos.chg 2; neg.chg 0
	H-region: length 7; peak value -5.66
	PSG score: -10.06
	
	GvH: von Heijne's method for signal seq. recognition
	GvH score (threshold: -2.1): 0.93
	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: 2
	INTEGRAL Likelihood = -5.68 Transmembrane 22 - 38
	INTEGRAL Likelihood = -2.23 Transmembrane 166 - 182
	PERIPHERAL Likelihood = 4.14 (at 44)
	ALOM score: -5.68 (number of TMSs: 2)
	
	MTOP: Prediction of membrane topology (Hartmann et al.)
	Center position for calculation: 29
	Charge difference: -2.0 C(1.5) - N(3.5)
	N >= C: N-terminal side will be inside
	
	>>> membrane topology: type 3a
	
	MITDISC: discrimination of mitochondrial targeting seq
	R content: 3 Hyd Moment(75): 9.09
	Hyd Moment(95): 7.60 G content: 2
	D/E content: 1 S/T content: 5
	Score: -1.11
	
	Gavel: prediction of cleavage sites for mitochondrial preseq
	R-2 motif at 61 SRG|IA
	
	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: 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): 47.8 %: mitochondrial 30.4 %: endoplasmic reticulum 17.4 %: nuclear 4.3 %: cytoplasmic >> prediction for CG161497-01 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.

5

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

AAU74509	Human cytotoxic T-lymphocyte-associated protein 4 (CTLA4) variant - Homo sapiens, 223 aa. [WO200190122-A2, 29-NOV-2001]	1..223 1..223	222/223 (99%) 223/223 (99%)	e-129
AAU74508	Human cytotoxic T-lymphocyte-associated protein 4 (CTLA4) - Homo sapiens, 223 aa. [WO200190122-A2, 29-NOV-2001]	1..223 1..223	221/223 (99%) 222/223 (99%)	e-128
AAY15129	Human CTLA-4 protein - Homo sapiens, 223 aa. [WO9957266-A2, 11-NOV-1999]	1..223 1..223	221/223 (99%) 222/223 (99%)	e-128
ABB79934	Human CTLA4 - Homo sapiens, 223 aa. [US2002115209-A1, 22-AUG-2002]	1..223 1..223	221/223 (99%) 222/223 (99%)	e-128
AAU00687	Human CTLA4 protein - Homo sapiens, 223 aa. [WO200130966-A2, 03-MAY-2001]	1..223 1..223	221/223 (99%) 222/223 (99%)	e-128

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
Q8WXJ1	Cytotoxic T-lymphocyte-associated protein 4 - Homo sapiens (Human), 223 aa.	1..223 1..223	222/223 (99%) 223/223 (99%)	e-128
P16410	Cytotoxic T-lymphocyte protein 4 precursor (Cytotoxic T-lymphocyte-associated antigen 4) (CTLA-4) (CD152 antigen) - Homo sapiens (Human), 223 aa.	1..223 1..223	221/223 (99%) 222/223 (99%)	e-128
AAB59385	CYTOTOXIC T-LYMPHOCYTE-ASSOCIATED PROTEIN 4 - Homo sapiens (Human), 223 aa.	1..223 1..223	221/223 (99%) 222/223 (99%)	e-128
Q9BDN7	CD152 protein precursor - Papio anubis (Olive baboon), 223 aa.	1..223 1..223	215/223 (96%) 218/223 (97%)	e-122
Q9BDC4	CD152 protein precursor - Macaca mulatta (Rhesus macaque),, 223 aa.	1..223 1..223	214/223 (95%) 218/223 (96%)	e-122

PFam analysis predicts 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**

Example 5.

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

Table 5A. NOV5 Sequence Analysis			
	SEQ ID NO: 23	2610 bp	
NOV5a, CG169401-01 DNA Sequence	GGACTGTTAACTGTTTCTGGCAAAC**ATG**AAGTCAGGCCTCTGGTATTTCTTTCTCTTCTGCTTG CGCATTAAAGTTTTAACAGGAGAAATCAATGGTTCTGCCAATTATGAGATGTTTATATTTCACA ACGGAGGTGTACAAATTTTATGCAAATATCCTGACATTGTCCAGCAATTTAAAATGCAGTTGCT GAAAGGGGGGCAAATACTCTGCGATCTCACTAAGACAAAAGGAAGTGGAAACACAGTGTCCATT AAGAGTCTGAAATTCTGCCATTCTCAGTTATCCAACAACAGTGTCTCTTTTTTTCTATACAACT TGGACCATTCTCATGCCAACTATTACTTCTGCAACCTATCAATTTTTGATCCTCCTCCTTTTAA AGTAACTCTTACAGGAGGATATTTGCATATTTATGAATCACAACTTTGTTGCCAGCTGAAGTTC TGGTTACCCATAGGATGTGCAGCCTTTGTTGTAGTCTGCATTTTGGGATGCATACTTATTTGTT GGCTTACAAAAAAGAAGTATTCATCCAGTGTGCACGACCCTAACGGTGAATACATGTTCATGAG AGCAGTGAACACAGCCAAAAAATCTAGACTCACAGATGTGACCCTA**TAA**TATGGAACTCTGGCA CCCAGGCATGAAGCACGTTGGCCAGTTTTCCTCAACTTGAAGTGCAAGATTCTCTTATTTCCGG GACCACGGAGAGTCTGACTTAACTACATACATCTTCTGCTGGTGTTTTGTTCAATCTGGAAGAA TGACTGTATCAGTCAATGGGGATTTTAACAGACTGCCTTGGTACTGCCGAGTCCTCTCAAAACA AACACCCTCTTGCAACCAGCTTTGGAGAAAGCCCAGCTCCTGTGTGCTCACTGGGAGTGGAATC CCTGTCTCCACATCTGCTCCTAGCAGTGCATCAGCCAGTAAAACAAACACATTTACAAGAAAAA TGTTTTAAAGATGCCAGGGGTACTGAATCTGCAAAGCAAATGAGCAGCCAAGGACCAGCATCTG TCCGCATTTCACTATCATACTACCTCTTCTTTCTGTAGGGATGAGAATTCCTCTTTTAATCAGT CAAGGGAGATGCTTCAAAGCTGGAGCTATTTTATTTCTGAGATGTTGATGTGAACTGTACATTA GTACATACTCAGTACTCTCCTTCAATTGCTGAACCCCAGTTGACCATTTTACCAAGACTTTAGA TGCTTTCTTGTGCCCTCAATTTTCTTTTTTAAAAATACTTCTACATGACTGCTTGACAGCCCAAC AGCCACTCTCAATAGAGAGCTATGTCTTACATTCTTTCCTCTGCTGCTCAATAGTTTTATATAT CTATGCATACATATATACACACATATGTATATAAAATTCATAATGAATATATTTGCCTATATTC TCCCTACAAGAATATTTTTGCTCCAGAAAGACATGTTCTTTTCTCAAATTCAGTTAAAATGGTT TACTTTGTTCAAGTTAGTGGTAGGAAACATTGCCCGGAATTGAAAGCAAATTTATTTTATTATC CTATTTTCTACCATTATCTATGTTTTCATGGTGCTATTAATTACAAGTTTAGTTCTTTTTGTAG ATCATATTAAAATTGCAAACAAAATCATCTTTAATGGGCCAGCATTCTCATGGGGTAGAGCAGA ATATTCATTTAGCCTGAAAGCTGCAGTTACTATAGGTTGCTGTCAGACTATACCCATGGTGCCT CTGGGCTTGACAGGTCAAAATGGTCCCCATCAGCCTGGAGCAGCCCTCCAGACCTGGGTGGAAT TCCAGGGTTGAGAGACTCCCCTGAGCCAGAGGCCACTAGGTATTCTTGCTCCCAGAGGCTGAAG TCACCCTGGGAATCACAGTGGTCTACCTGCATTCATAATTCCAGGATCTGTGAAGAGCACATAT GTGTCAGGGCACAATTCCCTCTCATAAAAACCACACAGCCTGGAAATTGGCCCTGGCCCTTCAA GATAGCCTTCTTTAGAATATGATTTGGCTAGAAAGATTCTTAAATATGTGGAATATGATTATTC TTAGCTGGAATATTTTCTCTACTTCCTGTCTGCATGCCCAAGGCTTCTGAAGCAGCCAATGTCG ATGCAACAACATTTGTAACTTTAGGTAAACTGGGATTATGTTGTAGTTTAACATTTTGTAACTG TGTGCTTATAGTTTACAAGTGAGACCCGATATGTCATTATGCATACTTATATTATCTTAAGCAT GTGTAATGCTGGATGTGTACAGTACAGTACTTAACTTGTAATTTGAATCTAGTATGGTGTTCTG TTTTCAGCTGACTTGGACAACCTGACTGGCTTTGCACAGGTGTTCCCTGAGTTGTTTGCAGGTT TCTGTGTGTGGGGTGGGGTATGGGGAGGAGAACCTTCATGGTGGCCCACCTGGCCTGGTTGTCC AAGCTGTGCCTCGACACATCCTCATCCCAAGCATGGGACACCTCAAGATGAATAATAATTCACA AAATTTCTGTGAAATCAAATCCAGTTTTAAGAGGAGCCACTTATCAAAGAGATTTTAACAGTAG TAAGAAGGCAAAGAATAAACATTTGATATTCAGCAACTGAAAAAAAAAAA		
	ORF Start: ATG at 26		ORF Stop: TAA at 623

	SEQ ID NO: 24	199 aa	MW at 22624.4kD
NOV5a, CG169401-01 Protein Sequence	MKSGLWYFFLFCLRIKVLTGEINGSANYEMFIFHNGGVQILCKYPDIVQQFKMQLLKGGQILCD LTKTKGSGNTVSIKSLKFCHSQLSNNSVSFFLYNLDHSHANYYFCNLSIFDPPPFKVTLTGGYL HIYESQLCCQLKFWLPIGCAAFVVVCILGCILICWLTKKKYSSSVHDPNGEYMFMRAVNTAKKS RLTDVTL		
	SEQ ID NO: 25	548 bp	
NOV5b, CG169401-02 DNA Sequence	GGACTGTTAACTGTTTCTGGCAAAC**ATG**AAGTCAGGCCTCTGGTATTTCTTTCTCTTCTGCTTG CGCATTAAAGTTTTAACAGGAGAAATCAATGGTTCTGCCAATTATGAGATGTTTATATTTCACA ACGGAGGTGTACAAATTTTATGCAAATATCCTGACATTGTCCAGCAATTTAAAATGCAGTTGCT GAAAGGGGGGCAAATACTCTGCGATCTCACTAAGACAAAAGGAAGTGGAAACACAGTGTCCATT AAGAGTCTGAAATTCTGCCATTCTCAGTTATCCAACAACAGTGTCTCTTTTTTTCTATACAACT TGGACCATTCTCATGCCAACTATTACTTCTGCAACCTATCAATTTTTGATCCTCCTCCTTTTAA AGTAACTCTTACAGGAGGATATTTGCATATTTATGAATCACAACTTTGTTGCCAGCTGAAGTTC TGGTTACCCATAGGATGTGCAGCCTTTGTTGTAGTCTGCATTTTGGGATGCATACTTATTTGTT GGCTTACAAAAAAAGATGTGACCCTA**TAA**TATGGAA		
	ORF Start: ATG at 26		ORF Stop: TAA at 539
	SEQ ID NO: 26	171 aa	MW at 19423.8kD
NOV5b, CG169401-02 Protein Sequence	MKSGLWYFFLFCLRIKVLTGEINGSANYEMFIFHNGGVQILCKYPDIVQQFKMQLLKGGQILCD LTKTKGSGNTVSIKSLKFCHSQLSNNSVSFFLYNLDHSHANYYFCNLSIFDPPPFKVTLTGGYL HIYESQLCCQLKFWLPIGCAAFVVVCILGCILICWLTKKDVTL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 5B.

Table 5B. Comparison of NOV5a against NOV5b.

Protein Sequence	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV5b	1..167 1..167	167/167 (100%) 167/167 (100%)

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 20 and 21
PSORT II analysis:	PSG: a new signal peptide prediction method N-region: length 2; pos.chg 1; neg.chg 0 H-region: length 11; peak value 10.41 PSG score: 6.01 GvH: von Heijne's method for signal seq. recognition GvH score (threshold: -2.1): -5.25 possible cleavage site: between 20 and 21 >>> 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 =-12.95 Transmembrane 145 - 161 PERIPHERAL Likelihood = 2.23 (at 4)

```
      ALOM score: -12.95  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 152
      Charge difference:  1.0   C( 1.5) - N( 0.5)
      C > N: C-terminal side will be inside

>>>Caution: Inconsistent mtop result with signal peptide
>>> membrane topology: type 1b  (cytoplasmic tail 145 to 199)

MITDISC: discrimination of mitochondrial targeting seq
      R content:        1       Hyd Moment(75):  5.54
      Hyd Moment(95):  6.77     G content:       2
      D/E content:      1       S/T content:     2
      Score: -4.44

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 24  LRI|KV

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   9.5%
      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: 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 CG169401-01 is nuc (k=23)

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.

5

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
ABP54902	Human inducible co-stimulator ICOS - Homo sapiens, 199 aa. [WO200274803-A2, 26-SEP-2002]	1..199 1..199	199/199 (100%) 199/199 (100%)	e-119
ABG63367	Human albumin fusion protein #42 - Homo sapiens, 199 aa. [WO200177137-A1, 18-OCT-2001]	1..199 1..199	199/199 (100%) 199/199 (100%)	e-119
ABG63366	Human albumin fusion protein #41 - Homo sapiens, 199 aa. [WO200177137-A1, 18-OCT-2001]	1..199 1..199	199/199 (100%) 199/199 (100%)	e-119
AAU99793	Human CD28 related protein-1 (CRP1) - Homo sapiens, 199 aa. [WO200244364-A2, 06-JUN-2002]	1..199 1..199	199/199 (100%) 199/199 (100%)	e-119
AAE03428	Human gene 2 encoded secreted protein HT2SG64, SEQ ID NO: 111 - Homo sapiens, 199 aa. [WO200132675-A1, 10-MAY-2001]	1..199 1..199	199/199 (100%) 199/199 (100%)	e-119

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

Q9Y6W8	Activation-inducible lymphocyte IMMUNOMEDIATORY molecule AILIM precursor (Inducible T-cell co-stimulator) (Inducible COSTIMULATOR precursor) - Homo sapiens (Human), 199 aa.	1..199 1..199	199/199 (100%) 199/199 (100%)	e-118
Q8N6W8	Similar to inducible T-cell co-stimulator - Homo sapiens (Human), 168 aa.	1..167 1..167	167/167 (100%) 167/167 (100%)	5e-99
Q9WVS0	Activation-inducible lymphocyte IMMUNOMEDIATORY molecule AILIM precursor (CCLP precursor) (Surface protein) - Mus musculus (Mouse), 200 aa.	1..198 1..199	138/199 (69%) 158/199 (79%)	1e-77
Q9JL17	CD28-related protein 1 (Inducible costimulatory protein) (Inducible T-cell co-stimulator) - Mus musculus (Mouse), 200 aa (fragment).	1..198 1..199	138/199 (69%) 157/199 (78%)	3e-77
Q9R1T7	Activation-inducible lymphocyte immunomediatory molecule AILIM precursor - Rattus norvegicus (Rat), 200 aa.	9..198 9..199	131/191 (68%) 148/191 (76%)	3e-73

PFam analysis predicts 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

Example 6. NOV 6, CG176697, Novel Plasma Membrane Protein.

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

Table 6A. NOV6 Sequence Analysis

	SEQ ID NO: 27	838 bp	
NOV6a, CG176697-02 DNA Sequence	CACCGGATCCACC**ATG**GCGGCGGGGGGCAGGATGGAGGACGGTTCCTTGGATATCACCCAGAGT ATTGAAGACGACCCACTTCTGGATGCCCAGCTTCTCCCACACCACTCATTACAAGCTCACTTTA GACCCCGATTCCATCCTCTTCCTACAGTCATCATAGTGAATCTTCTGTGGTTTATTCATCTCGT GTTTGTTGTTTTAGCATTTTTAACAGGTGTGCTTTGTTCTTATCCTAATCCAAATGAGGACAAG TGCCCAGGAAATTACACAAACCCATTGAAAGTTCAGACGGTTATAATCCTTGGGAAAGTTATTT TGTGGATTCTCCATTTACTCCTTGAATGCTACATCCAGTATCACCACAGCAAAATCAGAAACCG AGGCTATAACTTGATCTACCGATCAACAAGGCATCTCAAGAGACTTGCGTTGATGATACAGTCC TCTGGCAACACAGTGCTTCTCCTCATACTGTGCATGCAGCACTCCTTCCCAGAGCCTGGCAGAT TGTATCTTGACCTCATTCTGGCCATCTTGGCACTGGAACTCATCTGTTCCCTGATATGTCTCCT		

	CATTTACACAGTGAAAATCCGGAGATTTAATAAAGCTAAACCAGAGCCTGATATACTTGAAGAA GAAAAAATCTATGCTTACCCCAGCAATATTACCTCGGAGACTGGATTCAGAACTATTTCAAGCC TAGAAGAAATTGTTGAAAAGCAAGGAGACACCATTGAATACCTGAAGCGACACAATGCGCTGCT GAGTAAGCGATTGTTGGCTCTCACTTCCTCAGACCTGGGCTGTCAGCCAAGTAGAACG**TAG**GTC GACGGC		
	ORF Start: ATG at 14		ORF Stop: TAG at 827
	SEQ ID NO: 28	271 aa	MW at 30921.9kD
NOV6a, CG176697-02 Protein Sequence	MAAGGRMEDGSLDITQSIEDDPLLDAQLLPHHSLQAHFRPRFHPLPTVIIVNLLWFIHLVFVVL AFLTGVLCSYPNPNEDKCPGNYTNPLKVQTVIILGKVILWILHLLLECYIQYHHSKIRNRGYNL IYRSTRHLKRLALMIQSSGNTVLLLILCMQHSFPEPGRLYLDLILAILALELICSLICLLIYTV KIRRFNKAKPEPDILEEEKIYAYPSNITSETGFRTISSLEEIVEKQGDTIEYLKRHNALLSKRL LALTSSDLGCQPSRT		
	SEQ ID NO: 29	2961 bp	
NOV6b, CG176697-01 DNA Sequence	GCTTCCCAAAGTGTGGAATTACAAGCATGAGCCACCGTGCCTGGCCTTCTCCTTGTAAATTCTT GAACACATTTATAATAGTTGTTTTTAAATCTTTGTCCATTAATTCCGGTAACTGGTCATCTCTA AGTCGGCTACTGATTGTGCTTTCTTTAGATTTGGGGTCACATTTTCCTGCTTCTTTGCAGGTTT AAGGCCCAGAGGCTTACAGGTCCAGGTGGAGAGGCCGGGCTGGCCAGGGCTTCGGCCTCCGGCG TCGGGAA**ATG**GCGGCGGGGGGCAGGATGGAGGACGGTTCCTTGGATATCACCCAGAGTATTGAA GACGACCCACTTCTGGATGCCCAGCTTCTCCCACACCACTCATTACAAGCTCACTTTAGACCCC GATTCCATCCTCTTCCTACAGTCATCATAGTGAATCTTCTGTGGTTTATTCATCTCGTGTTTGT TGTTTTAGCATTTTTAACAGGTGTGCTTTGTTCTTATCCTAATCCAAATGAGGACAAGTGCCCA GGAAATTACACAAACCCATTGAAAGTTCAGACGGTTATAATCCTTGGGAAAGTTATTTTGTGGA TTCTCCATTTACTCCTTGAATGCTACATCCAGTATCACCACAGCAAAATCAGAAACCGAGGCTA TAACTTGATCTACCGATCAACAAGGCATCTCAAGAGACTTGCGTTGATGATACAGTCCTCTGGC AACACAGTGCTTCTCCTCATACTGTGCATGCAGCACTCCTTCCCAGAGCCTGGCAGATTGTATC TTGACCTCATTCTGGCCATCTTGGCACTGGAACTCATCTGTTCCCTGATATGTCTCCTCATTTA CACAGTGAAAATCCGGAGATTTAATAAAGCTAAACCAGAGCCTGATATACTTGAAGAAGAAAAA ATCTATGCTTACCCCAGCAATATTACCTCGGAGACTGGATTCAGAACTATTTCAAGCCTAGAAG AAATTGTTGAAAAGCAAGGAGACACCATTGAATACCTGAAGCGACACAATGCGCTGCTGAGTAA GCGATTGTTGGCTCTCACTTCCTCAGACCTGGGCTGTCAGCCAAGTAGAACG**TGA**GAGGCTCAC GGTCATGACAGCAATTGCAGAGGAACCCAGAGTAATTGAGACTGACTGACCACCTGACAAGCTG CCACGGGGAACTGCAGCTTTTGCTGAATAGCATTTTACAGTGTTTGTTGGAAACCTGAATTTGG TTCTGACTTCTGTGGCTGTTTAAAATATAGGGCTTGTGGGTCACTTGAAAAGTACCTGTAGAAG CCCGGATAACTTGAGGGGAATGTCTGTTTGTACTTTTGGAAATTATTTAACTGCTTGGTTTATC CTAGGATCAGAGGCTAAACAACTCCATAGTCAACACTTTTCCACCTGACATTAGAACCCTGTAA TGATTTCATTATGGATAGGGGAAGTCTAACAAGACAATCGATTTTAAACAGGGTTTTAATCAAA TAAGTTCTTCCCACTTTTAAGCTGATGAAAGAAGTCATTATTTCTTGTGAATATTTTTTCCTGG AGAGCCTTATCATGTATTTTATATGCTTATGTGGTGTTGGATGACATCATGGACCATATAGCTT TTATAGAGAATTTTTCTCACCATAGGACTGAGGTCTCACCAGGTGATCTACTATGCAAATTCCT ACAGTTTTCTATTCTTAAGAAATAAGGGCTGGGCACGGCGGATCATGAGGTCAGGAAATTGAGA CCATCCTGGCTAGCACGGTGAAACCCTGTCTCTACTAAAAATACAAAAAAAATTAGCCGAGCAT GGTGGCGGGCACCTGTAGTCCCAGCCACCTGGGAGGCTGAGGCAGGAGAATGGTGTGGACCCGG GAGGCAGAGCTTGCAGTGAGCCGAGATCACGCCACTGCGCTCCAGCCTGGGCGACAGAGCGAGA CTCCGTCTCAAAAAAACAAAAAAGAAAAAGAAAAGAAATAAGATATGATGATTGATTACTTGGA CCAGTGTTAGACCGACCTGTGTTAGACGAACCTGTGTTGCCTATTATATTGGCAGCACCTTCCA TTAGCAATAGCTGAAAAATTCAGAGAAGAATTTAGAACTTTCACTTTATTTTTGCCATTCTTCA AGAAGTTGCTGTATATTTCAGTCTGGTTATACCTTCAATTGATCAATAACTTCAATTATAGGAA AGGCATCTTCTATAATTTTAAATTACTGTAAAAAAAAGATACCTTTTAAAGCCTTGGATTATAC TTAAAAATTGTAATGGAAGAAATGCATATTAAGTTATATGCAGATATTAAACATATCTGATTCT GAAGATAAAACTGAAGTATTTTTCAGAGCAAATGGCATTATTTCTCCATTTTTCCTAGTATGCC CCTTTAATTTGTCAGTTTGCTGTTAGCACTGTTATATGCTGCCTTAGAAAGACTTTTGGTTGAA GGCTGAATATGAGGGGAGGGGATATTTATGACTGTGTTAATCTTTTTTGATCATTAGCTGAATA TTCAATATTTGATATGTAACCATATAGGGATGTGTTCGATATAACTATATAGGCAAGGGATATA TTGATTTGGCCACATCTATTTGATTTATGGAGGGCAAATACCTGTTTCTTTTAGATACAGCCAT TGCCGGGGCATTCTAAAAGTTTTCTACTAAGATAAAGATGATTGTGTATGGCTGGGAGCAGTTG CTCATGCCTGTAATCTTAGTGCTTTGGGAGGCTGAGGCAGGAGGATCACTCAAGGCCAGGAGTT TGAGACCAGCCTGGGCAACAATGAGACACTGTCTCTCAAAAAAAAAAAAAAAAAGATCGTGTGTC ACCTGCACACAACATTCACAAACTAAAGCCAAATTGTATTTTTAAAATTTCCTTTCTCCCTTCC GTTTTTTTTTGTTTTTTGGGGGGTTTTTTTGTTGTTGTTGTTTTTGAGACAGAGCGAGACTCCG TCTCAAAAAAACAAAAA		
	ORF Start: ATG at 264		ORF Stop: TGA at 1077
	SEQ ID NO: 30	271 aa	MW at 30921.9kD
NOV6b,	MAAGGRMEDGSLDITQSIEDDPLLDAQLLPHHSLQAHFRPRFHPLPTVIIVNLLWFIHLVFVVL		

CG176697-01 Protein Sequence	AFLTGVLCSYPNPNEDKCPGNYTNPLKVQTVIILGKVILWILHLLLECYIQYHHSKIRNRGYNL IYRSTRHLKRLALMIQSSGNTVLLLILCMQHSFPEPGRLYLDLILAILALELICSLICLLIYTV KIRRFNKAKPEPDILEEEKIYAYPSNITSETGFRTISSLEEIVEKQGDTIEYLKRHNALLSKRL LALTSSDLGCQPSRT		
	SEQ ID NO: 31	838 bp	
NOV6c, 316784436 DNA Sequence	CACCGGATCCACCATGGCGGCGGGGGGCAGGATGGAGGACGGTTCCTTGGATATCACCCAGAGT ATTGAAGACGACCCACTTCTGGATGCCCAGCTTCTCCCACACCACTCATTACAAGCTCACTTTA GACCCCGATTCCATCCTCTTCCTACAGTCATCATAGTGAATCTTCTGTGGTTTATTCATCTCGT GTTTGTTGTTTTAGCATTTTTAACAGGTGTGCTTTGTTCTTATCCTAATCCAAATGAGGACAAG TGCCCAGGAAATTACACAAACCCATTGAAAGTTCAGACGGTTATAATCCTTGGGAAAGTTATTT TGTGGATTCTCCATTTACTCCTTGAATGCTACATCCAGTATCACCACAGCAAAATCAGAAACCG AGGCTATAACTTGATCTACCGATCAACAAGGCATCTCAAGAGACTTGCGTTGATGATACAGTCC TCTGGCAACACAGTGCTTCTCCTCATACTGTGCATGCAGCACTCCTTCCCAGAGCCTGGCAGAT TGTATCTTGACCTCATTCTGGCCATCTTGGCACTGGAACTCATCTGTTCCCTGATATGTCTCCT CATTTACACAGTGAAAATCCGGAGATTTAATAAAGCTAAACCAGAGCCTGATATACTTGAAGAA GAAAAAATCTATGCTTACCCCAGCAATATTACCTCGGAGACTGGATTCAGAACTATTTCAAGCC TAGAAGAAATTGTTGAAAAGCAAGGAGACACCATTGAATACCTGAAGCGACACAATGCGCTGCT GAGTAAGCGATTGTTGGCTCTCACTTCCTCAGACCTGGGCTGTCAGCCAAGTAGAACG**TAG**GTC GACGGC		
	ORF Start: at 2		ORF Stop: TAG at 827
	SEQ ID NO: 32	275 aa	MW at 31268.2kD
NOV6c, 316784436 Protein Sequence	TGSTMAAGGRMEDGSLDITQSIEDDPLLDAQLLPHHSLQAHFRPRFHPLPTVIIVNLLWFIHLV FVVLAFLTGVLCSYPNPNEDKCPGNYTNPLKVQTVIILGKVILWILHLLLECYIQYHHSKIRNR GYNLIYRSTRHLKRLALMIQSSGNTVLLLILCMQHSFPEPGRLYLDLILAILALELICSLICLL IYTVKIRRFNKAKPEPDILEEEKIYAYPSNITSETGFRTISSLEEIVEKQGDTIEYLKRHNALL SKRLLALTSSDLGCQPSRT		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 6B.

Table 6B. Comparison of NOV6a against NOV6b and NOV6c.

Protein Sequence	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV6b	1..271 1..271	271/271 (100%) 271/271 (100%)
NOV6c	1..271 5..275	271/271 (100%) 271/271 (100%)

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

Table 6C. Protein Sequence Properties NOV6a

SignalP analysis:	Cleavage site between residues 66 and 67
PSORT II analysis:	PSG: a new signal peptide prediction method N-region: length 9; pos.chg 1; neg.chg 2 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): -5.62 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:   4
      INTEGRAL    Likelihood =-11.04   Transmembrane   48 -  64
      INTEGRAL    Likelihood = -7.01   Transmembrane   94 - 110
      INTEGRAL    Likelihood = -1.65   Transmembrane  139 - 155
      INTEGRAL    Likelihood =-11.68   Transmembrane  171 - 187
      PERIPHERAL  Likelihood =  7.11 (at 250)
      ALOM score: -11.68  (number of TMSs: 4)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 55
      Charge difference: -3.5   C( 0.5) - N( 4.0)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq
      R content:        1       Hyd Moment(75):  1.10
      Hyd Moment(95):  2.86     G content:       2
      D/E content:     2        S/T content:     0
      Score: -8.16

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 16  GRM|ED

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: 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):

        52.2 %: endoplasmic reticulum
        26.1 %: mitochondrial
        17.4 %: nuclear
         4.3 %: vesicles of secretory system

>> prediction for CG176697-02 is end (k=23)
```

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
ABJ10917	Human secreted protein (SECP) #13 - Homo sapiens, 271 aa. [WO200270669-A2, 12-SEP-2002]	1..271 1..271	271/271 (100%) 271/271 (100%)	e-156
ABB89474	Human polypeptide SEQ ID NO 1850 - Homo sapiens, 197 aa. [WO200190304-A2, 29-NOV-2001]	1..194 1..194	192/194 (98%) 192/194 (98%)	e-109
AAB56877	Human prostate cancer antigen protein sequence SEQ ID NO:1455 - Homo sapiens, 230 aa. [WO200055174-A1, 21-SEP-2000]	1..194 34..227	192/194 (98%) 192/194 (98%)	e-109
AAY13149	Human secreted protein encoded by 5' EST SEQ ID NO: 163 - Homo sapiens, 89 aa. [WO9906552-A2, 11-FEB-1999]	1..89 1..89	74/89 (83%) 74/89 (83%)	4e-37
ABB63338	Drosophila melanogaster polypeptide SEQ ID NO 16806 - Drosophila melanogaster, 260 aa. [WO200171042-A2, 27-SEP-2001]	57..256 49..244	49/209 (23%) 92/209 (43%)	1e-11

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**
AAH36301	Hypothetical protein FLJ38482 - Homo sapiens (Human), 271 aa.	1..271 1..271	271/271 (100%) 271/271 (100%)	e-156
Q8N928	Hypothetical protein FLJ38482 - Homo sapiens (Human), 267 aa.	10..271 6..267	261/262 (99%) 261/262 (99%)	e-149
Q9CXT7	3110005G23Rik protein - Mus musculus (Mouse), 266 aa.	1..271 1..266	211/271 (77%) 233/271 (85%)	e-119
Q8VDU8	Similar to RIKEN cDNA 3110005G23 gene - Mus musculus (Mouse), 221 aa.	1..271 1..221	177/271 (65%) 192/271 (70%)	5e-91
Q9VEH2	CG7523 protein - Drosophila melanogaster (Fruit fly), 260 aa.	57..256 49..244	49/209 (23%) 92/209 (43%)	3e-11

PFam analysis predicts 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**

Example 7. NOV 7: CG178810 POLIOVIRUS RECEPTOR.

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

Table 7A. NOV10 Sequence Analysis			
	SEQ ID NO: 33	1301 bp	
NOV7a, CG178810-03 DNA Sequence	AGGCCCAGCTGCTCGGAGCAACTGGC**ATG**GCCCGAGCCATGGCCGCCGCGTGGCCGCTGCTGCT GGTGGCGCTACTGGTGCTGTCCTGGCCACCCCCAGGAACCGGGGACGTCGTCGTGCAGGCGCCC ACCCAGGTGCCCGGCTTCTTGGGCGACTCCGTGACGCTGCCCTGCTACCTACAGGTGCCCAACA TGGAGGTGACGCATGTGTCACAGCTGACTTGGGCGCGGCATGGTGAATCTGGCAGCATGGCCGT CTTCCACCAAACGCAGGGCCCCAGCTATTCGGAGTCCAAACGGCTGGAATTCGTGGCAGCCAGA CTGGGCGCGGAGCTGCGGAATGCCTCGCTGAGGATGTTCGGGTTGCGCGTAGAGGATGAAGGCA ACTACACCTGCCTGTTCGTCACGTTCCCGCAGGGCAGCAGGAGCGTGGATATCTGGCTCCGAGT GCTTGCCAAGCCCCAGAACACAGCTGAGGTTCAGAAGGTCCAGCTCACTGGAGAGCCAGTGCCC ATGGCCCGCTGCGTCTCCACAGGGGGTCGCCCGCCAGCCCAAATCACCTGGCACTCAGACCTGG GCGGGATGCCCAATACGAGCCAGGTGCCAGGGTTCCTGTCTGGCACAGTCACTGTCACCAGCCT CTGGATATTGGTGCCCTCAAGCCAGGTGGACGGCAAGAATGTGACCTGCAAGGTGGAGCACGAG		

	AGCTTTGAGAAGCCTCAGCCGCTGACTGTGAACCTCACCGTGTACTACCCCCCAGAGGTATCCA TCTCTGGCTATGATAACAACTGGTACCTTGGCCAGAATGAGGCCACCCTGACCTGCGATGCTCG CAGCAACCCAGAGCCCACAGGCTATAATTGGAGCACGACCATGGGTCCCCTGCCACCCTTTGCT GTGGCCCAGGGCGCCCAGCTCCTGATCCGTCCTGTGGACAAACCAATCAACACAACTTTAATCT GCAACGTCACCAATGCCCTAGGAGCTCGCCAGGCAGAACTGACCGTCCAGGTCAAAGAGGGACC TCCCAGTGAGCACTCAGGCATGTCCCGTAACGCCATCATCTTCCTGGTTCTGGGAATCCTGGTT TTTCTGATCCTGCTGGGGATCGGGATTTATTTCTATTGGTCCAAATGTTCCCGTGAGGTCCTTT GGCACTGTCATCTGTGTCCCTCGAGTGAGCATCACCAGAGCTGCCGTACAGAGCATGCCAGCGC CTCAGCTAATGGGCATGTCTCCTATTCAGCTGTGAGCAGAGAGAACAGCTCTTCCCAGGATCCA CAGACAGAGGGCACAAGG**TGA**		
	ORF Start: ATG at 27		ORF Stop: TGA at 1299
	SEQ ID NO: 34	424 aa	MW at 46182.1kD
NOV7a, CG178810-03 Protein Sequence	MARAMAAAWPLLLVALLVLSWPPPGTGDVVVQAPTQVPGFLGDSVTLPCYLQVPNMEVTHVSQL TWARHGESGSMAVFHQTQGPSYSESKRLEFVAARLGAELRNASLRMFGLRVEDEGNYTCLFVTF PQGSRSVDIWLRVLAKPQNTAEVQKVQLTGEPVPMARCVSTGGRPPAQITWHSDLGGMPNTSQV PGFLSGTVTVTSLWILVPSSQVDGKNVTCKVEHESFEKPQPLTVNLTVYYPPEVSISGYDNNWY LGQNEATLTCDARSNPEPTGYNWSTTMGPLPPFAVAQGAQLLIRPVDKPINTTLICNVTNALGA RQAELTVQVKEGPPSEHSGMSRNAIIFLVLGILVFLILLGIGIYFYWSKCSREVLWHCHLCPSS EHHQSCRTEHASASANGHVSYSAVSRENSSSQDPQTEGTR		
	SEQ ID NO: 35	1121 bp	
NOV7b, CG178810-01 DNA Sequence	AGGCCCAGCTGCTCGGAGCAACTGGC**ATG**GCCCGAGCCATGGCCGCCGCGTGGCCGCTGCTGCT GGTGGCGCTACTGGTGCTGTCCTGGCCACCCCCAGGAACCGGGGACGTCGTCGTGCAGGCGCCC ACCCAGGTGCCCGGCTTCTTGGGCGACTCCGTGACGCTGCCCTGCTACCTACAGGTGCCCAACA TGGAGGTGACGCATGTGTCACAGCTGACTTGGGCGCGGCATGGTGAATCTGGCAGCATGGCCGT CTTCCACCAAACGCAGGGCCCCAGCTATTCGGAGTCCAAACGGCTGGAATTCGTGGCAGCCAGA CTGGGCGCGGAGCTGCGGAATGCCTCGCTGAGGATGTTCGGGTTGCGCGTAGAGGATGAAGGCA ACTACACCTGCCTGTTCGTCACGTTCCCGCAGGGCAGCAGGAGCGTGGATATCTGGCTCCGAGT GCTTGCCAAGCCCCAGAACACAGCTGAGGTTCAGAAGGTCCAGCTCACTGGAGAGCCAGTGCCC ATGGCCCGCTGCGTCTCCACAGGGGGTCGCCCGCCAGCCCAAATCACCTGGCACTCAGACCTGG GCGGGATGCCCAATACGAGCCAGGTGCCAGGGTTCCTGTCTGGCACAGTCACTGTCACCAGCCT CTGGATATTGGTGCCCTCAAGCCAGGTGGACGGCAAGAATGTGACCTGCAAGGTGGAGCACGAG AGCTTTGAGAAGCCTCAGCTGCTGACTGTGAACCTCACCGTGTACTACCCCCCAGAGGTATCCA TCTCTGGCTATGATAACAACTGGTACCTTGGCCAGAATGAGGCCACCCTGACCTGCGATGCTCG CAGCAACCCAGAGCCCACAGGCTATAATTGGAGCACGACCATGGGTCCCCTGCCACCCTTTGCT GTGGCCCAGGGCGCCCAGCTCCTGATCCGTCCTGTGGACAAACCAATCAACACAACTTTAATCT GCAACGTCACCAATGCCCTAGGAGCTCGCCAGGCAGAACTGACCGTCCAGGTCAAAGGTACAGA GCATGCCAGCGCCCCAGCTAATGGGCATGTCTCCTATTCAGCTGTGAGCAGAGAGAACAGCTCT TCCCAGGATCCACAGACAGAGGGCACAAGG**TGA**		
	ORF Start: ATG at 27		ORF Stop: TGA at 1119
	SEQ ID NO: 36	364 aa	MW at 39314.2kD
NOV7b, CG178810-01 Protein Sequence	MARAMAAAWPLLLVALLVLSWPPPGTGDVVVQAPTQVPGFLGDSVTLPCYLQVPNMEVTHVSQL TWARHGESGSMAVFHQTQGPSYSESKRLEFVAARLGAELRNASLRMFGLRVEDEGNYTCLFVTF PQGSRSVDIWLRVLAKPQNTAEVQKVQLTGEPVPMARCVSTGGRPPAQITWHSDLGGMPNTSQV PGFLSGTVTVTSLWILVPSSQVDGKNVTCKVEHESFEKPQLLTVNLTVYYPPEVSISGYDNNWY LGQNEATLTCDARSNPEPTGYNWSTTMGPLPPFAVAQGAQLLIRPVDKPINTTLICNVTNALGA RQAELTVQVKGTEHASAPANGHVSYSAVSRENSSSQDPQTEGTR		
	SEQ ID NO: 37	1145 bp	
NOV7c, CG178810-02 DNA Sequence	AGGCCCAGCTGCTCGGAGCAACTGGC**ATG**GCCCGAGCCATGGCCGCCGCGTGGCCGCTGCTGCT GGTGGCGCTACTGGTGCTGTCCTGGCCACCCCCAGGAACCGGGGACGTCGTCGTGCAGGCGCCC ACCCAGGTGCCCGGCTTCTTGGGCGACTCCGTGACGCTGCCCTGCTACCTACAGGTGCCCAACA TGGAGGTGACGCATGTGTCACAGCTGACTTGGGCGCGGCATGGTGAATCTGGCAGCATGGCCGT CTTCCACCAAACGCAGGGCCCCAGCTATTCGGAGTCCAAACGGCTGGAATTCGTGGCAGCCAGA CTGGGCGCGGAGCTGCGGAATGCCTCGCTGAGGATGTTCGGGTTGCGCGTAGAGGATGAAGGCA ACTACACCTGCCTGTTCGTCACGTTCCCGCAGGGCAGCAGGAGCGTGGATATCTGGCTCCGAGT GCTTGCCAAGCCCCAGAACACAGCTGAGGTTCAGAAGGTCCAGCTCACTGGAGAGCCAGTGCCC ATGGCCCGCTGCGTCTCCACAGGGGGTCGCCCGCCAGCCCAAATCACCTGGCACTCAGACCTGG GCGGGATGCCCAATACGAGCCAGGTGCCAGGGTTCCTGTCTGGCACAGTCACTGTCACCAGCCT CTGGATATTGGTGCCCTCAAGCCAGGTGGACGGCAAGAATGTGACCTGCAAGGTGGAGCACGAG AGCTTTGAGAAGCCTCAGCTGCTGACTGTGAACCTCACCGTGTACTACCCCCCAGAGGTATCCA TCTCTGGCTATGATAACAACTGGTACCTTGGCCAGAATGAGGCCACCCTGACCTGCGATGCTCG CAGCAACCCAGAGCCCACAGGCTATAATTGGAGCACGACCATGGGTCCCCTGCCACCCTTTGCT GTGGCCCAGGGCGCCCAGCTCCTGATCCGTCCTGTGGACAAACCAATCAACACAACTTTAATCT GCAACGTCACCAATGCCCTAGGAGCTCGCCAGGCAGAACTGACCGTCCAGGTCAAAGAGGGACC		

	TCCCAGTGAGCACTCAGGTACAGAGCATGCCAGCGCCTCAGCTAATGGGCATGTCTCCTATTCA GCTGTGAGCAGAGAGAACAGCTCTTCCCAGGATCCACAGACAGAGGGCACAAGG**TGA**
	ORF Start: ATG at 27 | | ORF Stop: TGA at 1143
	SEQ ID NO: 38 | 372 aa | MW at 40125.0kD
NOV7c, CG178810-02 Protein Sequence	MARAMAAAWPLLLVALLVLSWPPPGTGDVVVQAPTQVPGFLGDSVTLPCYLQVPNMEVTHVSQL TWARHGESGSMAVFHQTQGPSYSESKRLEFVAARLGAELRNASLRMFGLRVEDEGNYTCLFVTF PQGSRSVDIWLRVLAKPQNTAEVQKVQLTGEPVPMARCVSTGGRPPAQITWHSDLGGMPNTSQV PGFLSGTVTVTSLWILVPSSQVDGKNVTCKVEHESFEKPQLLTVNLTVYYPPEVSISGYDNNWY LGQNEATLTCDARSNPEPTGYNWSTTMGPLPPFAVAQGAQLLIRPVDKPINTTLICNVTNALGA RQAELTVQVKEGPPSEHSGTEHASASANGHVSYSAVSRENSSSQDPQTEGTR

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 7B.

Table 7B. Comparison of NOV7a against NOV7b and NOV7c.

Protein Sequence	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV7b	1..424 1..364	352/424 (83%) 354/424 (83%)
NOV7c	1..424 1..372	371/424 (87%) 371/424 (87%)

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

Table 7C. Protein Sequence Properties NOV7a

SignalP analysis:	Cleavage site between residues 28 and 29

PSORT II analysis:	PSG: a new signal peptide prediction method
	N-region: length 3; pos.chg 1; neg.chg 0
	H-region: length 24; peak value 10.88
	PSG score: 6.47
	
	GvH: von Heijne's method for signal seq. recognition
	GvH score (threshold: -2.1): 1.88
	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: 1
	Number of TMS(s) for threshold 0.5: 1
	INTEGRAL Likelihood =-16.77 Transmembrane 345 - 361
	PERIPHERAL Likelihood = 1.38 (at 192)
	ALOM score: -16.77 (number of TMSs: 1)
	
	MTOP: Prediction of membrane topology (Hartmann et al.)
	Center position for calculation: 13
	Charge difference: -3.0 C(-1.0) - N(2.0)
	N >= C: N-terminal side will be inside
	
	>>> membrane topology: type 1a (cytoplasmic tail 362 to 424)

```
MITDISC: discrimination of mitochondrial targeting seq
      R content:          1       Hyd Moment(75):  8.51
      Hyd Moment(95):  9.43       G content:       2
      D/E content:        1       S/T content:     2
      Score: -3.71

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 78  ARH|GE

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: none
      bipartite: none
      content of basic residues:   6.6%
      NLS Score: -0.47

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

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

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: 76.7

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

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

Final Results (k = 9/23):

        55.6 %: endoplasmic reticulum
        22.2 %: Golgi
        11.1 %: plasma membrane
        11.1 %: extracellular, including cell wall
```

	>> prediction for CG178810-03 is end (k=9)

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 7D.

Table 7D. Geneseq Results for NOV7a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW18669	Poliovirus receptor (45 kda) - Homo sapiens, 417 aa. [US5631407-A, 20-MAY-1997]	1..424 1..417	416/424 (98%) 416/424 (98%)	0.0
AAE23298	Human poliovirus receptor (PVR) alpha protein - Homo sapiens, 417 aa. [WO200228902-A2, 11-APR-2002]	1..424 1..417	415/424 (97%) 416/424 (97%)	0.0
AAR07131	H20A receptor, 416 aa. [WO9010699-A, 20-SEP-1990]	1..424 1..416	415/424 (97%) 415/424 (97%)	0.0
AAW18668	Poliovirus receptor (45 kDa) - Homo sapiens, 392 aa. [US5631407-A, 20-MAY-1997]	1..391 1..391	390/391 (99%) 390/391 (99%)	0.0
AAR07130	H20B receptor, 392 aa. [WO9010699-A, 20-SEP-1990]	1..391 1..391	390/391 (99%) 390/391 (99%)	0.0

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 7E.

Table 7E. Public BLASTP Results for NOV7a

Protein Accession Number	Protein/Organism/Length	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96BJ1	Poliovirus receptor - Homo sapiens (Human), 417 aa.	1..424 1..417	416/424 (98%) 416/424 (98%)	0.0
RWHUPA	poliovirus receptor splice form alpha precursor - human, 417 aa.	1..424 1..417	415/424 (97%) 415/424 (97%)	0.0
P15151	Poliovirus receptor precursor (CD155 antigen) - Homo sapiens (Human), 417 aa.	1..424 1..417	415/424 (97%) 416/424 (97%)	0.0
RWHUPD	poliovirus receptor splice form delta precursor - human, 392 aa.	1..391 1..391	389/391 (99%) 389/391 (99%)	0.0

P32506	Poliovirus receptor precursor - Cercopithecus aethiops (Green monkey) (Grivet), 417 aa.	1..424 1..417	376/424 (88%) 392/424 (91%)	0.0

PFam analysis predicts that the NOV7a protein contains the domains shown in the Table 7F.

Table 7F. Domain Analysis of NOV7a

Pfam Domain	NOV7a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ig	42..125	18/85 (21%) 52/85 (61%)	0.0077
ig	259..314	13/58 (22%) 38/58 (66%)	0.0025

Example 8. NOV 8: CG179299 Novel Membrane Protein.

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

Table 8A. NOV8 Sequence Analysis

	SEQ ID NO: 39	9591 bp	
NOV8a, CG179299-01 DNA Sequence	**ATG**TCCGGCACCAGCAGCCCCGAGGCGGTGAAGAAGCTGCTGGAGAATATGCAGAGCGACTTGC GCGCCTTGTCACTGGAGTGCAAGAAGAAATTCCCACCTGTCAAAGAGTTTGTAGCAGAAGATGG AGGAGAACCAAAGGACAGAGTGTTAGGAGAAAAGCATGGGAAAGCAGTAGTTAAGGCGTATAGC TTAGATACAGAGTCTGGAGTGTCTGTCCAAGCACTGAAAGAGAACAGCTCAGAGGTTGTACAGC CTTTTTTAATGGGTTGTGGAACCAAGGAACCGAAGATCACTCAGCTATGTTTGGCTGCTATTCA GAGACTCATGTCACATGAAGTCGTGTCTGAGACTGCAGCTGGAAATATAATTAACATGCTTTGG CAGCTAATGGAGAATAGTCTTGAAGAACTTAAGCTACTTCAAACAGTTCTTGTTCTTTTAACAA CCAATACAGTAGTTCATGATGAGGCACTTTCTAAGGCAATCGTTCTTTGTTTTCGACTACACTT CACAAAAGATAATATTACAAATAATACAGCTGCTGCTACAGTGCGACAAGTTGTTACTGTTGTT TTTGAGAGGATGGTTGCTGAAGATGAACGACACAGAGATATTATAGAACAACCAGTACTGGTAC AAGGAAATAGTAACAGAAGATCTGTCAGTACCCTCAAACCTTGTGCTAAAGATGCATATATGCT TTTCCAGGATCTTTGTCAGTTGGTTAATGCTGATGCTCCTTATTGGCTAGTGGGCATGACAGAA ATGACTCGGACGTTTGGCCTCGAATTACTTGAGTCAGTCCTCAATGATTTTCCGCAGGTCTTTT TACAACACCAAGAATTTAGTTTCCTCCTCAAAGAAAGGGTATGTCCTCTTGTGATAAAGCTCTT TTCTCCAAATATAAAGTTCAGACAAGGTTCCAGCACCTCATCTTCTCCAGCACCAGTTGAAAAA CCATATTTTCCTATCTGCATGCGTTTGCTGAGAGTAGTATCTGTTCTGATTAAGCAGTTTTACA GTCTTTTGGTAACTGAATGTGAGATATTTCTGTCACTTCTGGTGAAATTTCTGGATGCAGATAA ACCACAGTGGCTACGAGCTGTTGCGGTGGAATCAATACACAGATTCTGTGTGCAGCCTCAACTA TTAAGGTCATTTTGTCAGTCCTATGATATGAAACAGCATTCTACCAAGGTTTTTCGTGATATTG TAAATGCACTGGGATCTTTTATACAGTCCTTGTTTCTTGTCCCCCCTACTGGAAATCCTGCAAC AAGCAACCAAGCTGATGAAGCTGCCACTGAGAATATTTTAAAAGCTGAACTGACTATGGCTGCT CTTTGTGGAAGACTGGGCCTTGTAACTTCAAGAGATGCCTTTATAACTGCAATATGCAAAGGTT CCCTGCCTCCCCATTATGCTCTTACTGTATTGAATACCACCACTGCAGCTACACTTTCCAACAA ATCATATTCCGTTCAGGGCCAAAGTGTTATGATGATAAGTCCATCAAGTGAATCTCACCAACAA GTTGTGGCAGTGGGTCAACCTTTAGCAGTCCAGCCTCAAGGGACAGTAATGCTGACTTCCAAAA ATATCCAGTGTATGAGGACTTTACTTAACTTGGCGCATTGCCATGGGGCTGTTCTTGGAACATC ATGGCAACTTGTCTTGGCAACTCTTCAGCATCTTGTGTGGATTCTGGGATTAAAGCCTAGTAGT GGCGGTGCCTTGAAACCTGGGAGAGCTGTAGAAGGACCCAGTACAGTTCTAACAACAGCAGTGA		

TGACAGATTTACCAGTGATTTCCAATATACTTTCAAGATTGTTTGAAAGCTCACAGTATCTTGA
TGATGTATCACTGCATCATTTAATAAATGCACTTTGCTCCTTGTCTCTAGAAGCAATGGATATG
GCCTATGGAAATAATAAGGAACCATCTCTTTTTGCTGTTGCCAAATTGTTAGAAACTGGTTTAG
TTAATATGCACCGAATAGAAATTCTGTGGAGACCTCTGACTGGCCATCTACTTGAGGTCTGCCA
GCATCCAAACTCTCGAATGAGAGAATGGGGAGCAGAAGCTTTAACTTCTCTTATTAAAGCAGGA
TTAACATTTAACCATGATCCTCCACTCTCACAAAACCAGAGGCTGCAGTTGCTTTTATTGAACC
CGTTAAAGGAGATGTCCAATATTAATCATCCAGATATTCGACTCAAGCAGTTAGAATGCGTGTT
GCAGATTCTGCAGAGTCAGGGAGACAGTCTTGGGCCTGGATGGCCATTAGTGCTTGGAGTCATG
GGAGCAATCAGAAATGATCAAGGAGAATCCTTGATACGAACTGCATTCCAGTGTCTTCAGTTGG
TTGTGACAGATTTTCTACCAACAATGCCTTGTACTTGCCTGCAAATAGTTGTAGATGTTGCAGG
TAGCTTTGGCCTCCATAACCAAGAACTCAATATTAGTTTAACTTCAATAGGTTTATTGTGGAAT
ATTTCAGATTATTTTTTCCAAAGAGGGGAAACTATTGAAAAAGAACTAAATAAGGAAGAGGCAG
CACAGCAAAAGCAGGCAGAAGAGAAAGGAGTTGTTTTAAATCGGCCATTCCACCCTGCACCGCC
ATTTGATTGCTTGTGGTTATGTCTTTATGCAAAATTGGGTGAACTATGTGTGGATCCCCGTCCT
GCTGTCAGGAAGAGTGCAGGGCAAACTCTGTTTTCTACAATTGGTGCGCATGGAACTTTATTAC
AGCATTCAACCTGGCACACTGTTATCTGGAAGGTACTCTTTCATCTACTGGACAGAGTTCGAGA
GTCCTCTACCACTGCAGACAAAGAAAAGATTGAGTCTGGAGGTGGCAATATTCTCATTCATCAT
TCAAGGGACACCGCCGAGAAGCAATGGGCTGAGACGTGGGTATTAACATTGGCTGGAGTAGCAA
GGATCTTCAACACTAGAAGATATTTGCTGCAGCCTTTAGGAGATTTTTCAAGAGCTTGGGATGT
TCTTCTTGACCATATACAGTCAGCAGCACTCAGCAAAAACAATGAAGTATCTCTGGCTGCTCTG
AAAAGCTTCCAGGAAATTTTACAGATTGTGTCCCCTGTCAGAGACTCAGATAAGCCTGAGACAC
CACCTGTAGTTAATGTACCTGTGCCTGTTCTTATAGGGCCCATATCAGGCATGAGCAGGCCATT
TGTAAGAACAGATTCCATTGGAGAAAAACTAGGAAGATATAGTAGCTCTGAGCCACCCATTGTT
ACTGATGAGCTTGAAGATTTGAATCTATGGTGGGCTGCGTGGAATACCTGGTATAGAATTGGAT
CTGAAAGTACTAAGCCTCCTATTACTTTTGATAAACTAACTTTTATTCCTAGCCAGCCTTTTCT
TACAGCTTTAATTCAGATATTTCCAGCTCTCTACCAACACATAAAAACTGGTTTCAATATGGAT
GACTTGCAAAAGTTGGGAGTCATATTGCACAGTGCTATTTCAGTCCCAATAAGTTCAGATGCAT
CCCCTTTTATTCTTCCATCTTATACCGAAGCAGTTTTGACAAGTTTACAGGAAGCTGTACTTAC
AGCTTTAGATGTTCTCCAAAAGGCCATTTGTGTAGGACCAGAAAACATGCAGATAATGTATCCA
GCTATATTTGACCAGTTGTTGGCATTTGTAGAATTTTCCTGTAAACCTCCACAGTATGGACAGC
TGGAAACAAAGCACATTGCAAATGCAAAATATAATCAGATCCAACTATTTGCACCGGCGGAATG
GGTAGCCTTGAATTATGTGCCGTTTGCTGAAAGGTCTTTAGAAGTAGTTGTGGATTTATACCAA
AAAACAGCGTGTCACAAAGCAGTGGTGAATGAGAAAGTGCTCCAGAATATTATTAAGACTCTTA
GGGTTCCTCTCAGTTTGAAGTATTCCTGCCCTTCTGAAAGCACATGGAAACTAGCAGTATCCTC
TCTCCTCAGAGTTCTTTCTATTGGGCTACCTGTTGCCCGGCAGCATGCTTCTTCTGGAAAATTT
GACAGTATGTGGCCAGAACTAGCCAATACTTTTGAAGATTTTCTCTTTACTAAAAGCATACCTC
CAGATAATCTCTCTATTCAAGAGTTTCAAAGAAATGAAAATATTGATGTCGAGGTAGTTCAACT
TATCAGCAATGAGATACTACCTTATGCCAATTTTATTCCTAAGGAATTTGTTGGTCAAATAATG
ACAATGCTTAACAAGGGCTCAATACATTCTCAGTCATCTTCATTTACAGAAGCAGAGATTGATA
TTCGTTTGAGAGAGGAATTTTCTAAAATGTGTTTTGAAACATTACTCCAGTTTTCCTTCAGTAA
TAAAGTCACAACACCTCAAGAAGGCTACATCTCACGAATGGCACTCTCAGTGCTTTTAAAGAGG
TCCCAAGATGTACTACATCGCTATATAGAGGATGAAAGATTAAGTGGTAAATGCCCTCTTCCAA
GGCAACAAGTAACAGAAATTATATTTGTTTTAAAAGCAGTCAGTACTCTTATTGATTCACTTAA
GAAAACTCAGCCTGAGAATGTTGATGGAAATACCTGGGCACAAGTAATTGCCTTATACCCAACT
TTAGTAGAATGCATCACCTGTTCTTCTTCAGAAGTCTGTTCTGCACTTAAAGAGGCACTAGTTC
CTTTTAAGGATTTCATGCAGCCACCAGCATCCAGAGTTCAAAATGGAGAATCT**TGA**CCGGCTAC
AATATATTTGAAAGCAGGAAGATAGTCTAAAAAATGTTTGCTCCTAATTGAGCACGAGGTGAGA
AGGACATTTCTTACTGCAAATAATTCTTGGCAGCTGTTGTTGGCCTCCTTTAAATTCTACTTAC
CTGAGTTCAGTAATTCATATTACAGGCTTGCACATCAACAAAGGCTCCTGAATGAACAGCAGTG
TAAGGCTTTAATAAATTAAACTGATGGGAGGGATAATTAACACTACAGTATACATGCTACCATA
TCTCCAGTTGGTGATTTAAAGTGAGCTTATGTACAGTTTGTGGTGTATGTGTTAATGATGTACT
TTTTAAAAAGAAAGAAGAGATATTTCAATTCAGTCAGATTTATTAGTCTGGTGTTTTTGCACCC
TTTTTCAAGTACAAAATCGTACTAGAATTTTATGCAAGATGGTACTGTAACATTCCATATTATC
TATAACCAGCCTTTGTTAACAAAGGGAACTGATATACTTGTGTGTATAATAAATGGTACAGTTC
TGTAAAAAAAAAAAAAAAAAAAATAGTGCATTTATTTAAATTTTAAAAGTATTGATAATGTTAA
ATGCTTAAAGCTCTATTTATTATTAATACAAAATTGTTTGCTTACATTTTTACTTATAATTTGC
CTTCATATGTGGCGGATAAGCTCACCATATGATCATGCAGTTAGCTTCATGCTTATTTTAAATG
TATTATTAGTGACCATTAAACATCTGACCAGTAAGGTCATGTGAACACAGCAGCAAATAGTTTA
TGATTTGCTGATTTTGGAGCTTTGAAATATAGGTTCTTAATACATTGATACATATTGTAGCACT
ATGACTTCATCATACCTCATTTCTTTAAACAGCTCTCCAAGCTTTCACTGAAGTCTGTCTGTTT
TTTATATTGGCTGTCTGGATTTTAAAGACTTTTCATATTTTATATTTCTACTGATTTTGTTTCC
CCTAACAACATTTGTCACTGTCTTTGAATTATGACCCAGGCAAGATGATTTCAGATTTTCTAAA
ATCTTGCCTGTGAGGTTTTGTTCATATCAGTGCTTCATTTTGTAATGTCTTCTCAAGAAAAATA
CCTATGTTAACTCACAAGTATAAAATATGTGTGTATTATAAAACAATGAAAAGTGTATTTTTGG
AGATAGTCAAGCATTTAGAAGTGCAGTGAACTTGCTGTCACGGAGTAAAATGCTAATTATGTTT
CACTTTCCTAGCCTAGTGAAAAAGAAAAGTGCTCTTGAGTACAATACCTTAATTATTTCTTAAA
ATACTGACTTTGACCTAGCTCACTGTATTTTTTATTTAATGGATTATGGATTACAGTATTTTTC

	TTCTGAGTTAAATTTTCATAATTTATGTGAAGACACAAAGATGTTTAAAACAATGATTATTCAT AAGAAATCATGATGGTCTCAGTATTATTTTAGTGTATTGGAAGGTCTTTGATCTTAATAGAATT TATAAATTTCAGCTTCTCCAGAATAATCATAAAACTGCAAAAAGATATTATAATTGAGTCATGA TTGAGATACAGTTTTGAGGCTATTATAATTGTATAATTATTTAATTTGCATTATCTGTAAAATG TAGTAAGGTCTTTGAGGGGATATTTTTTATTTACATGAATTACTGAATTTCTATTTTATTATTT CACCTAAAATTAAGGTAAAATATGGCATTTCATAAGTTCTGCTTTCAGCATTTTCCTTAAAGTT GTAAAAAATCAAGCTATGTACTTATTTTCTATATTTGGGTGTGTTAAATTGAGGATTAGAAAAA TCCACATAATCACTGATAAAGCATTGAAACAGAATAACCCAAGGGTAGTGTACCGATTCAGTAA CATGTTAAAAATATTGCTATGCATTTATTCAAAGGAAAATGGTCTATTCTTGAGAAATAAAAGA TCAGTTGCAATTAGGATAATTAAATAGTTAAATATGAAAAAAAAGTATGCAATATACATTTATA TGAACCAAAGCTTGCTTTATCAGGACCATGCCCTACAGTTCAAAACATAAACATAGTGAATGTG TTAATATCATATAATAAGGTAATAAATGCCAGTCTTAGTGTGAAGCAAGTGGGTGGCCCCCTTG GTAGTATAATTGGACAGGATTTTCCTCCAGAATATTTCCTGTCACCCTCCAAGAGTCACTACAG TAATTGATTGCTGGCATGGAACACATTGCCCTTGTCTTGTTAGTATGAATTGGGTTCTCCAGTG ACTAGAAGAACTGGGGTGTGTGAAAGTATTCGATGCCAGGAGATTCAAAAAGGAAGCTCTCAAA GATAAGATCATTTTATGGCACAATTGAGTCTATAACCAGCCCTTTAAGCAGTAGTAAAAATGTC CTTTGTCATACTTACTAGAAATACTATGAGTTTTTTTTTTTTTTTTCATTTGAGACGGAGTCTC GCTCTGTCACCCAGGCTGGAATGCAGTGGCACGATCTTGGCTCACTGCAACCTCTGCTTCCCGG GTTCAAGCAGTTTCCTGCTTCAGACTTCCAAGTAGCTGGGATTACAGACATGCCACCATGCCAG GCTAATTTTTTTAATATTTTTAGTAGAGATGGGGTTTCACCATTTTGGCCATTCTAGTCTTGAA CTCCTGACCTCAGGTGATCTGCCCGCCTTGGCCTCCCAAAGTGCTGACATTACAGGCCTGAGCC ACTGCGCCCAGCCAATACCATGAGTTTTAAGCCTCACATCGTCACTTGCTGTCACTGCCAGTGC CTGTTTTATTCATATTGCTGGACAACAGACATATGCCACCAATTGTATGATTAATAAAGTCTTT TTCTGGCCATTTTGTCCATTATAAAGGAAATAACTAATTGTTAACTTGCATAGAATACTTCTTA GTTTCCTATGCTACCACCACTGCCAAGGGAGAAAAAAATACATCATTTTGTAATGTCTTTAGTA TTTCTTTATAACTAGTGTTAAGGTTTTGTTAATTTTATTGTATACATTTGTAACATTTATTAGG AGCCTTTTAGGTTCCAAAACAAACAAAAGGCATAAAAAAGTCTAGCTTAGAACCACTTTTCACT TGCTTTCATTTTTAATTTTATTCACTTAACAGCTAACATCTTTCTTGTTTCTTGTTTTTTCCAT TATATGGTTATCGATTCAACTCTTGCTATATTCCTTAAATTTGTATGTATCATCAGAAGAAAGA GATGAACAATTTAGTGTAGATATTTTATTCTGGAGAATAATATTCAATTAAATTATTTCTACAG CAGGCCAGTAACAACTAGATTATTTGTCCTTTCTCAGTATAATTTTAAAGAGCATTTTGTTTTA TTGTCACAATTTGGTACCACTAGTCCCAGGTAACCATTGGGCCAAAGGATCAGTTGAGAAACAG TTAAGGATGAATTAGCATAAGTTATGGAACAGTGTTAGAAAACAACTCAAAAGTATATTCTTTA TTAATGAGGTGGTCATTATTACATTTGTGTCAATGAAGGGCAGTGTAGTTATTTTAAAATGACT AATATTTTCTCCCCAAATACAGAATAATTCAGATGGGCAACCAAGTTTTCAAGAGACTGCTGTA GGTGAAGTCTGTCTAGCCAAGGCAGAACACTTACAGGAGTCCCTAACTGTGCCACCCTTGGAAT GGGTTAGTGTACAGGCTCAGAATATTGTGGATTACAGTTTTTCAGAGAAAACTACCACAGATGT AGACAAAAATGATCTCTGAAAGCATTGCCAGCAGCCAGGTATGTTCCTTAGATTTCCACTTAGG TTTGGCATTTTGGCAGATAAGCTAATCTTGTATAAAGCATCACATTTTACTATGCTTAGTGTTC CTGGGTTGTATTTATCTACATTATTAGAGGGAATTTTTATTTTAAAAAAATTGTCATTCATGAG AAGAATGGGAGTTCATGCCACATAGTATTTTACCAATTTATATAAAGTGGGAAAAGTCTTTAAT ACTTCATGATCACTTGAATTAAAGTTTTTGTATCTCTGGAAAGTAGAATAGTGCTTTCATTTGA ATGAAAAGTGTTTATAGATTCAGAAAGAGAGATGATATCTTGTATCTTGATTTATATACAGACC ATTTCAGAGGAAGTTAACATGTCTTACAAATTCCAATACTTTCTAATGCTCTAACAGTGTTGGC TATTTTAAACGAACATGTGGCAAGTTCTATATGAATATTCTTGGTCATCTCGACTAATTCTGAG GCAATGATGGACAGAGATGCTACTTCTTATTTAACTCTAGGCATGTTGACTTTTCAAAGCGGTT TCCTTATTTCTAAACAGAGATGATGATCAATGAGTTACTAATTCTTTAGAGGAAAAAATGAATA ATTTGAGTGTGGAGTTGATTTTAATGACAGGGTAATTCAAGTTGTTTGATAAATTTATTACTAT ATTGTAAGAGAGATCTTTGACCATTTTTCTTCCTTTTTCTTGGACATCACTTTCTTCCCTCCCC TTCTCTCTTTTATGTTTTTATCCTTGTTAAATTTTATGTTTACGTTACCATCTTTCTTATACTT TCCCCTGATTTTTCTCTTTTAATTCCTCTTTCATTCTCTGCCTCTTCTCTTTCAGCTCTTTCTC TAATTGTGCCTATTCCTTGTTCATAAGAAGTGGAGCTGTTAGTGGTAGAACCACTGCTCATGGT TCTACCACTACAAAGTGGAAAGTAGAAATACTTTGCACTTTGGCCACTGTTGCGTTTTTGCCAA GGTAAAGTTCCCCTGCCATTTTGAAATAGCTGAACAAGTTAAAGTAATATGTTCCAAAAACTGG AAGTGCCATAAAAAACTAAAAATAAAAAAAAATTGTGACTAAAAAAAAAAAAAAA
	ORF Start: ATG at 1 | | ORF Stop: TGA at 4726
	SEQ ID NO: 40 | 1575 aa | MW at 175348.8kD
NOV8a, CG179299-01 Protein Sequence	MSGTSSPEAVKKLLENMQSDLRALSLECKKKFPPVKEFVAEDGGEPKDRVLGEKHGKAVVKAYS LDTESGVSVQALKENSSEVVQPFLMGCGTKEPKITQLCLAAIQRLMSHEVVSETAAGNIINMLW QLMENSLEELKLLQTVLVLLTTNTVVHDEALSKAIVLCFRLHFTKDNITNNTAAATVRQVVTVV FERMVAEDERHRDIIEQPVLVQGNSNRRSVSTLKPCAKDAYMLFQDLCQLVNADAPYWLVGMTE MTRTFGLELLESVLNDFPQVFLQHQEFSFLLKERVCPLVIKLFSPNIKFRQGSSTSSSPAPVEK PYFPICMRLLRVVSVLIKQFYSLLVTECEIFLSLLVKFLDADKPQWLRAVAVESIHRFCVQPQL LRSFCQSYDMKQHSTKVFRDIVNALGSFIQSLFLVPPTGNPATSNQADEAATENILKAELTMAA LCGRLGLVTSRDAFITAICKGSLPPHYALTVLNTTTAATLSNKSYSVQGQSVMMISPSSESHQQ VVAVGQPLAVQPQGTVMLTSKNIQCMRTLLNLAHCHGAVLGTSWQLVLATLQHLVWILGLKPSS

	GGALKPGRAVEGPSTVLTTAVMTDLPVISNILSRLFESSQYLDDVSLHHLINALCSLSLEAMDM AYGNNKEPSLFAVAKLLETGLVNMHRIEILWRPLTGHLLEVCQHPNSRMREWGAEALTSLIKAG LTFNHDPPLSQNQRLQLLLLNPLKEMSNINHPDIRLKQLECVLQILQSQGDSLGPGWPLVLGVM GAIRNDQGESLIRTAFQCLQLVVTDFLPTMPCTCLQIVVDVAGSFGLHNQELNISLTSIGLLWN ISDYFFQRGETIEKELNKEEAAQQKQAEEKGVVLNRPFHPAPPFDCLWLCLYAKLGELCVDPRP AVRKSAGQTLFSTIGAHGTLLQHSTWHTVIWKVLFHLLDRVRESSTTADKEKIESGGGNILIHH SRDTAEKQWAETWVLTLAGVARIFNTRRYLLQPLGDFSRAWDVLLDHIQSAALSKNNEVSLAAL KSFQEILQIVSPVRDSDKPETPPVVNVPVPVLIGPISGMSRPFVRTDSIGEKLGRYSSSEPPIV TDELEDLNLWWAAWNTWYRIGSESTKPPITFDKLTFIPSQPFLTALIQIFPALYQHIKTGFNMD DLQKLGVILHSAISVPISSDASPFILPSYTEAVLTSLQEAVLTALDVLQKAICVGPENMQIMYP AIFDQLLAFVEFSCKPPQYGQLETKHIANAKYNQIQLFAPAEWVALNYVPFAERSLEVVVDLYQ KTACHKAVVNEKVLQNIIKTLRVPLSLKYSCPSESTWKLAVSSLLRVLSIGLPVARQHASSGKF DSMWPELANTFEDFLFTKSIPPDNLSIQEFQRNENIDVEVVQLISNEILPYANFIPKEFVGQIM TMLNKGSIHSQSSSFTEAEIDIRLREEFSKMCFETLLQFSFSNKVTTPQEGYISRMALSVLLKR SQDVLHRYIEDERLSGKCPLPRQQVTEIIFVLKAVSTLIDSLKKTQPENVDGNTWAQVIALYPT LVECITCSSSEVCSALKEALVPFKDFMQPPASRVQNGES

Further analysis of the NOV8a protein yielded the following properties shown in Table 8B.

Table 8B. Protein Sequence Properties NOV8a

SignalP analysis:	No Known Signal Sequence Predicted

PSORT II analysis:	PSG: a new signal peptide prediction method
	N-region: length 11; pos.chg 1; neg.chg 1
	H-region: length 0; peak value -8.54
	PSG score: -12.94
	
	GvH: von Heijne's method for signal seq. recognition
	GvH score (threshold: -2.1): -9.88
	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: 2
	INTEGRAL Likelihood = -2.39 Transmembrane 343 - 359
	INTEGRAL Likelihood = -2.07 Transmembrane 1319 -1335
	PERIPHERAL Likelihood = 0.53 (at 138)
	ALOM score: -2.39 (number of TMSs: 2)
	
	MTOP: Prediction of membrane topology (Hartmann et al.)
	Center position for calculation: 350
	Charge difference: 2.0 C(1.0) - 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): 5.70
	Hyd Moment(95): 3.66 G content: 1
	D/E content: 2 S/T content: 4
	Score: -6.66
	
	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:   8.9%
      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: 89

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

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

Final Results (k = 9/23):

        22.2 %: vacuolar
        22.2 %: nuclear
        22.2 %: endoplasmic reticulum
        11.1 %: Golgi
        11.1 %: mitochondrial
        11.1 %: vesicles of secretory system

>> prediction for CG179299-01 is vac (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 8C.

Table 8C. Geneseq Results for NOV8a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV8a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABP69631	Human polypeptide SEQ ID NO 1678 - Homo sapiens, 1216 aa. [WO200270539-A2, 12-SEP-2002]	381..1575 28..1216	1154/1196 (96%) 1164/1196 (96%)	0.0
AAB94172	Human protein sequence SEQ ID NO:14478 - Homo sapiens, 507 aa. [EP1074617-A2, 07-FEB-2001]	1063..1575 1..507	505/513 (98%) 506/513 (98%)	0.0
AAY02367	Polypeptide identified by the signal sequence trap method - Homo sapiens, 521 aa. [WO9918126-A1, 15-APR-1999]	432..930 22..521	495/500 (99%) 496/500 (99%)	0.0
ABG63631	Human albumin fusion protein #306 - Homo sapiens, 477 aa. [WO200177137-A1, 18-OCT-2001]	432..888 22..477	434/457 (94%) 437/457 (94%)	0.0
AAE03061	Human gene 10 encoded secreted protein HHFHI76, SEQ ID NO:95 - Homo sapiens, 477 aa. [WO200132676-A1, 10-MAY-2001]	432..888 22..477	434/457 (94%) 437/457 (94%)	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 8D.

Table 8D. Public BLASTP Results for NOV8a				
Protein Accession Number	**Protein/Organism/Length**	**NOV8a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8N3I5	Hypothetical protein - Homo sapiens (Human), 1500 aa (fragment).	381..1575 312..1500	1155/1196 (96%) 1164/1196 (96%)	0.0
Q8NHE2	SF21 protein - Homo sapiens (Human), 1717 aa.	381..1575 529..1717	1155/1196 (96%) 1164/1196 (96%)	0.0
Q8K4V3	SF21 protein - Mus musculus (Mouse), 1708 aa.	432..1575 571..1708	1087/1146 (94%) 1113/1146 (96%)	0.0
Q9UPW1	KIAA1040 protein - Homo sapiens (Human), 599 aa (fragment).	989..1575 13..599	582/587 (99%) 584/587 (99%)	0.0
Q8NAI0	Hypothetical protein FLJ35327 - Homo sapiens (Human), 586 aa.	989..1575 6..586	576/587 (98%) 578/587 (98%)	0.0

PFam analysis predicts that the NOV8a protein contains the domains shown in the Table 8E.

Table 8E. Domain Analysis of NOV8a			
Pfam Domain	**NOV8a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
DUF489	399..571	51/212 (24%) 98/212 (46%)	0.0049

Example 9. NOV 9, CG50301, humanTENM4.

The NOV9 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 9A.

Table 9A. NOV9 Sequence Analysis			
	SEQ ID NO: 41	1476 bp	
NOV9a, CG50301-07 DNA Sequence	**CAC**CTGAGCGGCCGCGCCACCATGGAGACAGACACACTCCTGCTATGGGTACTGCTGCTCTGGG TTCCAGGTTCCACTGGTGACCACCTGCAGCCGATGGAGGGGCAGATGTATGAGATCACGGAGGA CACAGCCAGCAGTTGGCCTGTGCCAACCGACGTCTCCCTATACCCCTCAGGGGGCACTGGCTTA GAGACCCCTGACAGGAAAGGCAAAGGAACCACAGAAGGAAAGCCCAGTAGTTTCTTTCCAGAGG ACAGTTTCATAGATTCTGGAGAAATTGATGTGGGAAGGCGAGCTTCCCAGAAGATTCCTCCTGG CACTTTCTGGAGATCTCAAGTGTTCATAGACCATCCTGTGCATCTGAAATTCAATGTGTCTCTG GGAAAGGCAGCCCTGGTTGGCATTTATGGCAGAAAAGGCCTCCCTCCTTCACATACACAGTTTG ACTTTGTGGAGCTGCTGGATGGCAGGAGGCTCCTAACCCAGGAGGCGCGGAGCCTAGAGGGGAC CCCGCGCCAGTCTCGGGGAACTGTGCCCCCCTCCAGCCATGAGACAGGCTTCATCCAGTATTTG GATTCAGGAATCTGGCACTTGGCTTTTTACAATGACGGAAAGGAGTCAGAAGTGGTTTCCTTTC TCACCACTGCCATTGAGTCGGTGGATAACTGCCCCAGCAACTGCTATGGCAATGGTGACTGCAT CTCTGGGACCTGCCACTGCTTCCTGGGTTTCCTGGGCCCCGACTGTGGCAGAGCCTCCTGCCCC GTGCTCTGTAGCGGAAATGGCCAATACATGAAAGGCAGATGCTTGTGCCACAGTGGCTGGAAAG GCGCTGAGTGCGATGTGCCCACCAACCAGTGTATCGATGTGGCCTGCAGCAACCATGGCACCTG CATCACGGGCACCTGCATCTGCAACCCTGGCTACAAGGGCGAGAGCTGTGAGGAAGTGGACTGC ATGGACCCCACATGTTCAGGCCGGGGTGTCTGCGTGAGAGGCGAATGCCACTGCTCTGTGGGAT GGGGAGGCACCAACTGCGAGACCCCCAGGGCCACATGCTTAGACCAGTGTTCAGGCCACGGAAC CTTCCTCCCGGACACCGGGCTTTGCAGCTGTGACCCAAGCTGGACTGGACACGACTGTTCTATC GAGATCTGTGCTGCCGACTGTGGTGGCCATGGCGTGTGCGTAGGGGGCACCTGCCGCTGCGAGG ATGGCTGGATGGGGGCAGCCTGCGACCAGCGGGCCTGCCACCCGCGCTGTGCCGAGCATGGGAC CTGCCGCGACGGCAAGTGCGAGTGCAGCCCTGGCTGGAATGGCGAACACTGCACCATCGCTCAC TATCTGGATAGGGTAGTTAAAGAGGGTTGCCCTGGGTTGTGCAATGGCAACGGCAGATGTACCT TAGACCTGAATGGTTGGCACTGCGTCTGCCAGCTGGGCTGGAGAGGAGCTGGCTGT**TGA**CTCGA GGGC		
	ORF Start: at 1		ORF Stop: TGA at 1465
	SEQ ID NO: 42	488 aa	MW at 52251.1kD
NOV9a, CG50301-07 Protein Sequence	HLSGRATMETDTLLLWVLLLWVPGSTGDHLQPMEGQMYEITEDTASSWPVPTDVSLYPSGGTGL ETPDRKGKGTTEGKPSSFFPEDSFIDSGEIDVGRRASQKIPPGTFWRSQVFIDHPVHLKFNVSL GKAALVGIYGRKGLPPSHTQFDFVELLDGRRLLTQEARSLEGTPRQSRGTVPPSSHETGFIQYL DSGIWHLAFYNDGKESEVVSFLTTAIESVDNCPSNCYGNGDCISGTCHCFLGFLGPDCGRASCP VLCSGNGQYMKGRCLCHSGWKGAECDVPTNQCIDVACSNHGTCITGTCICNPGYKGESCEEVDC MDPTCSGRGVCVRGECHCSVGWGGTNCETPRATCLDQCSGHGTFLPDTGLCSCDPSWTGHDCSI EICAADCGGHGVCVGGTCRCEDGWMGAACDQRACHPRCAEHGTCRDGKCECSPGWNGEHCTIAH YLDRVVKEGCPGLCNGNGRCTLDLNGWHCVCQLGWRGAGC		
	SEQ ID NO: 43	8354 bp	

NOV9b, CG50301-01 DNA Sequence	GTTTGTGGATGTGGAGGAGCGCGGGCCGGAGGCC**ATG**GACGTGAAGGAGAGGAAGCCTTACCGC TCGCTGACCCGGCGCCGCGACGCCGAGCGCCGCTACACCAGCTCGTCCGCGGACAGCGAGGAGG GCAAAGCCCCGCAGAAATCGTACAGCTCCAGCGAGACCCTGAAGGCCTACGACCAGGACGCCCG CCTAGCCTATGGCAGCCGCGTCAAGGACATTGTGCCGCAGGAGGCCGAGGAATTCTGCCGCACA GGTGCCAACTTCACCCTGCGGGAGCTGGGGCTGGAAGAAGTAACGCCCCCTCACGGGACCCTGT ACCGGACAGACATTGGCCTGCCCCAATGCGGCTACTCCATGGGGGCTGGCTCTGATGCCGACAT GGAGGCTGACACGGTGCTGTCCCCTGAGCACCCCGTGCGTCTGTGGGGCCGGAGCACACGGTCA GGGCGCAGCTCCTGCCTGTCCAGCCGGGCCAATTCCAATCTCACACTCACCGACACCGAGCATG AAAACACTGAGACTGATCATCCGGGCGGCCTGCAGAACCACGCGCGGCTCCGGACGCCGCCGCC GCCGCTCTCGCACGCCCACACCCCCAACCAGCACCACGCGGCCTCCATTAACTCCCTGAACCGG GGCAACTTCACGCCGAGGAGCAACCCCAGCCCGGCCCCCACGGACCACTCGCTCTCCGGAGAGC CCCCTGCCGGCGGCGCCCAGGAGCCTGCCCACGCCCAGGAGAACTGGCTGCTCAACAGCAACAT CCCCCTGGAGACCAGGAACCTAGGCAAGCAGCCATTCCTAGGGACATTGCAGGACAACCTCATT GAGATGGACATTCTCGGCGCCTCCCGCCATGATGGGGCTTACAGTGACGGGCACTTCCTCTTCA AGCCTGGAGGCACCTCCCCGCTCTTCTGCACCACATCACCAGGGTACCCACTGACGTCCAGCAC AGTGTACTCTCCTCCGCCCCGACCCCTGCCCCGCAGCACCTTCGCCCGGCCGGCCTTTAACCTC AAGAAGCCCTCCAAGTACTGTAACTGGAAGTGCGCAGCCCTGAGCGCCATCGTCATCTCAGCCA CTCTGGTCATCCTGCTGGCATACTTTGTGGCCATGCACCTGTTTGGCCTAAACTGGCACCTGCA GCCGATGGAGGGGCAGATGTATGAGATCACGGAGGACACAGCCAGCAGTTGGCCTGTGCCAACC GACGTCTCCCTATACCCCTCAGGGGGCACTGGCTTAGAGACCCCTGACAGGAAAGGCAAAGGAA CCACAGAAGGAAAGCCCAGTAGTTTCTTTCCAGAGGACAGTTTCATAGATTCTGGAGAAATTGA TGTGGGAAGGCGAGCCTCCCAGAAGATTCCTCCTGGCACTTTCTGGAGATCTCAAGTGTTCATA GACCATCCTGTGCATCTGAAATTCAATGTGTCTCTGGGAAAGGCAGCCCTGGTTGGCATTTATG GCAGAAAAGGCCTCCCTCCTTCACATACACAGTTTGACTTTGTGGAGCTGCTGGATGGCAGGAG GCTCCTAACCCAGGAGGCGCGGAGCCTAGAGGGGACCCCGCGCCAGTCTCGGGGAACTGTGCCC CCCTCCAGCCATGAGACAGGCTTCATCCAGTATTTGGATTCAGGAATCTGGCACTTGGCTTTTT ACAATGACGGAAAGGAGTCAGAAGTGGTTTCCTTTCTCACCACTGCCATTGAGTCGGTGGATAA CTGCCCCAGCAACTGCTATGGCAATGGTGACTGCATCTCTGGGACCTGCCACTGCTTCCTGGGT TTCCTGGGCCCCGACTGTGGCAGAGCCTCCTGCCCCGTGCTCTGTAGCGGAAATGGCCAATACA TGAAAGGCAGATGCTTGTGCCACAGTGGCTGGAAAGGCGCTGAGTGCGATGTGCCCACCAACCA GTGTATCGATGTGGCCTGCAGCAACCATGGCACCTGCATCACGGGCACCTGCATCTGCAACCCT GGCTACAAGGGCGAGAGCTGTGAGGAAGTGGACTGCATGGACCCCACATGTTCAGGCCGGGGTG TCTGCGTGAGAGGCGAATGCCATTGCTTTGTGGGATGGGGAGGCACCAACTGCGAGACCCCCAG GGCCACATGCTTAGACCAGTGTTCAGGCCACGGAACCTTCCTCCCGGACACCGGGCTTTGCAGC TGTGACCCAAGCTGGACTGGACACGACTGTTCTATCGAGATCTGTGCTGCCGACTGTGGTGGCC ATGGCGTGTGCGTAGGGGGCACCTGCCGCTGCGAGGATGGCTGGATGGGGGCAGCCTGCGACCA GCGGGCCTGCCACCCGCGCTGTGCCGAGCATGGGACCTGCCGCGACGGCAAGTGCGAGTGCAGC CCTGGCTGGAATGGCGAACACTGCACCATCGCTCACTATCTGGATAGGGTAGTTAAAGAGGGTT GCCCTGGGTTGTGCAATGGCAACGGCAGATGTACCTTAGACCTGAATGGTTGGCACTGCGTCTG CCAGCTGGGCTGGAGAGGAGCTGGCTGTGACACTTCCATGGAGACTGCCTGCGGTGACAGCAAA GACAATGATGGAGATGGCCTGGTGGACTGCATGGACCCTGACTGCTGCCTCCAGCCCCTGTGCC ATATCAACCCGCTGTGCCTTGGCTCCCCTAACCCTCTGGACATCATCCAGGAGACACAGGTCCC TGTGTCACAGCAGAACCTACACTCCTTCTATGACCGCATCAAGTTCCTCGTGGGCAGGGACAGC ACGCACATAATCCCCGGGGAGAACCCCTTTGATGGAGGGCATGCTTGTGTTATTCGTGGCCAAG TGATGACATCAGATGGAACCCCCCTGGTTGGTGTGAACATCAGTTTTGTCAATAACCCTCTCTT TGGATATACAATCAGCAGGCAAGATGGCAGCTTTGACTTGGTGACAAATGGCGGCATCTCCATC ATCCTGCGGTTCGAGCGGGCACCTTTCATCACACAGGAGCACACCCTGTGGCTGCCATGGGATC GCTTCTTTGTCATGGAAACCATCATCATGAGACATGAGGAGAATGAGATTCCCAGCTGTGACCT GAGCAATTTTGCCCGCCCCAACCCAGTCGTCTCTCCATCCCCACTGACGTCCTTCGCCAGCTCC TGTGCAGAGAAAGGCCCCATTGTGCCGGAAATTCAGGCTTTGCAGGAGGAAATCTCTATCTCTG GCTGCAAGATGAGGCTGAGCTACCTGAGCAGCCGGACCCCTGGCTACAAATCTGTCCTGAGGAT CAGCCTCACCCACCCGACCATCCCCTTCAACCTCATGAAGGTGCACCTCATGGTAGCGGTGGAG GGCCGCCTCTTCAGGAAGTGGTTCGCTGCAGCCCCAGACCTGTCCTATTATTTCATTTGGGACA AGACAGACGTCTACAACCAGAAGGTGTTTGGGCTTTCAGAAGCCTTTGTTTCCGTGGGTTATGA ATATGAATCCTGCCCAGATCTAATCCTGTGGGAAAAAAGAACAACAGTGCTGCAGGGCTATGAA ATTGACGCGTCCAAGCTTGGAGGATGGAGCCTAGACAAACATCATGCCCTCAACATTCAAAGTG GTATCCTGCACAAAGGGAATGGGGAGAACCAGTTTGTGTCTCAGCAGCCTCCTGTCATTGGGAG CATCATGGGCAATGGGCGCCGGAGAAGCATCTCCTGCCCCAGCTGCAACGGCCTTGCTGACGGC AACAAGCTCCTGGCCCCAGTGGCCCTCACCTGTGGCTCTGACGGGAGCCTCTATGTGGGTGATT TCAACTACATTAGAAGGATCTTCCCCTCTGGAAATGTCACCAACATCCTAGAGCTGAGGAATAA AGATTTCAGACATAGTCACAGTCCAGCACACAAATACTACCTGGCCACAGACCCCATGAGTGGG GCCGTCTTCCTTTCTGACAGCAACAGCCGGCGGGTCTTTAAAATCAAGTCCACTGTGGTGGTGA AGGACCTTGTCAAGAACTCTGAGGTGGTTGCGGGGACAGGTGACCAGTGCCTCCCCTTTGATGA CACTCGCTGCGGGGATGGTGGGAAGGCCACAGAAGCCACACTCACCAATCCCAGGGGTATTACA GTGGACAAGTTTGGGCTGATCTACTTCGTGGATGGCACCATGATCAGACGCATCGATCAGAATG GGATCATCTCCACCCTGCTCGGCTCTAATGATCTCACATCAGCCCGGCCACTCAGCTGTGATTC TGTCATGGATATTTCCCAGGTAAGACTGGAGTGGCCCACAGACTTAGCCATCAACCCAATGGAC

	AACTCACTTTATGTCCTCGACAACAATGTGGTCCTGCAAATCTCTGAAAACCACCAGGTGCGCA TTGTCGCCGGGAGGCCCATGCACTGCCAGGTCCCTGGCATTGACCACTTCCTGCTAAGCAAGGT GGCCATCCACGCAACCCTGGAGTCAGCCACCGCTTTGGCTGTTTCACACAATGGGGTCCTGTAT ATTGCTGAGACTGATGAGAAAAAGATCAACCGCATCAGGCAGGTCACCACTAGTGGAGAGATCT CACTCGTTGCTGGGGCCCCCAGTGGCTGTGACTGTAAAAATGATGCCAACTGTGATTGTTTTTC TGGAGACGATGGTTATGCCAAGGATGCAAAGTTAAATACCCCATCTTCCTTGGCTGTGTGTGCT GATGGGGAGCTCTACGTGGCCGACCTTGGGAACATCCGAATTCGGTTTATCCGGAAGAACAAGC CTTTCCTCAACACCCAGAACATGTATGAGCTGTCTTCACCAATTGACCAGGAGCTCTATCTGTT TGATACCACCGGCAAGCACCTGTACACCCAAAGCCTGCCCACAGGAGACTACCTGTACAACTTC ACCTACACTGGGGACGGCGACATCACACTCATCACAGACAACAATGGCAACATGGTAAATGTCC GCCGAGACTCTACTGGGATGCCCCTCTGGCTGGTGGTCCCAGATGGCCAGGTGTACTGGGTGAC CATGGGCACCAACAGTGCACTCAAGAGTGTGACCACACAAGGACACGAGTTGGCCATGATGACA TACCATGGCAATTCCGGCCTTCTGGCAACCAAAAGCAATGAAAACGGATGGACAACATTTTATG AGTACGACAGCTTTGGCCGCCTGACAAATGTGACCTTCCCTACTGGCCAGGTGAGCAGTTTCCG AAGTGATACAGACAGTTCAGTGCATGTCCAGGTAGAGACCTCCAGCAAGGATGATGTCACCATA ACCACCAACCTGTCTGCCTCAGGCGCCTTCTACACACTGCTGCAAGACCAAGTCCGGAACAGCT ACTACATCGGGGCCGATGGCTCCTTGCGGCTGCTGCTGGCCAACGGCATGGAGGTGGCGCTGCA GACTGAGCCCCACTTGCTGGCTGGCACCGTCAACCCCACCGTGGGCAAGAGGAATGTCACGCTG CCCATCGACAACGGCCTCAACCTGGTGGAGTGGCGCCAGCGCAAAGAGCAGGCTCGGGGCCAGG TCACTGTCTTTGGGCGCCGGCTGCGGGTGCACAACCGAAATCTCCTATCTCTGGACTTTGATCG CGTAACACGCACAGAGAAGATCTATGATGACCACCGCAAGTTCACCCTTCGGATTCTGTACGAC CAGGCGGGGCGGCCCAGCCTCTGGTCACCCAGCAGCAGGCTGAATGGTGTCAACGTGACATACT CCCCTGGGGGTTACATTGCTGGCATCCAGAGGGGCATCATGTCTGAAAGAATGGAATACGACCA GGCGGGCCGCATCACATCCAGGATCTTCGCTGATGGGAAGACATGGAGCTACACATACTTAGAG AAGTCCATGGTGCTGCTACTACACAGCCAGAGGCAGTATATCTTTGAGTTCGACAAGAATGACC GCCTCTCTTCTGTGACGATGCCCAACGTGGCGCGGCAGACACTAGAGACCATCCGCTCAGTGGG CTACTACAGAAACATCTATCAGCCCCCTGAGGGCAATGCCTCAGTCATACAGGACTTCACTGAG GATGGGCACCTCCTTCACACCTTCTACCTGGGCACTGGCCGCAGGGTGATATACAAGTATGGCA AACTGTCAAAGCTGGCAGAGACGCTCTATGACACCACCAAGGTCAGTTTCACCTATGACGAGAC GGCAGGCATGCTGAAGACCATCAACCTACAGAATGAGGGCTTCACCTGCACCATCCGCTACCGT CAGATTGGGCCCCTGATTGACCGACAGATCTTCCGCTTCACTGAGGAAGGCATGGTCAACGCCC GTTTTGACTACAACTATGACAACAGCTTCCGGGTGACCAGCATGCAGGCTGTGATCAACGAGAC CCCACTGCCCATTGATCTCTATCGCTATGATGATGTGTCAGGCAAGACAGAGCAGTTTGGGAAG TTTGGTGTCATTTACTATGACATTAACCAGATCATCACCACAGCTGTCATGACCCACACCAAGC ATTTTGATGCATATGGCAGGATGAAGGAAGTGCAGTATGAGATCTTCCGCTCGCTCATGTACTG GATGACCGTCCAGTATGATAACATGGGGCGAGTAGTGAAGAAGGAGCTGAAGGTAGGACCCTAC GCCAATACCACTCGCTACTCCTATGAGTATGATGCTGACGGCCAGCTGCAGACAGTCTCCATCA ATGACAAGCCACTCTGGCGCTACAGCTACGACCTCAATGGGAACCTGCACTTACTGAGCCCTGG GAACAGTGCACGGCTCACACCACTACGGTATGACATCCGCGACCGCATCACTCGGCTGGGTGAC GTGCAATACAAGATGGATGAGGATGGCTTCCTGAGGCAGCGGGGCGGTGATATCTTTGAGTACA ACTCAGCTGGCCTGCTCATCAAGGCCTACAACCGGGCTGGCAGCTGGAGTGTCAGGTACCGCTA CGATGGCCTGGGGCGGCGCGTGTCCAGCAAGAGCAGCCACAGCCACCACCTGCAGTTCTTCTAT GCAGACCTGACCAACCCCACCAAGGTCACCCACCTGTACAACCACTCCAGCTCTGAGATCACCT CCCTCTACTACGACTTGCAAGGACACCTCTTTGCCATGGAGCTGAGCAGTGGTGATGAGTTTTA CATAGCTTGTGACAACATCGGGACCCCTCTTGCTGTCTTTAGTGGAACAGGTTTGATGATCAAG CAAATCCTGTACACAGCCTATGGGGAGATCTACATGGATACCAACCCCAACTTTCAGATCATCA TAGGCTACCATGGTGGCCTCTATGATCCACTCACCAAGCTTGTCCACATGGGCCGGCGAGATTA TGATGTGCTGGCCGGACGCTGGACTAGCCCAGACCACGAGCTGTGGAAGCACCTTAGTAGCAGC AACGTCATGCCTTTTAATCTCTATATGTTCAAAAACAACAACCCCATCAGCAACTCCCAGGACA TCAAGTGCTTCATGACAGATGTTAACAGCTGGCTGCTCACCTTTGGATTCCAGCTACACAACGT GATCCCTGGTTATCCCAAACCAGACATGGATGCCATGGAACCCTCCTACGAGCTCATCCACACA CAGATGAAAACGCAGGAGTGGGACAACAGCAAGTCTATCCTCGGGGTACAGTGTGAAGTACAGA AGCAGCTCAAGGCCTTTGTCACCTTAGAACGGTTTGACCAGCTCTATGGCTCCACAATCACCAG CTGCCAGCAGGCTCCAAAGACCAAGAAGTTTGCATCCAGCGGCTCAGTCTTTGGCAAGGGGGTC AAGTTTGCCTTGAAGGATGGCCGAGTGACCACAGACATCATCAGTGTGGCCAATGAGGATGGGC GAAGGGTTGCTGCCATCTTGAACCATGCCCACTACCTAGAGAACCTGCACTTCACCATTGATGG GGTGGATACCCATTACTTTGTGAAACCAGGACCTTCAGAAGGTGACCTGGCCATCCTGGGCCTC AGTGGGGGGCGGCGAACCCTGGAGAATGGGGTCAACGTCACTGTGTCCCAGATCAACACAGTAC TTAATGGCAGGACTAGACGCTACACAGACATCCAGCTCCAGTACGGGGCACTGTGCTTGAACAC ACGCTACGGGACAACGTTGGATGAGGAGAAGGCACGGGTCCTGGAGCTGGCCCGGCAGAGAGCC GTGCGCCAAGCGTGGGCCCGCGAGCAGCAGAGACTGCGGGAAGGGGAGGAAGGCCTGCGGGCCT GGACAGAGGGGGAGAAGCAGCAGGTGCTGAGCACAGGGCGGGTGCAAGGCTACGACGGCTTTTT CGTGATCTCTGTCGAGCAGTACCCAGAACTGTCAGACAGCGCCAACAACATCCACTTCATGAGA CAGAGCGAGATGGGCCGGAGG**TGA**CAGAGAGGAC
	ORF Start: ATG at 35 ORF Stop: TGA at 8342
	SEQ ID NO: 44 2769 aa MW at 308023.4kD

NOV9b, CG50301-01 Protein Sequence	MDVKERKPYRSLTRRRDAERRYTSSSADSEEGKAPQKSYSSSETLKAYDQDARLAYGSRVKDIV PQEAEEFCRTGANFTLRELGLEEVTPPHGTLYRTDIGLPQCGYSMGAGSDADMEADTVLSPEHP VRLWGRSTRSGRSSCLSSRANSNLTLTDTEHENTETDHPGGLQNHARLRTPPPPLSHAHTPNQH HAASINSLNRGNFTPRSNPSPAPTDHSLSGEPPAGGAQEPAHAQENWLLNSNIPLETRNLGKQP FLGTLQDNLIEMDILGASRHDGAYSDGHFLFKPGGTSPLFCTTSPGYPLTSSTVYSPPPRPLPR STFARPAFNLKKPSKYCNWKCAALSAIVISATLVILLAYFVAMHLFGLNWHLQPMEGQMYEITE DTASSWPVPTDVSLYPSGGTGLETPDRKGKGTTEGKPSSFFPEDSFIDSGEIDVGRRASQKIPP GTFWRSQVFIDHPVHLKFNVSLGKAALVGIYGRKGLPPSHTQFDFVELLDGRRLLTQEARSLEG TPRQSRGTVPPSSHETGFIQYLDSGIWHLAFYNDGKESEVVSFLTTAIESVDNCPSNCYGNGDC ISGTCHCFLGFLGPDCGRASCPVLCSGNGQYMKGRCLCHSGWKGAECDVPTNQCIDVACSNHGT CITGTCICNPGYKGESCEEVDCMDPTCSGRGVCVRGECHCFVGWGGTNCETPRATCLDQCSGHG TFLPDTGLCSCDPSWTGHDCSIEICAADCGGHGVCVGGTCRCEDGWMGAACDQRACHPRCAEHG TCRDGKCECSPGWNGEHCTIAHYLDRVVKEGCPGLCNGNGRCTLDLNGWHCVCQLGWRGAGCDT SMETACGDSKDNDGDGLVDCMDPDCCLQPLCHINPLCLGSPNPLDIIQETQVPVSQQNLHSFYD RIKFLVGRDSTHIIPGENPFDGGHACVIRGQVMTSDGTPLVGVNISFVNNPLFGYTISRQDGSF DLVTNGGISIILRFERAPFITQEHTLWLPWDRFFVMETIIMRHEENEIPSCDLSNFARPNPVVS PSPLTSFASSCAEKGPIVPEIQALQEEISISGCKMRLSYLSSRTPGYKSVLRISLTHPTIPFNL MKVHLMVAVEGRLFRKWFAAAPDLSYYFIWDKTDVYNQKVFGLSEAFVSVGYEYESCPDLILWE KRTTVLQGYEIDASKLGGWSLDKHHALNIQSGILHKGNGENQFVSQQPPVIGSIMGNGRRRSIS CPSCNGLADGNKLLAPVALTCGSDGSLYVGDFNYIRRIFPSGNVTNILELRNKDFRHSHSPAHK YYLATDPMSGAVFLSDSNSRRVFKIKSTVVVKDLVKNSEVVAGTGDQCLPFDDTRCGDGGKATE ATLTNPRGITVDKFGLIYFVDGTMIRRIDQNGIISTLLGSNDLTSARPLSCDSVMDISQVRLEW PTDLAINPMDNSLYVLDNNVVLQISENHQVRIVAGRPMHCQVPGIDHFLLSKVAIHATLESATA LAVSHNGVLYIAETDEKKINRIRQVTTSGEISLVAGAPSGCDCKNDANCDCFSGDDGYAKDAKL NTPSSLAVCADGELYVADLGNIRIRFIRKNKPFLNTQNMYELSSPIDQELYLFDTTGKHLYTQS LPTGDYLYNFTYTGDGDITLITDNNGNMVNVRRDSTGMPLWLVVPDGQVYWVTMGTNSALKSVT TQGHELAMMTYHGNSGLLATKSNENGWTTFYEYDSFGRLTNVTFPTGQVSSFRSDTDSSVHVQV ETSSKDDVTITTNLSASGAFYTLLQDQVRNSYYIGADGSLRLLLANGMEVALQTEPHLLAGTVN PTVGKRNVTLPIDNGLNLVEWRQRKEQARGQVTVFGRRLRVHNRNLLSLDFDRVTRTEKIYDDH RKFTLRILYDQAGRPSLWSPSSRLNGVNVTYSPGGYIAGIQRGIMSERMEYDQAGRITSRIFAD GKTWSYTYLEKSMVLLLHSQRQYIFEFDKNDRLSSVTMPNVARQTLETIRSVGYYRNIYQPPEG NASVIQDFTEDGHLLHTFYLGTGRRVIYKYGKLSKLAETLYDTTKVSFTYDETAGMLKTINLQN EGFTCTIRYRQIGPLIDRQIFRFTEEGMVNARFDYNYDNSFRVTSMQAVINETPLPIDLYRYDD VSGKTEQFGKFGVIYYDINQIITTAVMTHTKHFDAYGRMKEVQYEIFRSLMYWMTVQYDNMGRV VKKELKVGPYANTTRYSYEYDADGQLQTVSINDKPLWRYSYDLNGNLHLLSPGNSARLTPLRYD IRDRITRLGDVQYKMDEDGFLRQRGGDIFEYNSAGLLIKAYNRAGSWSVRYRYDGLGRRVSSKS SHSHHLQFFYADLTNPTKVTHLYNHSSSEITSLYYDLQGHLFAMELSSGDEFYIACDNIGTPLA VFSGTGLMIKQILYTAYGEIYMDTNPNFQIIIGYHGGLYDPLTKLVHMGRRDYDVLAGRWTSPD HELWKHLSSSNVMPFNLYMFKNNNPISNSQDIKCFMTDVNSWLLTFGFQLHNVIPGYPKPDMDA MEPSYELIHTQMKTQEWDNSKSILGVQCEVQKQLKAFVTLERFDQLYGSTITSCQQAPKTKKFA SSGSVFGKGVKFALKDGRVTTDIISVANEDGRRVAAILNHAHYLENLHFTIDGVDTHYFVKPGP SEGDLAILGLSGGRRTLENGVNVTVSQINTVLNGRTRRYTDIQLQYGALCLNTRYGTTLDEEKA RVLELARQRAVRQAWAREQQRLREGEEGLRAWTEGEKQQVLSTGRVQGYDGFFVISVEQYPELS DSANNIHFMRQSEMGRR
	SEQ ID NO: 45 | 1392 bp
NOV9c, 172885447 DNA Sequence	**GGA**TCCCACCTGCAGCCGATGGAGGGGCAGATGTATGAGATCACGGAGGACACAGCCAGCAGTT GGCCTGTGCCAACCGACGTCTCCCTATACCCCTCAGGGGGCACTGGCTTAGAGACCCCTGACAG GAAAGGCAAAGGAACCACAGAAGGAAAGCCCAGTAGTTTCTTTCCAGAGGACAGTTTCATAGAT TCTGGAGAAATTGATGTGGGAAGGCGAGCTTCCCAGAAGATTCCTCCTGGCACTTTCTGGAGAT CTCAAGTGTTCATAGACCATCCTGTGCATCTGAAATTCAATGTGTCTCTGGGAAAGGCAGCCCT GGTTGGCATTTATGGCAGAAAAGGCCTCCCTCCTTCACATACACAGTTTGACTTTGTGGAGCTG CTGGATGGCAGGAGGCTCCTAACCCAGGAGGCGCGGAGCCTAGAGGGGACCCCGCGCCAGTCTC GGGGAACTGTGCCCCCCTCCAGCCATGAGACAGGCTTCATCCAGTATTTGGATTCAGGAATCTG GCACTTGGCTTTTTACAATGACGGAAAGGAGTCAGAAGTGGTTTCCTTTCTCACCACTGCCATT GAGTCGGTGGATAACTGCCCCAGCAACTGCTATGGCAATGGTGACTGCATCTCTGGGACCTGCC ACTGCTTCCTGGGTTTCCTGGGCCCCGACTGTGGCAGAGCCTCCTGCCCCGTGCTCTGTAGCGG AAATGGCCAATACATGAAAGGCAGATGCTTGTGCCACAGTGGCTGGAAAGGCGCTGAGTGCGAT GTGCCCACCAACCAGTGTATCGATGTGGCCTGCAGCAACCATGGCACCTGCATCACGGGCACCT GCATCTGCAACCCTGGCTACAAGGGCGAGAGCTGTGAGGAAGTGGACTGCATGGACCCCACATG TTCAGGCCGGGGTGTCTGCGTGAGAGGCGAATGCCACTGCTCTGTGGGATGGGGAGGCACCAAC TGCGAGACCCCCAGGGCCACATGCTTAGACCAGTGTTCAGGCCACGGAACCTTCCTCCCGGACA CCGGGCTTTGCAGCTGTGACCCAAGCTGGACTGGACACGACTGTTCTATCGAGATCTGTGCTGC CGACTGTGGTGGCCATGGCGTGTGCGTAGGGGGCACCTGCCGCTGCGAGGATGGCTGGATGGGG GCAGCCTGCGACCAGCGGGCCTGCCACCCGCGCTGTGCCGAGCATGGGACCTGCCGCGACGGCA AGTGCGAGTGCAGCCCTGGCTGGAATGGCGAACACTGCACCATCGCTCACTATCTGGATAGGGT AGTTAAAGAGGGTTGCCCTGGGTTGTGCAATGGCAACGGCAGATGTACCTTAGACCTGAATGGT

	TGGCACTGCGTCTGCCAGCTGGGCTGGAGAGGAGCTGGCTGTCTCGAG		
	ORF Start: at 1		ORF Stop: end of sequence
	SEQ ID NO: 46	464 aa	MW at 49573.0kD
NOV9c, 172885447 Protein Sequence	GSHLQPMEGQMYEITEDTASSWPVPTDVSLYPSGGTGLETPDRKGKGTTEGKPSSFFPEDSFID SGEIDVGRRASQKIPPGTFWRSQVFIDHPVHLKFNVSLGKAALVGIYGRKGLPPSHTQFDFVEL LDGRRLLTQEARSLEGTPRQSRGTVPPSSHETGFIQYLDSGIWHLAFYNDGKESEVVSFLTTAI ESVDNCPSNCYGNGDCISGTCHCFLGFLGPDCGRASCPVLCSGNGQYMKGRCLCHSGWKGAECD VPTNQCIDVACSNHGTCITGTCICNPGYKGESCEEVDCMDPTCSGRGVCVRGECHCSVGWGGTN CETPRATCLDQCSGHGTFLPDTGLCSCDPSWTGHDCSIEICAADCGGHGVCVGGTCRCEDGWMG AACDQRACHPRCAEHGTCRDGKCECSPGWNGEHCTIAHYLDRVVKEGCPGLCNGNGRCTLDLNG WHCVCQLGWRGAGCLE		
	SEQ ID NO: 47	829 bp	
NOV9d, 312712785 DNA Sequence	CACCGGATCCGAGTCGGTGGATAACTGCCCCAGCAACTGCTATGGCAATGGTGACTGCATCTCTG GGACCTGCCACTGCTTCCTGGGTTTCCTGGGCCCCGACTGTGGCAGAGCCTCCTGCCCCGTGCTC TGTAGCGGAAATGGCCAATACATGAAAGGCAGATGCTTGTGCCACAGTGGCTGGAAAGGCGCTGA GTGCGATGTGCCCACCAACCAGTGTATCGATGTGGCCTGCAGCAACCATGGCACCTGCATCACGG GCACCTGCATCTGCAACCCTGGCTACAAGGGCGAGAGCTGTGAGGAAGTGGACTGCATGGACCCC ACATGTTCAGGCCGGGGTGTCTGCGTGAGAGGCGAATGCCACTGCTCTGTGGGATGGGGAGGCAC CAACTGCGAGACCCCCAGGGCCACATGCTTAGACCAGTGTTCAGGCCACGGAACCTTCCTCCCGG ACACCGGGCTTTGCAGCTGTGACCCAAGCTGGACTGGACACGACTGTTCTATCGAGATCTGTGCT GCCGACTGTGGTGGCCATGGCGTGTGCGTAGGGGGCACCTGCCGCTGCGAGGATGGCTGGATGGG GGCAGCCTGCGACCAGCGGGCCTGCCACCCGCGCTGTGCCGAGCATGGGACCTGCCGCGACGGCA AGTGCGAGTGCAGCCCTGGCTGGAATGGCGAACACTGCACCATCGCTCACTATCTGGATAGGGTA GTTAAAGAGGGTTGCCCTGGGTTGTGCAATGGCAACGGCAGATGTACCTTAGACCTGAATGGTTG GCACTGCGTCTGCCAGCTGGGCTGGAGAGGAGCTGGCTGTCTCGAGGGC		
	ORF Start: at 2		ORF Stop: end of sequence
	SEQ ID NO: 48	276 aa	MW at 28794.9kD
NOV9d, 312712785 Protein Sequence	TGSESVDNCPSNCYGNGDCISGTCHCFLGFLGPDCGRASCPVLCSGNGQYMKGRCLCHSGWKGA ECDVPTNQCIDVACSNHGTCITGTCICNPGYKGESCEEVDCMDPTCSGRGVCVRGECHCSVGWG GTNCETPRATCLDQCSGHGTFLPDTGLCSCDPSWTGHDCSIEICAADCGGHGVCVGGTCRCEDG WMGAACDQRACHPRCAEHGTCRDGKCECSPGWNGEHCTIAHYLDRVVKEGCPGLCNGNGRCTLD LNGWHCVCQLGWRGAGCLEG		
	SEQ ID NO: 49	1399 bp	
NOV9e, 311748770 DNA Sequence	CACCGGATCCGAGTCGGTGGATAACTGCCCCAGCAACTGCTATGGCAATGGTGACTGCATCTCT GGGACCTGCCACTGCTTCCTGGGTTTCCTGGGCCCCGACTGTGGCAGAGCCTCCTGCCCCGTGC TCTGTAGCGGAAATGGCCAATACATGAAAGGCAGATGCTTGTGCCACAGTGGCTGGAAAGGCGC TGAGTGCGATGTGCCCACCAACCAGTGTATCGATGTGGCCTGCAGCAACCATGGCACCTGCATC ACGGGCACCTGCATCTGCAACCCTGGCTACAAGGGCGAGAGCTGTGAGGAAGTGGACTGCATGG ACCCCACATGTTCAGGCCGGGGTGTCTGCGTGAGAGGCGAATGCCACTGCTCTGTGGGATGGGG AGGCACCAACTGCGAGACCCCCAGGGCCACATGCTTAGACCAGTGTTCAGGCCACGGAACCTTC CTCCCGGACACCGGGCTTTGCAGCTGTGACCCAAGCTGGACTGGACACGACTGTTCTATCGAGA TCTGTGCTGCCGACTGTGGTGGCCATGGCGTGTGCGTAGGGGGCACCTGCCGCTGCGAGGATGG CTGGATGGGGGCAGCCTGCGACCAGCGGGCCTGCCACCCGCGCTGTGCCGAGCATGGGACCTGC CGCGACGGCAAGTGCGAGTGCAGCCCTGGCTGGAATGGCGAACACTGCACCATCGCTCACTATC TGGATAGGGTAGTTAAAGAGGGTTGCCCTGGGTTGTGCAATGGCAACGGCAGATGTACCTTAGA CCTGAATGGTTGGCACTGCGTCTGCCAGCTGGGCTGGAGAGGAGCTGGCTGTGACACTTCCATG GAGACTGCCTGCGGTGACAGCAAAGACAATGATGGAGATGGCCTGGTGGACTGCATGGACCCTG ACTGCTGCCTCCAGCCCCTGTGCCATATCAACCCGCTGTGCCTTGGCTCCCCTAACCCTCTGGA CATCATCCAGGAGACACAGGTCCCTGTGTCACAGCAGAACCTACACTCCTTCTATGACCGCATC AAGTTCCTCGTGGGCAGGGACAGCACGCACATAATCCCCGGGGAGAACCCCTTTGATGGAGGGC ATGCTTGTGTTATTCGTGGCCAAGTGATGACATCAGATGGAACCCCCCTGGTTGGTGTGAACAT CAGTTTTGTCAATAACCCTCTCTTTGGATATACAATCAGCAGGCAAGATGGCAGCTTTGACTTG GTGACAAATGGCGGCATCTCCATCATCCTGCGGTTCGAGCGGGCACCTTTCATCACACAGGAGC ACACCCTGTGGCTGCCATGGGATCGCTTCTTTGTCATGGAAACCATCATCATGAGACATGAGGA GAATGAGATTCCCAGCTGTGACCTGAGCAATTTTGCCCGCCCCAACCTCGAGGGC		
	ORF Start: at 2		ORF Stop: end of sequence
	SEQ ID NO: 50	466 aa	MW at 49933.5kD
NOV9e, 311748770 Protein Sequence	TGSESVDNCPSNCYGNGDCISGTCHCFLGFLGPDCGRASCPVLCSGNGQYMKGRCLCHSGWKGA ECDVPTNQCIDVACSNHGTCITGTCICNPGYKGESCEEVDCMDPTCSGRGVCVRGECHCSVGWG GTNCETPRATCLDQCSGHGTFLPDTGLCSCDPSWTGHDCSIEICAADCGGHGVCVGGTCRCEDG WMGAACDQRACHPRCAEHGTCRDGKCECSPGWNGEHCTIAHYLDRVVKEGCPGLCNGNGRCTLD		

	LNGWHCVCQLGWRGAGCDTSMETACGDSKDNDGDGLVDCMDPDCCLQPLCHINPLCLGSPNPLD IIQETQVPVSQQNLHSFYDRIKFLVGRDSTHIIPGENPFDGGHACVIRGQVMTSDGTPLVGVNI SFVNNPLFGYTISRQDGSFDLVTNGGISIILRFERAPFITQEHTLWLPWDRFFVMETIIMRHEE NEIPSCDLSNFARPNLEG		
	SEQ ID NO: 51	1371 bp	
NOV9f, 311748779 DNA Sequence	CACCGGATCCGAGTCGGTGGATAACTGCCCCAGCAACTGCTATGGCAATGGTGACTGCATCTCT GGGACCTGCCACTGCTTCCTGGGTTTCCTGGGCCCCGACTGTGGCAGAGCCTCCTGCCCCGTGC TCTGTAGCGGAAATGGCCAATACATGAAAGGCAGATGCTTGTGCCACAGTGGCTGGAAAGGCGC TGAGTGCGATGTGCCCACCAACCAGTGTATCGATGTGGCCTGCAGCAACCATGGCACCTGCATC ACGGGCACCTGCATCTGCAACCCTGGCTACAAGGGCGAGAGCTGTGAGGAAGTGGACTGCATGG ACCCCACATGTTCAGGCCGGGGTGTCTGCGTGAGAGGCGAATGCCACTGCTCTGTGGGATGGGG AGGCACCAACTGCGAGACCCCCAGGGCCACATGCTTAGACCAGTGTTCAGGCCACGGAACCTTC CTCCCGGACACCGGGCTTTGCAGCTGTGACCCAAGCTGGACTGGACACGACTGTTCTATCGAGA TCTGTGCTGCCGACTGTGGTGGCCATGGCGTGTGCGTAGGGGGCACCTGCCGCTGCGAGGATGG CTGGATGGGGGCAGCCTGCGACCAGCGGGCCTGCCACCCGCGCTGTGCCGAGCATGGGACCTGC CGCGACGGCAAGTGCGAGTGCAGCCCTGGCTGGAATGGCGAACACTGCACCATCGAGGGTTGCC CTGGGTTGTGCAATGGCAATGGCAGATGTACCTTAGACCTGAATGGTTGGCACTGCGTCTGCCA GCTGGGCTGGAGAGGAGCTGGCTGTGACACTTCCATGGAGACTGCCTGCGGTGACAGCAAAGAC AATGATGGAGATGGCCTGGTGGACTGCATGGACCCTGACTGCTGCCTCCAGCCCCTGTGCCATA TCAACCCGCTGTGCCTTGGCTCCCCTAACCCTCTGGACATCATCCAGGAGACACAGGTCCCTGT GTCACAGCAGAACCTACACTCCTTCTATGACCGCATCAAGTTCCTCGTGGGCAGGGACAGCACG CACATAATCCCCGGGGAGAACCCCTTTGATGGAGGGCATGCTTGTGTTATTCGTGGCCAAGTGA TGACATCAGATGGAACCCCCCTGGTTGGTGTGAACATCAGTTTTGTCAATAACCCTCTCTTTGG ATATACAATCAGCAGGCAAGATGGCAGCTTTGACTTGGTGACAAATGGCGGCATCTCCATCATT CTGTGGTTCGAGCGGGCACCTTTCATCACACAGGAGCACACCCTGTGGCTGCCATGGGATCGCT TCTTTGTCATGGAAACCATCATCATGAGACATGAGGAGAATGAGATTCCCAGCTGTGACCTGAG CAATTTTGCCCGCCCCAACCTCGAGGG		
	ORF Start: at 2		ORF Stop: at 1370
	SEQ ID NO: 52	456 aa	MW at 48824.2kD
NOV9f, 311748779 Protein Sequence	TGSESVDNCPSNCYGNGDCISGTCHCFLGFLGPDCGRASCPVLCSGNGQYMKGRCLCHSGWKGA ECDVPTNQCIDVACSNHGTCITGTCICNPGYKGESCEEVDCMDPTCSGRGVCVRGECHCSVGWG GTNCETPRATCLDQCSGHGTFLPDTGLCSCDPSWTGHDCSIEICAADCGGHGVCVGGTCRCEDG WMGAACDQRACHPRCAEHGTCRDGKCECSPGWNGEHCTIEGCPGLCNGNGRCTLDLNGWHCVCQ LGWRGAGCDTSMETACGDSKDNDGDGLVDCMDPDCCLQPLCHINPLCLGSPNPLDIIQETQVPV SQQNLHSFYDRIKFLVGRDSTHIIPGENPFDGGHACVIRGQVMTSDGTPLVGVNISFVNNPLFG YTISRQDGSFDLVTNGGISIILWFERAPFITQEHTLWLPWDRFFVMETIIMRHEENEIPSCDLS NFARPNLE		
	SEQ ID NO: 53	1392 bp	
NOV9g, CG50301-02 DNA Sequence	GGATCCCACCTGCAGCCGATGGAGGGGCAGATGTATGAGATCACGGAGGACACAGCCAGCAGTT GGCCTGTGCCAACCGACGTCTCCCTATACCCCTCAGGGGGCACTGGCTTAGAGACCCCTGACAG GAAAGGCAAAGGAACCACAGAAGGAAAGCCCAGTAGTTTCTTTCCAGAGGACAGTTTCATAGAT TCTGGAGAAATTGATGTGGGAAGGCGAGCTTCCCAGAAGATTCCTCCTGGCACTTTCTGGAGAT CTCAAGTGTTCATAGACCATCCTGTGCATCTGAAATTCAATGTGTCTCTGGGAAAGGCAGCCCT GGTTGGCATTTATGGCAGAAAAGGCCTCCCTCCTTCACATACACAGTTTGACTTTGTGGAGCTG CTGGATGGCAGGAGGCTCCTAACCCAGGAGGCGCGGAGCCTAGAGGGGACCCCGCGCCAGTCTC GGGGAACTGTGCCCCCCTCCAGCCATGAGACAGGCTTCATCCAGTATTTGGATTCAGGAATCTG GCACTTGGCTTTTTACAATGACGGAAAGGAGTCAGAAGTGGTTTCCTTTCTCACCACTGCCATT GAGTCGGTGGATAACTGCCCCAGCAACTGCTATGGCAATGGTGACTGCATCTCTGGGACCTGCC ACTGCTTCCTGGGTTTCCTGGGCCCCGACTGTGGCAGAGCCTCCTGCCCCGTGCTCTGTAGCGG AAATGGCCAATACATGAAAGGCAGATGCTTGTGCCACAGTGGCTGGAAAGGCGCTGAGTGCGAT GTGCCCACCAACCAGTGTATCGATGTGGCCTGCAGCAACCATGGCACCTGCATCACGGGCACCT GCATCTGCAACCCTGGCTACAAGGGCGAGAGCTGTGAGGAAGTGGACTGCATGGACCCCACATG TTCAGGCCGGGGTGTCTGCGTGAGAGGCGAATGCCACTGCTCTGTGGGATGGGGAGGCACCAAC TGCGAGACCCCCAGGGCCACATGCTTAGACCAGTGTTCAGGCCACGGAACCTTCCTCCCGGACA CCGGGCTTTGCAGCTGTGACCCAAGCTGGACTGGACACGACTGTTCTATCGAGATCTGTGCTGC CGACTGTGGTGGCCATGGCGTGTGCGTAGGGGGCACCTGCCGCTGCGAGGATGGCTGGATGGGG GCAGCCTGCGACCAGCGGGCCTGCCACCCGCGCTGTGCCGAGCATGGGACCTGCCGCGACGGCA AGTGCGAGTGCAGCCCTGGCTGGAATGGCGAACACTGCACCATCGCTCACTATCTGGATAGGGT AGTTAAAGAGGGTTGCCCTGGGTTGTGCAATGGCAACGGCAGATGTACCTTAGACCTGAATGGT TGGCACTGCGTCTGCCAGCTGGGCTGGAGAGGAGCTGGCTGTCTCGAG		
	ORF Start: at 7		ORF Stop: at 1387
	SEQ ID NO: 54	460 aa	MW at 49186.6kD

NOV9g, CG50301-02 Protein Sequence	HLQPMEGQMYEITEDTASSWPVPTDVSLYPSGGTGLETPDRKGKGTTEGKPSSFFPEDSFIDSG EIDVGRRASQKIPPGTFWRSQVFIDHPVHLKFNVSLGKAALVGIYGRKGLPPSHTQFDFVELLD GRRLLTQEARSLEGTPRQSRGTVPPSSHETGFIQYLDSGIWHLAFYNDGKESEVVSFLTTAIES VDNCPSNCYGNGDCISGTCHCFLGFLGPDCGRASCPVLCSGNGQYMKGRCLCHSGWKGAECDVP TNQCIDVACSNHGTCITGTCICNPGYKGESCEEVDCMDPTCSGRGVCVRGECHCSVGWGGTNCE TPRATCLDQCSGHGTFLPDTGLCSCDPSWTGHDCSIEICAADCGGHGVCVGGTCRCEDGWMGAA CDQRACHPRCAEHGTCRDGKCECSPGWNGEHCTIAHYLDRVVKEGCPGLCNGNGRCTLDLNGWH CVCQLGWRGAGC		
	SEQ ID NO: 55	8355 bp	
NOV9h, CG50301-03 DNA Sequence	**ATG**GACGTGAAGGAGAGGAAGCCTTACCGCTCGCTGACCCGGCGCCGCGACGCCGAGCGCCGCT ACACCAGCTCGTCCGCGGACAGCGAGGAGGGCAAAGCCCCGCAGAAATCGTACAGCTCCAGCGA GACCCTGAAGGCCTACGACCAGGACGCCCGCCTAGCCTATGGCAGCCGCGTCAAGGACATTGTG CCGCAGGAGGCCGAGGAATTCTGCCGCACAGGTGCCAACTTCACCCTGCGGGAGCTGGGGCTGG AAGAAGTAACGCCCCCTCACGGGACCCTGTACCGGACAGACATTGGCCTCCCCCACTGCGGCTA CTCCATGGGGGCTGGCTCTGATGCCGACATGGAGGCTGACACGGTGCTGTCCCCTGAGCACCCC GTGCGTCTGTGGGGCCGGAGCACACGGTCAGGGCGCAGCTCCTGCCTGTCCAGCCGGGCCAATT CCAATCTCACACTCACCGACACCGAGCATGAAAACACTGAGACTGATCATCCGGGCGGCCTGCA GAACCACGCGCGGCTCCGGACGCCGCCGCCGCCGCTCTCGCACGCCCACACCCCCAACCAGCAC CACGCGGCCTCCATTAACTCCCTGAACCGGGGCAACTTCACGCCGAGGAGCAACCCCAGCCCGG CCCCCACGGACCACTCGCTCTCCGGAGAGCCCCCTGCCGGCGGCGCCCAGGAGCCTGCCCACGC CCAGGAGAACTGGCTGCTCAACAGCAACATCCCCCTGGAGACCAGAAACCTAGGCAAGCAGCCA TTCCTAGGGACATTGCAGGACAACCTCATTGAGATGGACATTCTCGGCGCCTCCCGCCATGATG GGGCTTACAGTGACGGGCACTTCCTCTTCAAGCCTGGAGGCACCTCCCCGCTCTTCTGCACCAC ATCACCAGGGTACCCACTGACGTCCAGCACAGTGTACTCTCCTCCGCCCCGACCCCTGCCCCGC AGCACCTTCGCCTGGCCGGCCTTTAACCTCAAGAAGCCCTCCAAGTACTGTAACTGGAAGTGCG CAGCCCTGAGCGCCATCGTCATCTCAGCCACTCTGGTCATCCTGCTGGCATACTTTGTGGCCAT GCACCTGTTTGGCCTAAACTGGCACCTGCAGCCGATGGAGGGGCAGATGTATGAGATCACGGAG GACACAGCCAGCAGTTGGCCTGTGCCAACCGACGTCTCCCTATACCCCTCAGGGGGCACTGGCT TAGAGACCCCTGACAGGAAAGGCAAAGGAACCACAGAAGGAAAGCCCAGTAGTTTCTTTCCAGA GGCCAGTTTCATAGATTCTGGAGAAATTGATGTGGGAAGGCGAGCTTCCCAGAAGATTCCTCCT GGCACTTTCTGGAGATCTCAAGTGTTCATAGACCATCCTGTGCATCTGAAATTCAATGTGTCTC TGGGAAAGGCAGCCCTGGTTGGCATTTATGGCAGAAAAGGCCTCCCTCCTTCACATACACAGTT TGACTTTGTGGAGCTGCTGGATGGCAGGAGGCTCCTAACCCAGGAGGCGCGGAGCCTAGAGGGG ACCCCGCGCCAGTCTCGGGGAACTGTGCCCCCCTCCAGCCATGAGACAGGCTTCATCCAGTATT TGGATTCAGGAATCTGGCACTTGGCTTTTTACAATGACGGAAAGGAGTCAGAAGTGGTTTCCTT TCTCTCCCCAGAGTCGGTGGATAACTGCCCCAGCAACTGCTATGGCAATGGTGACTGCATCTCT GGGACCTGCCACTGCTTCCTGGGTTTCCTGGGCCCCGACTGTGGCAGAGCCTCCTGCCCCGTGC TCTGTAGCGGAAATGGCCAATACATGAAAGGCAGATGCTTGTGCCACAGTGGCTGGAAAGGCGC TGAGTGCGATGTGCCCACCAACCAGTGTATCGATGTGGCCTGCAGCAACCATGGCACCTGCATC ACGGGCACCTGCATCTGCAACCCTGGCTACAAGGGCGAGAGCTGTGAGGAAGACTGCATGGACC CCACATGTTCAGGCCGGGGTGTCTGCGTGAGAGGCGAATGCCACTGCTCTGTGGGATGGGGAGG CACCAACTGCGAGACCCCCAGGGCCACATGCTTAGACCAGTGTTCAGGCCACGGAACCTTCCTC CCGGACACCGGGCTTTGCAGCTGTGACCCAAGCTGGACTGGACACGACTGTTCTATCGAGATCT GTGCTGCCGACTGTGGTGGCCATGGCGTGTGCGTAGGGGGCACCTGCCGCTGCGAGGATGGCTG GATGGGGGCAGCCTGCGACCAGCGGGCCTGCCACCCGCGCTGTGCCGAGCATGGGACCTGCCGC GACGGCAAGTGCGAGTGCAGCCCTGGCTGGAATGGCGAACACTGCACCATCTCCCTAGCTCACT ATCTGGATAGGGTAGTTAAACTTTCAGAGGGTTGCCCTGGGTTGTGCAATGGCAACGGCAGATG TACCTTAGACCTGAATGGTTGGCACTGCGTCTGCCAGCTGGGCTGGAGAGGAGCTGGCTGTGAC ACTTCCATGGAGACTGCCTGCGGTGACAGCAAAGACAATGATGGAGATGGCCTGGTGGACTGCA TGGACCCTGACTGCTGCCTCCAGCCCCTGTGCCATATCAACCCGCTGTGCCTTGGCTCCCCTAA CCCTCTGGACATCATCCAGGAGACACAGGTCCCTGTGTCACAGCAGAACCTACACTCCTTCTAT GACCGCATCAAGTTCCTCGTGGGCAGGGACAGCACGCACATAATCCCCGGGGAGAACCCCTTTG ATGGAGGGCATGCTTGTGTTATTCGTGGCCAAGTGATGACATCAGATGGAACCCCCCTGGTTGG TGTGAACATCAGTTTTGTCAATAACCCTCTCTTTGGATATACAATCAGCAGGCAAGATGGCAGC TTTGACTTGGTGACAAATGGCGGCATCTCCATCATCCTGCGGTTCGAGCGGGCACCTTTCATCA CACAGGAGCACACCCTGTGGCTGCCATGGGATCGCTTCTTTGTCATGGAAACCATCATCATGAG ACATGAGGAGAATGAGATTCCCAGCTGTGACCTGAGCAATTTTGCCCGCCCCAACCCAGTCGTC TCTCCATCCCCACTGACGTCCTTCGCCAGCTCCTGTGCAGAGAAAGGCCCCATTGTGCCGGAAA TTCAGGCTTTGCAGGAGGAAATCTCTATCTCTGGCTGCAAGATGAGGCTGAGCTACCTGAGCAG CCGGACCCCTGGCTACAAATCTGTCCTGAGGATCAGCCTCACCCACCCGACCATCCCCTTCAAC CTCATGAAGGTGCACCTCATGGTAGCGGTGGAGGGCCGCCTCTTCAGGAAGTGGTTCGCTGCAG CCCCAGACCTGTCCTATTATTTCATTTGGGACAAGACAGACGTCTACAACCAGAAGGTGTTTGG GCTTTCAGAAGCCTTTGTTTCCGTGGGTTATGAATATGAATCCTGCCCAGATCTAATCCTGTGG GAAAAAAGAACAACAGTGCTGCAGGGCTATGAAATTGATGCGTCCAAGCTTGGAGGATGGAGCC TAGACAAACATCATGCCCTCAACATTCAAAGTGGTGGCATCCTGCACAAAGGGAATGGGGAGAA CCAGTTTGTGTCTCAGCAGCCTCCTGTCATTGGGAGCATCATGGGCAATGGGCGCCGGAGAAGC		

ATCTCCTGCCCCAGCTGCAACGGCCTTGCTGACGGCAACAAGCTCCTGGCCCCAGTGGCCCTCA
CCTGTGGCTCTGACGGGAGCCTCTATGTGGGTGATTTCAACTACATTAGAAGGATCTTCCCCTC
TGGAAATGTCACCAACATCCTAGAGCTGAGGGTCAGAAATAAAGATTTCAGACATAGTCACAGT
CCAGCACACAAATACTACCTGGCCACAGACCCCATGAGTGGGGCCGTCTTCCTTTCTGACAGCA
ACAGCCGGCGGGTCTTTAAAATCAAGTCCACTGTGGTGGTGAAGGACCTTGTCAAGAACTCTGA
GGTGGTTGCGGGGACAGGTGACCAGTGCCTCCCCTTTGATGACACTCGCTGCGGGGATGGTGGG
AAGGCCACAGAAGCCACACTCACCAATCCCAGGGGTCCCCCAGGCATTACAGTGGACAAGTTTG
GGCTGATCTACTTCGTGGATGGCACCATGATCAGACGCATCGATCAGAATGGGATCATCTCCAC
CCTGCTCGGCTCTAATGATCTCACATCAGCCCGGCCACTCAGCTGTGATTCTGTCATGGATATT
TCCCAGGTAAGACAGGTTCACCTGGAGTGGCCCACAGACTTAGCCATCAACCCAATGGACAACT
CACTTTATGTCCTCGACAACAATGTGGTCCTGCAAATCTCTGAAAACCACCAGGTGCGCATTGT
CGCCGGGAGGCCCATGCACTGCCAGGTCCCTGGCATTGACCACTTCCTGCTAAGCAAGGTGGCC
ATCCACGCAACCCTGGAGTCAGCCACCGCTTTGGCTGTTTCACACAATGGGGTCCTGTATATTG
CTGAGACTGATGAGAAAAAGATCAACCGCATCAGGCAGGTCACCACTAGTGGAGAGATCTCACT
CGTTGCTGGGGCCCCCAGTGGCTGTGACTGTAAAAATGATGCCAACTGTGATTGTTTTTCTGGA
GACGATGGTTATGCCAAGGATGCAAAGTTAAATACCCCATCTTCCTTGGCTGTGTGTGCTGATG
GGGAGCTCTACGTGGCCGACCTTGGGAACATCCGAATTCGGTTTATCCGGAAGAACAAGCCTTT
CCTCAACACCCAGAACATGTATGAGCTGTCTTCACCAATTGACCAGGAGCTCTATCTGTTTGAT
ACCACCGGCAAGCACCTGTACACCCAAAGCCTGCCCACAGGAGACTACCTGTACAACTTCACCT
ACACTGGGGACGGCGACATCACACTCATCACAGACAACAATGGCAACATGGTAAATGTCCGCCG
AGACTCTACTGGGATGCCCCTCTGGCTGGTGGTCCCAGATGGCCAGGTGTACTGGGTGACCATG
GGCACCAACAGTGCACTCAAGAGTGTGACCACACAAGGACACGAGTTGGCCATGATGACATACC
ATGGCAATTCCGGCCTTCTGGCAACCAAAAGCAATGAAAACGGATGGACAACATTTTATGAGTA
CGACAGCTTTGGCCGCCTGACAAATGTGACCTTCCCTACTGGCCAGGTGAGCAGTTTCCGAAGT
GATACAGACAGTTCAGTGCATGTCCAGGTAGAGACCTCCAGCAAGGATGATGTCACCATAACCA
CCAACCTGTCTGCCTCAGGCGCCTTCTACGACCAAGTCCGGAACAGCTACTACATCGGGGCCGA
TGGCTCCTTGCGGCTGCTGCTGGCCAACGGCATGGAGGTGGCGCTGCAGACTGAGCCCCACTTG
CTGGCTGGCACCGTCAACCCCACCGTGGGCAAGAGGAATGTCACGCTGCCCATCGACAACGGCC
TCAACCTGGTGGAGTGGCGCCAGCGCAAAGAGCAGGCTCGGGGCCAGGTCACTGTCTTTGGGCG
CCGGCTGCGGGTGCTCCAGGTTCACAACCGAAATCTCCTATCTCTGGACTTTGATCGCGTAACA
CGCACAGAGAAGATCTATGATGACCACCGCAAGTTCACCCTTCGGATTCTGTACGACCAGGCGG
GGCGGCCCAGCCTCTGGTCACCCAGCAGCAGGCTGAATGGTGTCAACGTGACATACTCCCCTGG
GGGTTACATTGCTGGCATCCAGAGGGGCATCATGTCTGAAAGAATGGAATACGACCAGGCGGGC
CGCATCACATCCAGGATCTTCGCTGATGGGAAGACATGGAGCTACACATACTTAGAGAAGGCAG
GTGTCGAGTTCGACAAGAATGACCGCCTCTCTTCTGTGACGATGCCCAACGTGGCGCGGCAGAC
ACTAGAGACCATCCGCTCAGTGGGCTACTACAGAAACATCTATCAGCCCCCTGAGGGCAATGCC
TCAGTCATACAGGACTTCACTGAGGATGGGCACCTCCTTCACACCTTCTACCTGGGCACTGGCC
GCAGGGTGATATACAAGTATGGCAAACTGTCAAAGCTGGCAGAGACGCTCTATGACACCACCAA
GGTCAGTTTCACCTATGACGAGACGGCAGGCATGCTGAAGACCATCAACCTACAGAATGAGGGC
TTCACCTGCACCATCCGCTACCGTCAGATTGGGCCCCTGATTGACCGACAGATCTTCCGCTTCA
CTGAGGAAGGCATGGTCAACGCCCGTTTTGACTACAACTATGACAACAGCTTCCGGGTGACCAG
CATGCAGGCTGTGATCAACGAGACCCCACTGCCCATTGATCTCTATCGCTATGATGATGTGTCA
GGCAAGACAGAGCAGTTTGGGAAGTTTGGTGTCATTTACTATGACATTAACCAGATCATCACCA
CAGCTGTCATGACCCACACCAAGCATTTTGATGCATATGGCAGGATGAAGGAAGTGCAGTATGA
GATCTTCCGCTCGCTCATGTACTGGATGACCGTCCAGTATGATAACATGGGGCGAGTAGTGAAG
AAGGAGCTGAAGGTAGGACCCTACGCCAATACCACTCGCTACTCCTATGAGTATGATGCTGACG
GCCAGCTGCAGACAGTCTCCATCAATGACAAGCCACTCTGGCGCTACAGCTACGACCTCAATGG
GAACCTGCACTTACTGAGCCCTGGGAACAGTGCACGGCTCACACCACTACGGTATGACATCCGC
GACCGCATCACTCGGCTGGGTGACGTGCAATACAAGATGGATGAGGATGGCTTCCTGAGGCAGC
GGGGCGGTGATATCTTTGAGTACAACTCAGCTGGCCTGCTCATCAAGGCCTACAACCGGGCTGG
CAGCTGGAGTGTCAGGTACCGCTACGATGGCCTGGGGCGGCGCGTGTCCAGCAAGAGCAGCCAC
AGCCACCACCTGCAGTTCTTCTATGCAGACCTGACCAACCCCACCAAGGTCACCCACCTGTACA
ACCACTCCAGCTCTGAGATCACCTCCCTCTACTACGACTTGCAAGGACACCTCTTTGCCATGGA
GCTGAGCAGTGGTGATGAGTTTTACATAGCTTGTGACAACATCGGGACCCCTCTTGCTGTCTTT
AGTGGAACAGGTTTGATGATCAAGCAAATCCTGTACACAGCCTATGGGGAGATCTACATGGATA
CCAACCCCAACTTTCAGATCATCATAGGCTACCATGGTGGCCTCTATGATCCACTCACCAAGCT
TGTCCACATGGGCCGGCGAGATTATGATGTGCTGGCCGGACGCTGGACTAGCCCAGACCACGAG
CTGTGGAAGCACCTTAGTAGCAGCAACGTCATGCCTTTTAATCTCTATATGTTCAAAAACAACA
ACCCCATCAGCAACTCCCAGGACATCAAGTGCTTCATGACAGATGTTAACAGCTGGCTGCTCAC
CTTTGGATTCCAGCTACACAACGTGATCCCTGGTTATCCCAAACCAGACATGGATGCCATGGAA
CCCTCCTACGAGCTCATCCACACACAGATGAAAACGCAGGAGTGGGACAACAGCAAGGTAATTC
CTGCACAAGGCTGCCAGTCTATCCTCGGGGTACAGTGTGAAGTACAGAAGCAGCTCAAGGCCTT
TGTCACCTTAGAACGGTTTGACCAGCTCTATGGCTCCACAATCACCAGCTGCCAGCAGGCTCCA
AAGACCAAGAAGTTTGCATCCAGCGGCTCAGTCTTTGGCAAGGGGGTCAAGTTTGCCTTGAAGG
ATGGCCGAGTGACCACAGACATCATCAGTGTGGCCAATGAGGATGGGCGAAGGGTTGCTGCCAT
CTTGAACCATGCCCACTACCTAGAGAACCTGCACTTCACCATTGATGGGGTGGATACCCATTAC
TTTGTGAAACCAGGACCTTCAGAAGGTGACCTGGCCATCCTGGGCCTCAGTGGGGGGCGGCGAA

	CCCTGGAGAATGGGGTCAACGTCACTGTGTCCCAGATCAACACAGTACTTAATGGCAGGACTAG ACGCTACACAGACATCCAGCTCCAGTACGGGGCACTGTGCTTGAACACACGCTACGGGACAACG TTGGATGAGGAGAAGGCACGGGTCCTGGAGCTGGCCCGGCAGAGAGCCGTGCGCCAAGCGTGGG CCCGCGAGCAGCAGAGACTGCGGGAAGGGGAGGAAGGCCTGCGGGCCTGGACAGAGGGGGAGAA GCAGCAGGTGCTGAGCACAGGGCGGGTGCAAGGCTACGACGGCTTTTTCGTGATCTCTGTCGAG CAGTACCCAGAACTGTCAGACAGCGCCAACAACATCCACTTCATGAGACAGAGCGAGATGGGCC GGAGG**TGA**CAGAGAGGACCAAGGACTTCTTGCCAA
	ORF Start: ATG at 1 | ORF Stop: TGA at 8326
	SEQ ID NO: 56 | 2775 aa | MW at 308177.6kD
NOV9h, CG50301-03 Protein Sequence	MDVKERKPYRSLTRRRDAERRYTSSSADSEEGKAPQKSYSSSETLKAYDQDARLAYGSRVKDIV PQEAEEFCRTGANFTLRELGLEEVTPPHGTLYRTDIGLPHCGYSMGAGSDADMEADTVLSPEHP VRLWGRSTRSGRSSCLSSRANSNLTLTDTEHENTETDHPGGLQNHARLRTPPPPLSHAHTPNQH HAASINSLNRGNFTPRSNPSPAPTDHSLSGEPPAGGAQEPAHAQENWLLNSNIPLETRNLGKQP FLGTLQDNLIEMDILGASRHDGAYSDGHFLFKPGGTSPLFCTTSPGYPLTSSTVYSPPPRPLPR STFAWPAFNLKKPSKYCNWKCAALSAIVISATLVILLAYFVAMHLFGLNWHLQPMEGQMYEITE DTASSWPVPTDVSLYPSGGTGLETPDRKGKGTTEGKPSSFFPEASFIDSGEIDVGRRASQKIPP GTFWRSQVFIDHPVHLKFNVSLGKAALVGIYGRKGLPPSHTQFDFVELLDGRRLLTQEARSLEG TPRQSRGTVPPSSHETGFIQYLDSGIWHLAFYNDGKESEVVSFLSPESVDNCPSNCYGNGDCIS GTCHCFLGFLGPDCGRASCPVLCSGNGQYMKGRCLCHSGWKGAECDVPTNQCIDVACSNHGTCI TGTCICNPGYKGESCEEDCMDPTCSGRGVCVRGECHCSVGWGGTNCETPRATCLDQCSGHGTFL PDTGLCSCDPSWTGHDCSIEICAADCGGHGVCVGGTCRCEDGWMGAACDQRACHPRCAEHGTCR DGKCECSPGWNGEHCTISLAHYLDRVVKLSEGCPGLCNGNGRCTLDLNGWHCVCQLGWRGAGCD TSMETACGDSKDNDGDGLVDCMDPDCCLQPLCHINPLCLGSPNPLDIIQETQVPVSQQNLHSFY DRIKFLVGRDSTHIIPGENPFDGGHACVIRGQVMTSDGTPLVGVNISFVNNPLFGYTISRQDGS FDLVTNGGISIILRFERAPFITQEHTLWLPWDRFFVMETIIMRHEENEIPSCDLSNFARPNPVV SPSPLTSFASSCAEKGPIVPEIQALQEEISISGCKMRLSYLSSRTPGYKSVLRISLTHPTIPFN LMKVHLMVAVEGRLFRKWFAAAPDLSYYFIWDKTDVYNQKVFGLSEAFVSVGYEYESCPDLILW EKRTTVLQGYEIDASKLGGWSLDKHHALNIQSGGILHKGNGENQFVSQQPPVIGSIMGNGRRRS ISCPSCNGLADGNKLLAPVALTCGSDGSLYVGDFNYIRRIFPSGNVTNILELRVRNKDFRHSHS PAHKYYLATDPMSGAVFLSDSNSRRVFKIKSTVVVKDLVKNSEVVAGTGDQCLPFDDTRCGDGG KATEATLTNPRGPPGITVDKFGLIYFVDGTMIRRIDQNGIISTLLGSNDLTSARPLSCDSVMDI SQVRQVHLEWPTDLAINPMDNSLYVLDNNVVLQISENHQVRIVAGRPMHCQVPGIDHFLLSKVA IHATLESATALAVSHNGVLYIAETDEKKINRIRQVTTSGEISLVAGAPSGCDCKNDANCDCFSG DDGYAKDAKLNTPSSLAVCADGELYVADLGNIRIRFIRKNKPFLNTQNMYELSSPIDQELYLFD TTGKHLYTQSLPTGDYLYNFTYTGDGDITLITDNNGNMVNVRRDSTGMPLWLVVPDGQVYWVTM GTNSALKSVTTQGHELAMMTYHGNSGLLATKSNENGWTTFYEYDSFGRLTNVTFPTGQVSSFRS DTDSSVHVQVETSSKDDVTITTNLSASGAFYDQVRNSYYIGADGSLRLLLANGMEVALQTEPHL LAGTVNPTVGKRNVTLPIDNGLNLVEWRQRKEQARGQVTVFGRRLRVLQVHNRNLLSLDFDRVT RTEKIYDDHRKFTLRILYDQAGRPSLWSPSSRLNGVNVTYSPGGYIAGIQRGIMSERMEYDQAG RITSRIFADGKTWSYTYLEKAGVEFDKNDRLSSVTMPNVARQTLETIRSVGYYRNIYQPPEGNA SVIQDFTEDGHLLHTFYLGTGRRVIYKYGKLSKLAETLYDTTKVSFTYDETAGMLKTINLQNEG FTCTIRYRQIGPLIDRQIFRFTEEGMVNARFDYNYDNSFRVTSMQAVINETPLPIDLYRYDDVS GKTEQFGKFGVIYYDINQIITTAVMTHTKHFDAYGRMKEVQYEIFRSLMYWMTVQYDNMGRVVK KELKVGPYANTTRYSYEYDADGQLQTVSINDKPLWRYSYDLNGNLHLLSPGNSARLTPLRYDIR DRITRLGDVQYKMDEDGFLRQRGGDIFEYNSAGLLIKAYNRAGSWSVRYRYDGLGRRVSSKSSH SHHLQFFYADLTNPTKVTHLYNHSSSEITSLYYDLQGHLFAMELSSGDEFYIACDNIGTPLAVF SGTGLMIKQILYTAYGEIYMDTNPNFQIIIGYHGGLYDPLTKLVHMGRRDYDVLAGRWTSPDHE LWKHLSSSNVMPFNLYMFKNNNPISNSQDIKCFMTDVNSWLLTFGFQLHNVIPGYPKPDMDAME PSYELIHTQMKTQEWDNSKVIPAQGCQSILGVQCEVQKQLKAFVTLERFDQLYGSTITSCQQAP KTKKFASSGSVFGKGVKFALKDGRVTTDIISVANEDGRRVAAILNHAHYLENLHFTIDGVDTHY FVKPGPSEGDLAILGLSGGRRTLENGVNVTVSQINTVLNGRTRRYTDIQLQYGALCLNTRYGTT LDEEKARVLELARQRAVRQAWAREQQRLREGEEGLRAWTEGEKQQVLSTGRVQGYDGFFVISVE QYPELSDSANNIHFMRQSEMGRR
	SEQ ID NO: 57 | 802 bp
NOV9i, CG50301-04 DNA Sequence	CACCGGATCC**GAG**TCGGTGGATAACTGCCCCAGCAACTGCTATGGCAATGGTGACTGCATCTCT GGGACCTGCCACTGCTTCCTGGGTTTCCTGGGCCCCGACTGTGGCAGAGCCTCCTGCCCCGTGC TCTGTAGCGGAAATGGCCAATACATGAAAGGCAGATGCTTGTGCCACAGTGGCTGGAAAGGCGC TGAGTGCGATGTGCCCACCAACCAGTGTATCGATGTGGCCTGCAGCAACCATGGCACCTGCATC ACGGGCACCTGCATCTGCAACCCTGGCTACAAGGGCGAGAGCTGTGAGGAAGTGGACTGCATGG ACCCCACATGTTCAGGCCGGGGTGTCTGCGTGAGAGGCGAATGCCACTGCTCTGTGGGATGGGG AGGCACCAACTGCGAGACCCCCAGGGCCACATGCTTAGACCAGTGTTCAGGCCACGGAACCTTC CTCCCGGACACCGGGCTTTGCAGCTGTGACCCAAGCTGGACTGGACACGACTGTTCTATCGAGA TCTGTGCTGCCGACTGTGGTGGCCATGGCGTGTGCGTAGGGGGCACCTGCCGCTGCGAGGATGG CTGGATGGGGGCAGCCTGCGACCAGCGGGCCTGCCACCCGCGCTGTGCCGAGCATGGGACCTGC CGCGACGGCAAGTGCGAGTGCAGCCCTGGCTGGAATGGCGAACACTGCACCATCGAGGGTTGCC

	CTGGGTTGTGCAATGGCAACGGCAGATGTACCTTAGACCTGAATGGTTGGCACTGCGTCTGCCA GCTGGGCTGGAGAGGAGCTGGCTGT**CTC**AAGGGC		
	ORF Start: at 11		ORF Stop: at 794
	SEQ ID NO: 58	261 aa	MW at 27168.1kD
NOV9i, CG50301-04 Protein Sequence	ESVDNCPSNCYGNGDCISGTCHCFLGFLGPDCGRASCPVLCSGNGQYMKGRCLCHSGWKGAECD VPTNQCIDVACSNHGTCITGTCICNPGYKGESCEEVDCMDPTCSGRGVCVRGECHCSVGWGGTN CETPRATCLDQCSGHGTFLPDTGLCSCDPSWTGHDCSIEICAADCGGHGVCVGGTCRCEDGWMG AACDQRACHPRCAEHGTCRDGKCECSPGWNGEHCTIEGCPGLCNGNGRCTLDLNGWHCVCQLGW RGAGC		
	SEQ ID NO: 59	1399 bp	
NOV9j, CG50301-05 DNA Sequence	CACCGGATCC**GAG**TCGGTGGATAACTGCCCCAGCAACTGCTATGGCAATGGTGACTGCATCTCT GGGACCTGCCACTGCTTCCTGGGTTTCCTGGGCCCCGACTGTGGCAGAGCCTCCTGCCCCGTGC TCTGTAGCGGAAATGGCCAATACATGAAAGGCAGATGCTTGTGCCACAGTGGCTGGAAAGGCGC TGAGTGCGATGTGCCCACCAACCAGTGTATCGATGTGGCCTGCAGCAACCATGGCACCTGCATC ACGGGCACCTGCATCTGCAACCCTGGCTACAAGGGCGAGAGCTGTGAGGAAGTGGACTGCATGG ACCCCACATGTTCAGGCCGGGGTGTCTGCGTGAGAGGCGAATGCCACTGCTCTGTGGGATGGGG AGGCACCAACTGCGAGACCCCCAGGGCCACATGCTTAGACCAGTGTTCAGGCCACGGAACCTTC CTCCCGGACACCGGGCTTTGCAGCTGTGACCCAAGCTGGACTGGACACGACTGTTCTATCGAGA TCTGTGCTGCCGACTGTGGTGGCCATGGCGTGTGCGTAGGGGGCACCTGCCGCTGCGAGGATGG CTGGATGGGGGCAGCCTGCGACCAGCGGGCCTGCCACCCGCGCTGTGCCGAGCATGGGACCTGC CGCGACGGCAAGTGCGAGTGCAGCCCTGGCTGGAATGGCGAACACTGCACCATCGCTCACTATC TGGATAGGGTAGTTAAAGAGGGTTGCCCTGGGTTGTGCAATGGCAACGGCAGATGTACCTTAGA CCTGAATGGTTGGCACTGCGTCTGCCAGCTGGGCTGGAGAGGAGCTGGCTGTGACACTTCCATG GAGACTGCCTGCGGTGACAGCAAAGACAATGATGGAGATGGCCTGGTGGACTGCATGGACCCTG ACTGCTGCCTCCAGCCCCTGTGCCATATCAACCCGCTGTGCCTTGGCTCCCCTAACCCTCTGGA CATCATCCAGGAGACACAGGTCCCTGTGTCACAGCAGAACCTACACTCCTTCTATGACCGCATC AAGTTCCTCGTGGGCAGGGACAGCACGCACATAATCCCCGGGGAGAACCCCTTTGATGGAGGGC ATGCTTGTGTTATTCGTGGCCAAGTGATGACATCAGATGGAACCCCCCTGGTTGGTGTGAACAT CAGTTTTGTCAATAACCCTCTCTTTGGATATACAATCAGCAGGCAAGATGGCAGCTTTGACTTG GTGACAAATGGCGGCATCTCCATCATCCTGCGGTTCGAGCGGGCACCTTTCATCACACAGGAGC ACACCCTGTGGCTGCCATGGGATCGCTTCTTTGTCATGGAAACCATCATCATGAGACATGAGGA GAATGAGATTCCCAGCTGTGACCTGAGCAATTTTGCCCGCCCCAAC**CTC**GAGGGC		
	ORF Start: at 11		ORF Stop: at 1391
	SEQ ID NO: 60	460 aa	MW at 49388.9kD
NOV9j, CG50301-05 Protein Sequence	ESVDNCPSNCYGNGDCISGTCHCFLGFLGPDCGRASCPVLCSGNGQYMKGRCLCHSGWKGAECD VPTNQCIDVACSNHGTCITGTCICNPGYKGESCEEVDCMDPTCSGRGVCVRGECHCSVGWGGTN CETPRATCLDQCSGHGTFLPDTGLCSCDPSWTGHDCSIEICAADCGGHGVCVGGTCRCEDGWMG AACDQRACHPRCAEHGTCRDGKCECSPGWNGEHCTIAHYLDRVVKEGCPGLCNGNGRCTLDLNG WHCVCQLGWRGAGCDTSMETACGDSKDNDGDGLVDCMDPDCCLQPLCHINPLCLGSPNPLDIIQ ETQVPVSQQNLHSFYDRIKFLVGRDSTHIIPGENPFDGGHACVIRGQVMTSDGTPLVGVNISFV NNPLFGYTISRQDGSFDLVTNGGISIILRFERAPFITQEHTLWLPWDRFFVMETIIMRHEENEI PSCDLSNFARPN		
	SEQ ID NO: 61	829 bp	
NOV9k, CG50301-06 DNA Sequence	CACCGGATCC**GAG**TCGGTGGATAACTGCCCCAGCAACTGCTATGGCAATGGTGACTGCATCTCT GGGACCTGCCACTGCTTCCTGGGTTTCCTGGGCCCCGACTGTGGCAGAGCCTCCTGCCCCGTGC TCTGTAGCGGAAATGGCCAATACATGAAAGGCAGATGCTTGTGCCACAGTGGCTGGAAAGGCGC TGAGTGCGATGTGCCCACCAACCAGTGTATCGATGTGGCCTGCAGCAACCATGGCACCTGCATC ACGGGCACCTGCATCTGCAACCCTGGCTACAAGGGCGAGAGCTGTGAGGAAGTGGACTGCATGG ACCCCACATGTTCAGGCCGGGGTGTCTGCGTGAGAGGCGAATGCCACTGCTCTGTGGGATGGGG AGGCACCAACTGCGAGACCCCCAGGGCCACATGCTTAGACCAGTGTTCAGGCCACGGAACCTTC CTCCCGGACACCGGGCTTTGCAGCTGTGACCCAAGCTGGACTGGACACGACTGTTCTATCGAGA TCTGTGCTGCCGACTGTGGTGGCCATGGCGTGTGCGTAGGGGGCACCTGCCGCTGCGAGGATGG CTGGATGGGGGCAGCCTGCGACCAGCGGGCCTGCCACCCGCGCTGTGCCGAGCATGGGACCTGC CGCGACGGCAAGTGCGAGTGCAGCCCTGGCTGGAATGGCGAACACTGCACCATCGCTCACTATC TGGATAGGGTAGTTAAAGAGGGTTGCCCTGGGTTGTGCAATGGCAACGGCAGATGTACCTTAGA CCTGAATGGTTGGCACTGCGTCTGCCAGCTGGGCTGGAGAGGAGCTGGCTGT**CTC**GAGGGC		
	ORF Start: at 11		ORF Stop: at 821
	SEQ ID NO: 62	270 aa	MW at 28250.4kD
NOV9k, CG50301-06 Protein Sequence	ESVDNCPSNCYGNGDCISGTCHCFLGFLGPDCGRASCPVLCSGNGQYMKGRCLCHSGWKGAECD VPTNQCIDVACSNHGTCITGTCICNPGYKGESCEEVDCMDPTCSGRGVCVRGECHCSVGWGGTN CETPRATCLDQCSGHGTFLPDTGLCSCDPSWTGHDCSIEICAADCGGHGVCVGGTCRCEDGWMG		

	AACDQRACHPRCAEHGTCRDGKCECSPGWNGEHCTIAHYLDRVVKEGCPGLCNGNGRCTLDLNG WHCVCQLGWRGAGC

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 9B.

Table 9B. Comparison of NOV9a against NOV9b through NOV9k.

Protein Sequence	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV9b	29..488 371..830	459/460 (99%) 459/460 (99%)
NOV9c	27..488 1..462	461/462 (99%) 461/462 (99%)
NOV9d	216..488 1..273	271/273 (99%) 271/273 (99%)
NOV9e	216..488 1..273	271/273 (99%) 271/273 (99%)
NOV9f	216..488 1..264	262/273 (95%) 262/273 (95%)
NOV9g	29..488 1..460	460/460 (100%) 460/460 (100%)
NOV9h	29..488 371..831	454/464 (97%) 455/464 (97%)
NOV9i	219..488 1..261	261/270 (96%) 261/270 (96%)
NOV9j	219..488 1..270	270/270 (100%) 270/270 (100%)
NOV9k	219..488 1..270	270/270 (100%) 270/270 (100%)

Further analysis of the NOV9a protein yielded the following properties shown in Table 9C.

Table 9C. Protein Sequence Properties NOV9a

SignalP analysis:	Cleavage site between residues 28 and 29
PSORT II analysis:	PSG: a new signal peptide prediction method N-region: length 11; 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): 5.51 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 =  1.54 (at 6)
      ALOM score:   1.54  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       1       Hyd Moment(75): 10.04
      Hyd Moment(95):  8.14    G content:       1
      D/E content:     2       S/T content:     3
      Score: -4.92

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 15  GRA|TM

NUCDISC: discrimination of nuclear localization signals
      pat4: none
      pat7: PDRKGKG (4) at   67
      bipartite: none
      content of basic residues:   7.4%
      NLS Score: -0.13

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

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

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):

        47.8 %: mitochondrial
        13.0 %: extracellular, including cell wall
        13.0 %: cytoplasmic
        13.0 %: nuclear
         4.3 %: vacuolar
         4.3 %: endoplasmic reticulum
         4.3 %: peroxisomal

>> prediction for CG50301-07 is mit (k=23)
```

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
ABG97359	Human CGDD10, INCYTE 7488573CD1 - Homo sapiens, 2758 aa. [WO200272830-A2, 19-SEP-2002]	29..488 367..826	459/460 (99%) 459/460 (99%)	0.0
ABG70388	Human TEN-M4-like protein - Homo sapiens, 2769 aa. [WO200257453-A2, 25-JUL-2002]	29..488 371..830	459/460 (99%) 459/460 (99%)	0.0
ABG97359	Human CGDD10, INCYTE 7488573CD1 - Homo sapiens, 2758 aa. [WO200272830-A2, 19-SEP-2002]	29..488 367..826	459/460 (99%) 459/460 (99%)	0.0
ABB98401	Human NOV1, a TEN-M4 like protein - Homo sapiens, 2794 aa. [WO200255704-A2, 18-JUL-2002]	29..488 371..831	398/464 (85%) 419/464 (89%)	0.0
ABP53588	Human NOV15c protein SEQ ID NO:40 - Homo sapiens, 2628 aa. [WO200262999-A2, 15-AUG-2002]	29..488 239..695	312/466 (66%) 354/466 (75%)	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

Q9WTS7	Ten-m4 - Mus musculus (Mouse), 2771 aa.	29..488 371..832	449/462 (97%) 451/462 (97%)	0.0
O70465	DOC4 - Mus musculus (Mouse), 2825 aa.	29..488 441..893	438/462 (94%) 441/462 (94%)	0.0
Q9W7R3	Ten-m4 - Brachydanio rerio (Zebrafish) (Danio rerio), 2824 aa.	29..488 442..892	345/462 (74%) 392/462 (84%)	0.0
Q9WTS6	Ten-m3 - Mus musculus (Mouse), 2715 aa.	29..488 339..782	310/460 (67%) 352/460 (76%)	0.0
Q9JLC1	ODZ3 - Mus musculus (Mouse), 2346 aa (fragment).	50..488 5..420	297/439 (67%) 335/439 (75%)	0.0

PFam analysis predicts 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
EGF	288..315	15/47 (32%) 21/47 (45%)	0.014
EGF	320..347	11/47 (23%) 22/47 (47%)	0.21
EGF	418..444	13/47 (28%) 24/47 (51%)	0.0045
EGF	458..488	10/47 (21%) 21/47 (45%)	0.66

Example 10.

The NOV10 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 10A.

Table 10A. NOV10 Sequence Analysis

	SEQ ID NO: 63	1352 bp	
NOV10a, CG56201-03 DNA Sequence	CCTGCAGGTTCTCTCAGCCCCTTTTCACAATATTTGATTAGGAATTTGGGGCGGGACCCTGGTC TGGCACAGGCACGCACACTCTCAGTAGACTCTTTCACTCCTCTCTCTCTTCCTCTCTCACACGT TCTCCAACCCAAGGAGGCCAGACAGAGGGACGTGGTCACTCTCTGAAAAGTTCAACTTGAGAGA CAAA**ATG**CAGTGGACCTCCCTCCTGCTGCTGGCAGGGCTCTTCTCCCTCTCCCAGGCCCAGTAT GAAGATGACCCTCATTGGTGGTTCCACTACCTCCGCAGCCAGCAGTCCACCTACTACGATCCCT ATGACCCTTACCCGTATGAGACCTACGAGCCTTACCCCTATGGGGTGGATGAAGGGCCAGCCTA CACCTACGGCTCTCCATCCCCTCCAGATCCCCGCGACTGCCCCCAGGAATGCGACTGCCCACCC AACTTCCCCACGGCCATGTACTGTGACAATCGCAACCTCAAGTACCTGCCCTTCGTTCCCTCCC GCATGAAGTATGTGTACTTCCAGAACAACCAGATCACCTCCATCCAGGAAGGCGTCTTTGACAA TGCCACAGGGCTGCTCTGGATTGCTCTCCACGGCAACCAGATCACCAGTGATAAGGTGGGCAGG AAGGTCTTCTCCAAGCTGAGGCACCTGGAGAGGCTGTACCTGGACCACAACAACCTGACCCGGA TGCCCGGTCCCCTGCCTCGATCCCTGAGAGAGCTCCATCTCGACCACAACCAGATCTCACGGGT		

	CCCCAACAATGCTCTGGAGGGGCTGGAGAACCTCACGGCCTTGTACCTCCAACACAATGAGATC CAGGAAGTGGGCAGTTCCATGAGGGGCCTCCGGTCACTGATCTTGCTGGACCTGAGTTATAACC ACCTTCGGAAGGTGCCTGATGGGCTGCCCTCAGCTCTTGAGCAGCTGTACATGGAGCACAACAA TGTCTACACCGTCCCCGATAGCTACTTCCGGGGGGCGCCCAAGCTGCTGTATGTGCGGCTGTCC CACAACAGTCTAACCAACAATGGCCTGGCCTCCAACACCTTCAATTCCAGCAGCCTCCTTGAGC TAGACCTCTCCTACAACCAGCTGCAGAAGATCCCCCCAGTCAACACCAACCTGGAGAACCTCTA CCTCCAAGGCAATAGGATCAATGAGTTCTCCATCAGCAGCTTCTGCACCGTGGTGGACGTCGTG AACTTCTCCAAGCTGCAGGTGCTGCGCCTGGACGGGAACGAGATCAAGCGCAGCGCCATGCCTG CCGACGCGCCCCTCTGCCTGCGCCTTGCCAGCCTCATCGAGATC**TGA**GCAGGCCTGTGGCATTG TCAAAGAA		
	ORF Start: ATG at 197		ORF Stop: TGA at 1325
	SEQ ID NO: 64	376 aa	MW at 43178.3kD
NOV10a, CG56201-03 Protein Sequence	MQWTSLLLLAGLFSLSQAQYEDDPHWWFHYLRSQQSTYYDPYDPYPYETYEPYPYGVDEGPAYT YGSPSPPDPRDCPQECDCPPNFPTAMYCDNRNLKYLPFVPSRMKYVYFQNNQITSIQEGVFDNA TGLLWIALHGNQITSDKVGRKVFSKLRHLERLYLDHNNLTRMPGPLPRSLRELHLDHNQISRVP NNALEGLENLTALYLQHNEIQEVGSSMRGLRSLILLDLSYNHLRKVPDGLPSALEQLYMEHNNV YTVPDSYFRGAPKLLYVRLSHNSLTNNGLASNTFNSSSLLELDLSYNQLQKIPPVNTNLENLYL QGNRINEFSISSFCTVVDVVNFSKLQVLRLDGNEIKRSAMPADAPLCLRLASLIEI		
	SEQ ID NO: 65	1139 bp	
NOV10b, CG56201-04 DNA Sequence	AAGGAGGCCAGACAGAGGGACGTGGTCACTTCTCTGAAAAGTTCAACTTGAGCAAA**ATG**CAGTG GACCTCCCTCCTGCTGCTGGCAGGGCTCTTCTCCCTCTCCCAGGCCCAGTATGAAGATGACCCT CATTGGTGGTTCCACTACCTCCGCAGCCAGCAGTCCACCTACTACGATCCCTATGACCCTTACC CGTATGAGACCTACGAGCCTTACCCCTATGGGGTGGATGAAGGGCCAGCCTACACCTACGGCTC TCCATCCCCTCCAGATCCCCGCGACTGCCCCCAGGAGTGCGACTGCCCACCCAACTTCCCCACG GCCATGTACTGTGACAATCGCAACCTCAAGTACCTGCCCTTCGTTCCCTCCCGCATGAAGTATG TGTACTTCCAGAACAACCAGATCACCTCCATCCAGGAAGGCGTCTTTGACAATGCCACAGGGCT GCTCTGGATTGCTCTCCACGGCAACCAGATCACCAGTGATAAGGTGGGCAGGAAGGTCTTCTCC AAGCTGAGGCACCTGGAGAGGCTGTACCTGGACCACAACAACCTGACCCGGATGCCCGGTCCCC TGCCTCGATCCCTGAGAGAGCTCCATCTCGACCACAACCAGATCTCACGGGTCCCCAACAATGC TCTGGAGGGGCTGGAGAACCTCACGGCCTTGTACCTCCAACACAATGAGATCCAGGAAGTGGGC AGTTCCATGAGGGGCCTCCGGTCACTGTACTTGCTGGACCTGAGTTATAACCACCTTCGGAAGG TGCCTGATGGGCTGCCCTCAGCTCTTGAGCAGCTGTACATGGAGCACAACAATGTCTACACCGT CCCCGATAGCTACTTCCGGGGGGCGCCCAAGCTGCTGTATGTGCGGCTGTCCCACAACAGTCTA ACCAACAATGGCCTGGCCTCCAACACCTTCAATTCCAGCAGCCTCCTTGAGCTAGACCTCTCCT ACAACCAGCTGCAGAAGATCCCCCCAGTCAACACCATCAGCAGCTTCTGCACCGTGGTGGACGT CGTGAACTTCTCCCAGCTGCAGGTCGTGCGGCTGGACGGGAACGAGATGAAGCGGAGCGCCATG CCTGCCGAGGCGCCCCTCTGCCTGCGCCTTGCCAGCCTCATCGAGATC**TGA**		
	ORF Start: ATG at 57		ORF Stop: TGA at 1137
	SEQ ID NO: 66	360 aa	MW at 41340.2kD
NOV10b, CG56201-04 Protein Sequence	MQWTSLLLLAGLFSLSQAQYEDDPHWWFHYLRSQQSTYYDPYDPYPYETYEPYPYGVDEGPAYT YGSPSPPDPRDCPQECDCPPNFPTAMYCDNRNLKYLPFVPSRMKYVYFQNNQITSIQEGVFDNA TGLLWIALHGNQITSDKVGRKVFSKLRHLERLYLDHNNLTRMPGPLPRSLRELHLDHNQISRVP NNALEGLENLTALYLQHNEIQEVGSSMRGLRSLYLLDLSYNHLRKVPDGLPSALEQLYMEHNNV YTVPDSYFRGAPKLLYVRLSHNSLTNNGLASNTFNSSSLLELDLSYNQLQKIPPVNTISSFCTV VDVVNFSQLQVVRLDGNEMKRSAMPAEAPLCLRLASLIEI		
	SEQ ID NO: 67	903 bp	
NOV10c, 247618257 DNA Sequence	**GGA**TCCCAGTATGAAGATGACCCTCATTGGTGGTTCCACTACCTCCGCAGCCAGCAGTCCACCTA CTACGATCCCTATGACCCTTACCCGTATGAGACCTACGAGCCTTACCCCTATGGGGTGGATGAAG GGCCAGCCTACACCTACGGTTCTCCATCCCCTCCAGATCCCCGCGACTGCCCCCAGGAGTGCGAC TGCCCACCCAACTTCCCCACGGCCATGTACTGTGACAATCGCAACCTCAAGTACCTGCCCTTCGT TCCCTCCCGCATGAAGTATGTGTACTTCCAGAACAACCAGATCACCTCCATCCAGGAAGGCGTCT TTGACAATGCCACAGGGCTGCTCTGGATTGCTCTCCACGGCAACCAGATCACCAGTGATAAGGTG GGCAGGAAGGTCTTCTCCAAGCTGAGGCACCTGGAGAGGCTGTACCTGGACCACAACAACCTGAC CCGGATGCCCGGTCCCCTGCCTCGATCCCTGAGAGAGCTCCATCTCGACCACAACCAGATCTCAC GGGTCCCCAACAATGCTCTGGAGGGGCTGGAGAACCTCACGGCCTTGTACCTCCAACACAATGAG ATCCAGGAAGTGGGCAGTTCCATGAGGGGCCTCCGGTCACTGATCTTGCTGGACCTGAGTTATAA CCACCTTCGGAAGGTGCCTGATGGGCTGCCCTCAGCTCTTGAGCAGCTGTACATGGAGCACAACA ATGTCTACACCGTCCCCGATAGCTACTTCCGGGGGGCGCCCAAGCTGCTGTATGTGCGGCTGTCC CACAACAGTCTAACCAACAATGGCCTGGCCTCCAACACCTTCAATTCCAGCAGCCTCCTTGAGCT AGACCTCTCCTACAACCAGCTGCAGAAGATCCCCCCAGTCAACACCATCAGCCTCGAG		
	ORF Start: at 1		ORF Stop: end of sequence
	SEQ ID NO: 68	301 aa	MW at 34768.5kD

NOV10c, 247618257 Protein Sequence	GSQYEDDPHWWFHYLRSQQSTYYDPYDPYPYETYEPYPYGVDEGPAYTYGSPSPPDPRDCPQEC DCPPNFPTAMYCDNRNLKYLPFVPSRMKYVYFQNNQITSIQEGVFDNATGLLWIALHGNQITSD KVGRKVFSKLRHLERLYLDHNNLTRMPGPLPRSLRELHLDHNQISRVPNNALEGLENLTALYLQ HNEIQEVGSSMRGLRSLILLDLSYNHLRKVPDGLPSALEQLYMEHNNVYTVPDSYFRGAPKLLY VRLSHNSLTNNGLASNTFNSSSLLELDLSYNQLQKIPPVNTISLE		
	SEQ ID NO: 69	1455 bp	
NOV10d, CG56201-01 DNA Sequence	CCTGCAGGTTCTCTCAGCCCCTTTTCACAATATTTGATTAGGAATTTGGGGCGGGACCCTGGTC TGGCACAGGCACGCACACTCTCAGTAGACTCTTTCACTCCTCTCTCTCTTCCTCTCTCACACGT TCTCCAACCCAAGGAGGCCAGACAGAGGGACGTGGTCACTCTCTGAAAAGTTCAACTTGAGAGA CAAA**ATG**CAGTGGACCTCCCTCCTGCTGCTGGCAGGGCTCTTCTCCCTCTCCCAGGCCCAGTAT GAAGATGACCCTCATTGGTGGTTCCACTACCTCCGCAGCCAGCAGTCCACCTACTACGATCCCT ATGACCCTTACCCGTATGAGACCTACGAGCCTTACCCCTATGGGGTGGATGAAGGGCCAGCCTA CACCTACGGCTCTCCATCCCCTCCAGATCCCCGCGACTGCCCCCAGGAATGCGACTGCCCACCC AACTTCCCCACGGCCATGTACTGTGACAATCGCAACCTCAAGTACCTGCCCTTCGTTCCCTCCC GCATGAAGTATGTGTACTTCCAGAACAACCAGATCACCTCCATCCAGGAAGGCGTCTTTGACAA TGCCACAGGGCTGCTCTGGATTGCTCTCCACGGCAACCAGATCACCAGTGATAAGGTGGGCAGG AAGGTCTTCTCCAAGCTGAGGCACCTGGAGAGGCTGTACCTGGACCACAACAACCTGACCCGGA TGCCCGGTCCCCTGCCTCGATCCCTGAGAGAGCTCCATCTCGACCACAACCAGATCTCACGGGT CCCCAACAATGCTCTGGAGGGGCTGGAGAACCTCACGGCCTTGTACCTCCAACACAATGAGATC CAGGAAGTGGGCAGTTCCATGAGGGGCCTCCGGTCACTGATCTTGCTGGACCTGAGTTATAACC ACCTTCGGAAGGTGCCTGATGGGCTGCCCTCAGCTCTTGAGCAGCTGTACATGGAGCACAACAA TGTCTACACCGTCCCCGATAGCTACTTCCGGGGGGCGCCCAAGCTGCTGTATGTGCGGCTGTCC CACAACAGTCTAACCAACAATGGCCTGGCCTCCAACACCTTCAATTCCAGCAGCCTCCTTGAGC TAGACCTCTCCTACAACCAGCTGCAGAAGATCCCCCCAGTCAACACCAACCTGGAGAACCTCTA CCTCCAAGGCAATAGGATCAATGAGTTCTCCATCCAGGAAGGCGTCTTTGACAATGCCACAGGG CTGCTCTGGATTGCTCTCCACGGCAACTTCTCCACGGCCATGTACTGTGACAATCGCAACCTCA AGTACCTGCCCTTCGTTCCCTCCCGCATGAAGTATGTGTACTTCCAGAACAACCAGATCACCTC CAAGCTGCAGGTGCTGCGCCTGGACGGGAACGAGATCAAGCGCAGCGCCATGCCTGCCGACGCG CCCCTCTGCCTGCGCCTTGCCAGCCTCATCGAGATC**TGA**GCAGCCCT		
	ORF Start: ATG at 197		ORF Stop: TGA at 1445
	SEQ ID NO: 70	416 aa	MW at 47952.7kD
NOV10d, CG56201-01 Protein Sequence	MQWTSLLLLAGLFSLSQAQYEDDPHWWFHYLRSQQSTYYDPYDPYPYETYEPYPYGVDEGPAYT YGSPSPPDPRDCPQECDCPPNFPTAMYCDNRNLKYLPFVPSRMKYVYFQNNQITSIQEGVFDNA TGLLWIALHGNQITSDKVGRKVFSKLRHLERLYLDHNNLTRMPGPLPRSLRELHLDHNQISRVP NNALEGLENLTALYLQHNEIQEVGSSMRGLRSLILLDLSYNHLRKVPDGLPSALEQLYMEHNNV YTVPDSYFRGAPKLLYVRLSHNSLTNNGLASNTFNSSSLLELDLSYNQLQKIPPVNTNLENLYL QGNRINEFSIQEGVFDNATGLLWIALHGNFSTAMYCDNRNLKYLPFVPSRMKYVYFQNNQITSK LQVLRLDGNEIKRSAMPADAPLCLRLASLIEI		
	SEQ ID NO: 71	965 bp	
NOV10e, CG56201-02 DNA Sequence	AAGGAGGCCAGACAGAGGGACGTGGTCACTTCTCTGAAAAGTTCAACTTGAGCAAA**ATG**CAGTG GACCTCCCTCCTGCTGCTGGCAGGGCTCTTCTCCCTCTCCCAGGCCCAGTATGAAGATGACCCT CATTGGTGGTTCCACTACCTCCGCAGCCAGCAGTCCACCTACTACGATCCCTATGACCCTTACC CGTATGAGACCTACGAGCCTTACCCCTATGGGGTGGATGAAGGGCCAGCCTACACCTACGGCTC TCCATCCCCTCCAGATCCCCGCGACTGCCCCCAGGAATGCGACTGCCCACCCAACTTCCCCACG GCCATGTACTGTGACAATCGCAACCTCAAGTACCTGCCTCGATCCCTGAGAGAGCTCCATCTCG ACCACAACCAGATCTCACGGGTCCCCAACAATGCTCTGGAGGGGCTGGAGAACCTCACGGCCTT GTACCTCCAACACAATGAGATCCAGGAAGTGGGCAGTTCCATGAGGGGCCTCCGGTCACTGTAC TTGCTGGACCTGAGTTATAACCACCTTCGGAAGGTGCCTGATGGGCTGCCCTCAGCTCTTGAGC AGCTGTACATGGAGCACAACAATGTCTACACCGTCCCCGATAGCTACTTCCGGGGGGCGCCCAA GCTGCTGTATGTGCGGCTGTCCCACAACAGTCTAACCAACAATGGCCTGGCCTCCAACACCTTC AATTCCAGCAGCCTCCTTGAGCTAGACCTCTCCTACAACCAGCTGCAGAAGATCCCCCCAGTCA ACACCAACCTGGAGAACCTCTACCTCCAAGGCAATAGGATCAATGAGTTCTCCATCAGCAGCTT CTGCACCGTGGTGGACGTCGTGAACTTCTCCCAGCTGCAGGTCGTGCGGCTGGACGGGAACGAG ATGAAGCGGAGCGCCATGCCTGCCGAGGCGCCCCTCTGCCTGCGCCTTGCCAGCCTCATCGAGA TC**TGA**		
	ORF Start: ATG at 57		ORF Stop: TGA at 963
	SEQ ID NO: 72	302 aa	MW at 34678.5kD
NOV10e, CG56201-02 Protein Sequence	MQWTSLLLLAGLFSLSQAQYEDDPHWWFHYLRSQQSTYYDPYDPYPYETYEPYPYGVDEGPAYT YGSPSPPDPRDCPQECDCPPNFPTAMYCDNRNLKYLPRSLRELHLDHNQISRVPNNALEGLENL TALYLQHNEIQEVGSSMRGLRSLYLLDLSYNHLRKVPDGLPSALEQLYMEHNNVYTVPDSYFRG APKLLYVRLSHNSLTNNGLASNTFNSSSLLELDLSYNQLQKIPPVNTNLENLYLQGNRINEFSI SSFCTVVDVVNFSQLQVVRLDGNEMKRSAMPAEAPLCLRLASLIEI		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 10B.

Table 10B. Comparison of NOV10a against NOV10b through NOV10e.

Protein Sequence	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV10b	1..376 1..360	355/376 (94%) 359/376 (95%)
NOV10c	18..313 2..297	295/296 (99%) 296/296 (99%)
NOV10d	1..376 1..416	367/416 (88%) 369/416 (88%)
NOV10e	1..376 1..302	293/376 (77%) 298/376 (78%)

Further analysis of the NOV10a protein yielded the following properties shown in Table 10C.

Table 10C. Protein Sequence Properties NOV10a

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 20; peak value 10.61 PSG score: 6.21 GvH: von Heijne's method for signal seq. recognition GvH score (threshold: -2.1): 8.80 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 = 3.66 (at 332) ALOM score: 3.66 (number of TMSs: 0) MTOP: Prediction of membrane topology (Hartmann et al.) Center position for calculation: 9 Charge difference: -2.0 C(-1.0) - N(1.0) N >= C: N-terminal side will be inside MITDISC: discrimination of mitochondrial targeting seq R content: 0 Hyd Moment(75): 0.11 Hyd Moment(95): 3.44 G content: 1 D/E content: 1 S/T content: 4 Score: -5.41 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 ***
         LYLDHNNLTRMPGPLPRSLREL at 160
         LDLSYNHLRKVPDGLPSALEQL at 228
 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):

        66.7 %: extracellular, including cell wall
        22.2 %: mitochondrial
        11.1 %: nuclear

>> prediction for CG56201-03 is exc (k=9)
```

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**
AAB43747	Human cancer associated protein sequence SEQ ID NO:1192 - Homo sapiens, 415 aa. [WO200055350-A1, 21-SEP-2000]	1..376 40..415	376/376 (100%) 376/376 (100%)	0.0
AAG78508	Human fibromodulin amino acid sequence - Homo sapiens, 376 aa. [US6277812-B1, 21-AUG-2001]	1..376 1..376	372/376 (98%) 374/376 (98%)	0.0
AAW26404	Human fibromodulin - Homo sapiens, 376 aa. [US5654270-A, 05-AUG-1997]	1..376 1..376	372/376 (98%) 374/376 (98%)	0.0
AAR36773	Human fibromodulin - Homo sapiens, 376 aa. [WO9309800-A, 27-MAY-1993]	1..376 1..376	372/376 (98%) 374/376 (98%)	0.0
ABB57321	Mouse ischaemic condition related protein sequence SEQ ID NO:892 - Mus musculus, 376 aa. [WO200188188-A2, 22-NOV-2001]	1..376 1..376	348/376 (92%) 364/376 (96%)	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**
AAH35281	Similar to fibromodulin - Homo sapiens (Human), 376 aa.	1..376 1..376	376/376 (100%) 376/376 (100%)	0.0
Q06828	Fibromodulin precursor (FM) (Collagen-binding 59 kDa protein) (Keratan sulfate proteoglycan fibromodulin) (KSPG fibromodulin) - Homo sapiens (Human), 376 aa.	1..376 1..376	371/376 (98%) 375/376 (99%)	0.0
S55275	fibromodulin precursor - human, 376 aa.	1..376 1..376	372/376 (98%) 374/376 (98%)	0.0

P50608	Fibromodulin precursor (FM) (Collagen-binding 59 kDa protein) (Keratan sulfate proteoglycan fibromodulin) (KSPG fibromodulin) - Mus musculus (Mouse), 376 aa.	1..376 1..376	348/376 (92%) 364/376 (96%)	0.0
BAC37953	10 days neonate cortex cDNA, RIKEN full-length enriched library, clone:A830016K08 product:fibromodulin, full insert sequence - Mus musculus (Mouse), 376 aa.	1..376 1..376	347/376 (92%) 364/376 (96%)	0.0

PFam analysis predicts 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
LRRNT	75..104	12/32 (38%) 27/32 (84%)	9.9e-11
LRR	177..200	10/25 (40%) 22/25 (88%)	0.003
LRR	294..313	11/25 (44%) 18/25 (72%)	0.65

Example 11. NOV 11, CG56653, Ficolin.

The NOV11 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 11A.

Table 11A. NOV11 Sequence Analysis

	SEQ ID NO: 73	1045 bp	
NOV11a, CG56653-11 DNA Sequence	GAGCTGGGGGACTCTTCAGAGTCAAAGGCCAGAGAGC**ATG**GAGCTGAGTGGAGCCACCATGGCC CGGGGGCTCGCTGTCCTGCTAGTCTTGTTCCTGCATATCAAGAACCTGCCTGCCCAGGCTGCGG ACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGACAAGCTCACCATTCTCCGAGG CTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCAGGTGTCATTGGAGAGAGAGGA GAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGACCAGTGGGGCCCAAAGGAGACCGAGGAG AGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTCGTGTGCGACAGGCCCACG CAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTGGCACACCATCTACCTGCCC GACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAGGGGGCTGGACCGTTTTCC AGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGCCGCATACAAGCAGGGCTTCGG CAGTCAGCTGGGGGAGTTCTGGCTGGGGAACGACAACATCCACGCCCTGACTGCCCAGGGAAGC AGCGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACCAGTTTGCTAAGTACAAATCAT TCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCCTTTGTCGGGGGCAGTGC GGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAAGACCAAGACAATGATGTG AGTTCTTCGAATTGTGCTGAGAAGTTCCAGGGAGCCTGGTGGTACGCCGACTGTCATGCTTCAA		

	ACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAATGGTATCAACTGGAGTGC GGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGCGGCCCGCC**TAG**ACGGGC CAGGACCCCTCCACATGCACC		
	ORF Start: ATG at 38		ORF Stop: TAG at 1016
	SEQ ID NO: 74	326 aa	MW at 35078.1kD
NOV11a, CG56653-11 Protein Sequence	MELSGATMARGLAVLLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKG EAGVIGERGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFL SGWHTIYLPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQGFGSQLGEFWLGNDN IHALTAQGSSELRVDLVDFEGNHQFAKYKSFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFF STKDQDNDVSSSNCAEKFQGAWWYADCHASNLNGLYLMGPHESYANGINWSAAKGYKYSYKVSE MKVRPA		
	SEQ ID NO: 75	1187 bp	
NOV11b, CG56653-01 DNA Sequence	CTGAGTGGAGCCACC**ATG**GCCCGGGGGCTCGCTGTCCTGCTAGTCTTGTTCCTGCATATCAAGA ACCTGCCTGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGA CAAGCTCACCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCA GGTGTCATTGGAGAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGACCAGTGG GGCCCAAAGGAGACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCA GTCGTGTGCGACAGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGC TGGCACACCATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACG GAGGGGGCTGGACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGC CGCATACAAGCAGGGCTTCGGCAGTCAGCTGGGGGAGTTCTGGCTGGGGAACGACAACATCCAC GCCCTGACTGCCCAGGGAAGCAGCGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACC AGTTTGCTAAGTACAAATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGG AGCCTTTGTCGGGGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACC AAAGACCAAGACAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAGGGAGCCTGGTGGT ACGCCGACTGTCATGCTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGC CAATGGTATCAACTGGAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAG GTGCGGCCCGCC**TAG**ACGGGCCAGGACCCCTCCACATGCACCTGCTAGTGGGGAGGCCACACCC ACAAGCGCTGCGTCGTGGAAGTCACCCCATTTCCCCAGCCAGACACACTCCCATGACGCCCACA GCTGCCCCTTTGCCCCCAGCTCAGTCAAGCCGCCACATGCCCACAACCTCACCAGAGGGAGAAT TATGTTTCTAAATATGTTTACTTTGGGACAGAAAA		
	ORF Start: ATG at 16		ORF Stop: TAG at 973
	SEQ ID NO: 76	319 aa	MW at 34388.3kD
NOV11b, CG56653-01 Protein Sequence	MARGLAVLLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKGEAGVIGE RGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFLSGWHTIY LPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQGFGSQLGEFWLGNDNIHALTAQ GSSELRVDLVDFEGNHQFAKYKSFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFFSTKDQDN DVSSSNCAEKFQGAWWYADCHASNLNGLYLMGPHESYANGINWSAAKGYKYSYKVSEMKVRPA		
	SEQ ID NO: 77	900 bp	
NOV11c, CG56653-05 DNA Sequence	GGATCC**GCG**GACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGACAAGCTCACCA TTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCAGGTGTCATTGG AGAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGACCAGTGGGGCCCAAAGGA GACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTCGTGTGCGA CAGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTGGCACACCAT CTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAGGGGGCTGG ACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGCCGCATACAAGC AGGGCTTCGGCAGTCAGCTGGGGGAGTTCTGGCTGGGGAACGACAACATCCACGCCCTGACTGC CCAGGGAAGCAGCGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACCAGTTTGCTAAG TACAAATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCCTTTGTCG GGGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAAGACCAAGA CAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAGGGAGCCTGGTGGTACGCCGACTGT CATGCTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAATGGTATCA ACTGGAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGGGCCCGCC**CT** **C**GAG		
	ORF Start: at 7		ORF Stop: at 895
	SEQ ID NO: 78	296 aa	MW at 31819.1kD
NOV11c, CG56653-05 Protein Sequence	ADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKGEAGVIGERGERGLPGAPGKAGPVGPKGDR GEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFLSGWHTIYLPDCRPLTVLCDMDTDGGGWTV FQRRMDGSVDFYRDWAAYKQGFGSQLGEFWLGNDNIHALTAQGSSELRVDLVDFEGNHQFAKYK SFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFFSTKDQDNDVSSSNCAEKFQGAWWYADCHA SNLNGLYLMGPHESYANGINWSAAKGYKYSYKVSEMKGPA		

	SEQ ID NO: 79	970 bp	
NOV11d, CG56653-07 DNA Sequence	GGGGCTCGCTGTCATGCTAGTCTTGTTCCTGCATATCAAGAACCTGCCTGCCCAGGCTGCGGAC ACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGACAAGCTCACCATTCTCCGAGGCT GCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCAGGTGTCATTGGAGAGAGAGGAGA ACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGACCAGTGGGGCCCAAAGGAGACCGAGGAGAG AAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTCGTGTGCGACAGGCCCACGCA ACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTGGCACACCATCTACCTGCCCGA CTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAGGGGGCTGGACCGTTTTCCAG CGGAGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGCCGCATACAAGCAGGGCTTCGGCA GTCAGCTGGGGGAGTTCTGGCTGGGGAATGACAACATCCACGCCTTGACTGCCCAGGGAAGCAG CGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACCAGTTTGCTAAGTACAAATCATTC AAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCCTTTGTCGGGGGCAGTGCGG GTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAAGACCAAGACAATGATGTGAG TTCTTCGAATTGTGCTGAGAAGTTCCAAGGAGCCTGGTGGTACGCCGACTGTCATGCTTCAAAC CTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAATGGTATCAACTGGAGTGCGG CGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGCGGCCCGCCTAGACGGGCCA GGACCCCTCC		
	ORF Start: at 2		ORF Stop: TAG at 950
	SEQ ID NO: 80	316 aa	MW at 34047.9kD
NOV11d, CG56653-07 Protein Sequence	GLAVMLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKGEAGVIGERGE RGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFLSGWHTIYLPD CRPLTVLCDMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQGFGSQLGEFWLGNDNIHALTAQGSS ELRVDLVDFEGNHQFAKYKSFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFFSTKDQDNDVS SSNCAEKFQGAWWYADCHASNLNGLYLMGPHESYANGINWSAAKGYKYSYKVSEMKVRPA		
	SEQ ID NO: 81	988 bp	
NOV11e, CG56653-06 DNA Sequence	GTTTCCTCTGCCCTGTTAGAGCTGGGGGACTCTTCAGAGTCAAAGGCCAGAGAGCATGGAGCTG AGTGGAGCCACCATGGCCCGGGGGCTCGCTGTCCTGCTAGTCTTGTTCCTGCATATCAAGAACC TGCCTGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGACAA GCTCACCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCAGGT GTCATTGGAGAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGACCAGTGGGGC CCAAAGGAGACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTC GTGTGCGACAGGTCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTGG CACACCATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAG GGGGCTGGACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTTGAGGGCAACCACCAGTT TGCTAAGTACAAATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCC TTTGTCGGGGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAAG ACCAAGACAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAAGGAGCCTGGTGGTACGC CGACTGTCATGCTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAAT GGTATCAACTGCAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGC GGCCCGCCTAGACGGGCCAGGACCCCTCCACATGCACCTGCTAGTGGGGAGGCCACACCCACAA GCGCTGCGTCGTGGAAGTCACCCATTTC		
	ORF Start: ATG at 56		ORF Stop: TAG at 905
	SEQ ID NO: 82	283 aa	MW at 30136.7kD
NOV11e, CG56653-06 Protein Sequence	MELSGATMARGLAVLLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKG EAGVIGERGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFL SGWHTIYLPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFEGNHQFAKYKSFKVADEAEKYKLV LGAFVGGSAGNSLTGHNNNFFSTKDQDNDVSSSNCAEKFQGAWWYADCHASNLNGLYLMGPHES YANGINCSAAKGYKYSYKVSEMKVRPA		
	SEQ ID NO: 83	871 bp	
NOV11f, 247675490 DNA Sequence	AGGCTCCGCGGCCGCCCCCTTCACCACCATGGCCCGGGGGCTCGCTGTCCTGCTAGTCTTGTTCC TGCATATCAAGAACCTGCCTGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTG GAGGGCTCTGACAAGCTCACCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAA GGGAGAGGCAGGTGTCATTGGAGAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAG GACCAGTGGGGCCCAAAGGAGACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGG CAGTCTCAGTCGTGTGCGACAGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCT GAGCGGCTGGCACACCATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACA CGGACGGAGGGGGCTGGACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTTGAGGGCAAC CACCAGTTTGCTAAGTACAAATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACT GGGAGCCTTTGTCGGGGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCA CCAAAGACCAAGACAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAAGGAGCCTGGTGG TACGCCGACTGTCATGCTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGC CAATGGTATCAACTGCAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGG		

	TGCGGCCCGCCAAGGGTGGGCGCGCC		
	ORF Start: at 2		ORF Stop: end of sequence
	SEQ ID NO: 84	290 aa	MW at 30720.4kD
NOV11f, 247675490 Protein Sequence	GSAAAPFTTMARGLAVLLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGP KGEAGVIGERGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGY FLSGWHTIYLPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFEGNHQFAKYKSFKVADEAEKYK LVLGAFVGGSAGNSLTGHNNNFFSTKDQDNDVSSSNCAEKFQGAWWYADCHASNLNGLYLMGPH ESYANGINCSAAKGYKYSYKVSEMKVRPAKGGRA		
	SEQ ID NO: 85	593 bp	
NOV11g, 247675499 DNA Sequence	AGGCTCCGCGGCCGCCAGTACTCGAGAAAGGGCGCGCCCACCCTTCACCGGCCCACGCAACTGC AAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTGGCACACCATCTACCTGCCCGACTGCC GGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAGGGGGCTGGACCGTTTTCCAGCGGAG GATGGATGGCTCTGTGGACTTTGAGGGCAACCACCAGTTTGCTAAGTACAAATCATTCAAGGTG GCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCCTTTGTCGGGGGCAGTGCGGGTAATT CTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAAGACCAAGACAATGATGTGAGTTCTTC GAATTGTGCTGAGAAGTTCCAAGGAGCCTGGTGGTACGCCGACTGTCATGCTTCAAACCTCAAT GGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAATGGTATCAACTGCAGTGCGGCGAAGG GGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGCGGCCCGCCAAGGGTGGGCGCGCCGA GCCAGGTTTCTTGTACA		
	ORF Start: at 2		ORF Stop: at 593
	SEQ ID NO: 86	197 aa	MW at 21748.0kD
NOV11g, 247675499 Protein Sequence	GSAAASTRERARPPFTGPRNCKDLLDRGYFLSGWHTIYLPDCRPLTVLCDMDTDGGGWTVFQRR MDGSVDFEGNHQFAKYKSFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFFSTKDQDNDVSSS NCAEKFQGAWWYADCHASNLNGLYLMGPHESYANGINCSAAKGYKYSYKVSEMKVRPAKGGRAE PGFLY		
	SEQ ID NO: 87	910 bp	
NOV11h, 277765267 DNA Sequence	GCCGGATCCGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGACAAGCTCA CCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCAGGTGTCAT TGGAGAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGACCAGTGGGGCCCAAA GGAGACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTCGTGTG CGACAGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTGGCACAC CATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAGGGGGC TGGACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGCCGCATACA AGCAGGGCTTCGGCAGTCAGCTGGGGGAGTTCTGGCTGGGGAATGACAACATCCACGCCTTGAC TGCCCAGGGAAGCAGCGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACCAGTTTGCT AAGTACAAATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCCTTTG TCGGGGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAAGACCA AGACAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAAGGAGCCTGGTGGTACGCCGAC TGTCATGCTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAATGGTA TCAACTGGAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGCGGCC CGCCCTCGAGGGTG		
	ORF Start: at 1		ORF Stop: at 910
	SEQ ID NO: 88	303 aa	MW at 32531.9kD
NOV11h, 277765267 Protein Sequence	AGSADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKGEAGVIGERGERGLPGAPGKAGPVGPK GDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFLSGWHTIYLPDCRPLTVLCDMDTDGGG WTVFQRRMDGSVDFYRDWAAYKQGFGSQLGEFWLGNDNIHALTAQGSSELRVDLVDFEGNHQFA KYKSFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFFSTKDQDNDVSSSNCAEKFQGAWWYAD CHASNLNGLYLMGPHESYANGINWSAAKGYKYSYKVSEMKVRPALEG		
	SEQ ID NO: 89	979 bp	
NOV11i, 277580705 DNA Sequence	GCCGGATCCACCATGGCCCGGGGGCTCGCTGTCATGCTAGTCTTGTTCCTGCATATCAAGAACC TGCCTGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGACAA GCTCACCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCAGGT GTCATTGGAGAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGACCAGTGGGGC CCAAAGGAGACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTC GTGTGCGACAGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTGG CACACCATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAG GGGGCTGGACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGCCGC ATACAAGCAGGGCTTCGGCAGTCAGCTGGGGGAGTTCTGGCTGGGGAATGACAACATCCACGCC TTGACTGCCCAGGGAAGCAGCGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACCAGT TTGCTAAGTACAAATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGC		

	CTTTGTCGGGGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAA GACCAAGACAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAAGGAGCCTGGTGGTACG CCGACTGTCATGCTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAA TGGTATCAACTGGAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTG CGGCCCGCCCTCGAGGGTG
	ORF Start: at 1 | | ORF Stop: at 979
	SEQ ID NO: 90 | 326 aa | MW at 35022.0kD
NOV11i, 277580705 Protein Sequence	AGSTMARGLAVMLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKGEAG VIGERGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFLSGW HTIYLPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQGFGSQLGEFWLGNDNIHA LTAQGSSELRVDLVDFEGNHQFAKYKSFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFFSTK DQDNDVSSSNCAEKFQGAWWYADCHASNLNGLYLMGPHESYANGINWSAAKGYKYSYKVSEMKV RPALEG
	SEQ ID NO: 91 | 970 bp |
NOV11j, 275623919 DNA Sequence	GCCGGATCCACCGGGCTCGCTGTCATGCTAGTCTTGTTCCTGCATATCAAGAACCTGCCTGCCC AGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGACAAGCTCACCAT TCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCAGGTGTCATTGGA GAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGACCAGTGGGGCCCAAAGGAG ACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTCGTGTGCGAC AGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTGGCACACCATC TACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAGGGGGCTGGA CCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGCCGCATACAAGCA GGGCTTCGGCAGTCAGCTGGGGGAGTTCTGGCTGGGGAATGACAACATCCACGCCTTGACTGCC CAGGGAAGCAGCGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACCAGTTTGCTAAGT ACAAATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCCTTTGTCGG GGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAAGACCAAGAC AATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAAGGAGCCTGGTGGTACGCCGACTGTC ATGCTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAATGGTATCAA CTGGAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGCGGCCCGCC CTCGAGGGTG
	ORF Start: at 1 | | ORF Stop: at 970
	SEQ ID NO: 92 | 323 aa | MW at 34663.5kD
NOV11j, 275623919 Protein Sequence	AGSTGLAVMLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKGEAGVIG ERGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFLSGWHTI YLPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQGFGSQLGEFWLGNDNIHALTA QGSSELRVDLVDFEGNHQFAKYKSFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFFSTKDQD NDVSSSNCAEKFQGAWWYADCHASNLNGLYLMGPHESYANGINWSAAKGYKYSYKVSEMKVRPA LEG
	SEQ ID NO: 93 | 991 bp |
NOV11k, 274056295 DNA Sequence	CACCAAGCTTCCCACCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAAG GAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTGAAGA TCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCACGCCAGAGA ATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAACCATAAAGACAAC TCAAGTATACGCAAGCTCCTTAGGAACCTGGACTTCTATGTCCTTCCAGTTCTTAACATAGATG GTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCCCGTTCACCCCATAATAATGG CACATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCATCTTGGTGTAGTATTGGTGCCTCT AGAAACTGCCAAGATCAAACATTCTGTGGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTG TTGCCAGCTTCATAGAGAGCAAGAAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGG GCAGTTAATTCTCACACCTTACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATT CAAGTTGGACAGAAGGCAGCAAATGCATTGAAAGCAAAGTATGGAACCAATTATAGAGTTGGAT CGAGTGCAGATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGAT TCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGGAACATATGGGTTTGTTCTGCCAGAAGCT CAGATCCAGCCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGTCCTGGATGA**TGT**GTATG CGAAACACTGGCACTCGGACAGTGCTGGAAGGGTGACATCTGCCACTATGCTGCTGGGCCTGCT GGTGTCCTGCATGTCTCTTCTCGTCGACGG**C**
	ORF Start: at 889 | | ORF Stop: at 991
	SEQ ID NO: 94 | 34 aa | MW at 3658.2kD
NOV11k, 274056295 Protein Sequence	CVCETLALGQCWKGDICHYAAGPAGVLHVSSRRR
	SEQ ID NO: 95 | 972 bp |

NOV11l, 268951998 DNA Sequence	**GGA**TCCACCATGGCCCGGGGGCTCGCTGTCCTGCTAGTCTTGTTCCTGCATATCAAGAACCTGCC TGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGACAAGCTCA CCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCAGGTGTCATT GGAGAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGACCAGTGGGGCCCAAAGG AGACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTCGTGTGCGA CAGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTGGCACACCATC TACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAGGGGGCTGGAC CGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGCCGCATACAAGCAGG GCTTCGGCAGTCAGCTGGGGGAGTTCTGGCTGGGGAACGACAACATCCACGCCCTGACTGCCCAG GGAAGCAGCGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACCAGTTTGCTAAGTACAA ATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCCTTTGTCGGGGGCA GTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAAGACCAAGACAATGAT GTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAGGGAGCCTGGTGGTACGCCGACTGTCATGCTTC AAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAATGGTATCAACTGGAGTG CGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGCGGCCCGCCCTCGAG
	ORF Start: at 1 | ORF Stop: end of sequence
	SEQ ID NO: 96 | 324 aa | MW at 34875.8kD
NOV11l, 268951998 Protein Sequence	GSTMARGLAVLLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKGEAGV IGERGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFLSGWH TIYLPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQGFGSQLGEFWLGNDNIHAL TAQGSSELRVDLVDFEGNHQFAKYKSFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFFSTKD QDNDVSSSNCAEKFQGAWWYADCHASNLNGLYLMGPHESYANGINWSAAKGYKYSYKVSEMKVR PALE
	SEQ ID NO: 97 | 1332 bp |
NOV11m, CG56653-02 DNA Sequence	TTTTAGGTCTGTTTGTCGTAGGCAGATGGAGCTTGTTATAATTATGCCTCATAGGGATAGTACA AGGAAGGGGTAGGCTATGTGTTTTGTCAGGGAGTTGAGAAACTGTGGCACAAGGCGAGAGCTGG TTTCCTCTGCCCTGTTAGAGCTGGGGGACTCTTCAGAGTCAAAGGCCAGAGAGC**ATG**GAGCTGA GTGGAGCCACCATGGCCCGGGGGCTCGCTGTCCTGCTAGTCTTGTTCCTGCATATCAAGAACCT GCCTGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGGCAAG CTCACCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCAGGTG TCATTGGAGAGAGAGGAGACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCA GTCTCAGTCGTGTGCGACAGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTG AGCGGCTGGCACACCATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACA CGGACGGAGGGGGCTGGACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGGA CTGGGCCGCATACAAGCAGGGCTTCGGCAGTCAGCTGGGGGAGTTCTGGCTGGGGAATGACAAC ATCCACGCCCTGACTGCCCAGGGAAGCAGCGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGCA ACCACCAGTTTGCTAAGTACAAATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGT ACTGGGAGCCTTTGTCGGGGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTC TCCACCAAAGACCAAGACAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAAGGAGCCT GGTGGTACGCCGACTGTCATGCTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAG CTATGCCAATGGTATCAACTGGAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAG ATGAAGGTGCGGCCCGCC**TAG**ACGGGCCAGGACCCCTCCACATGCACCTGCTAGTGGGGAGGCC ACACCCACAAGCGCTGCGTCGTGGAAGTCACCCCATTTCCCCAGCCAGACACACTCCCATGACG CCCACAGCTGCCCCTTTGCCCCCAGCTCAGTCAAGCCGCCACATGCCCACAACCTCACCAGAGG GAGAATTATGTTTCTAAATATGTTTACTTTTGGGGACAGAAAAAAAAAAAA
	ORF Start: ATG at 183 | | ORF Stop: TAG at 1107
	SEQ ID NO: 98 | 308 aa | MW at 33393.2kD
NOV11m, CG56653-02 Protein Sequence	MELSGATMARGLAVLLVLFLHIKNLPAQAADTCPEVKVVGLEGSGKLTILRGCPGLPGAPGPKG EAGVIGERGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFLSGWHTIYLPDCRPLTVLC DMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQGFGSQLGEFWLGNDNIHALTAQGSSELRVDLVD FEGNHQFAKYKSFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFFSTKDQDNDVSSSNCAEKF QGAWWYADCHASNLNGLYLMGPHESYANGINWSAAKGYKYSYKVSEMKVRPA
	SEQ ID NO: 99 | 728 bp |
NOV11n, CG56653-03 DNA Sequence	**CTG**CATATCAAGAACCTGCCTGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCC TGGAGGGCTCTGACAAGCTCACCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCC AAAGGGAGAGGCAGGTGTCATTGGAGAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAG GCAGGACCAGTGGGGCCCAAAGGAGACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACG CTGGGCAGTCTCAGTCGTGTGCGACAGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTA TTTCCTGAGCGGCTGGCACACCATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGAC ATGGACACGGACGGAGGGGGCTGGACCGTTTTCCAGGGAGCCTGGTGGTACGCCGACTGTCATG CTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAATGGTATCAACTG GAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGCGGCCCGCC**TAG** ACGGGCCAGGACCCCTCCACATGCACCTGCTAGTGGGGAGGCCACACCCACAAGCGCTGCGTCG

	TGGAAGTCACCCCATTTCCCCAGCCAGACACACTCCCATGACGCCCACAGCTGCCCCTTTGCCC CCAGCTCAGTCAAGCCGCCACATG		
	ORF Start: at 1		ORF Stop: TAG at 574
	SEQ ID NO: 100	191 aa	MW at 20220.7kD
NOV11n, CG56653-03 Protein Sequence	LHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKGEAGVIGERGERGLPGAPGK AGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFLSGWHTIYLPDCRPLTVLCD MDTDGGGWTVFQGAWWYADCHASNLNGLYLMGPHESYANGINWSAAKGYKYSYKVSEMKVRPA		
	SEQ ID NO: 101	1104 bp	
NOV11o, CG56653-04 DNA Sequence	CTGAGTGGAGCCACC**ATG**GCCCGGGGGCTCGCTGTCCTGCTAGTCTTGTTCCTGCATATCACGA ACCTGCCTGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGA CAAGCTCACCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGACCAAAGGGAGAGGCA GGTGTCATTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTCGTGTGCGACAGGCCCACGCAACT GCAAGGACCTGCTAGACCGAGGGTATTTCCTGAGCGGCTGGCACACCATCTACCTGCCCGACTG CCGGCCCCTTACTGTGCTCTGTGACATGGATACGGACGGAGGGGGCTGGACCGTTTTCCAGCGG AGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGCCGCATACAAGCAGGGCTTCGGCAGTC AGCTGGGGGAGTTCTGGCTGGGGAATGACAACATCCACGCCCTGACTGCCCAGGGAAGCAGCGA GCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACCAGTTTGCTAAGTACAAATCATTCAAG GTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCCTTTGTCGGGGGCAGTGCGGGTA ATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAAGACCAAGACAATGATGTGAGTTC TTCGAATTGTGCTGAGAAGTTCCAGGGAGCCTGGTGGTACGCCGACTGTCATGCTTCAAACCTC AATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAATGGTATCAACTGGAGTGCGGCGA AGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGCGGCCCGCC**TAG**ACGGGCCAGGA CCCCTCCACATGCACCTGCTAGTGGGGAGGCCACACCCACAAGCGCTGCGTCGTGGAAGTCACC CCATTTCCCCAGCCAGACACACTCCCATGACGCCCACAGCTGCCCCTTTGCCCCCAGCTCAGTC AAGCCGCCACATGCCCACAACCTCACCAGAGGGAGAATTATGTTTCTAAATATGTTTACTTTGG GACAGAAAAAAAAAAA		
	ORF Start: ATG at 16		ORF Stop: TAG at 883
	SEQ ID NO: 102	289 aa	MW at 31404.9kD
NOV11o, CG56653-04 Protein Sequence	MARGLAVLLVLFLHITNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKGEAGVIGE KGDAGQSQSCATGPRNCKDLLDRGYFLSGWHTIYLPDCRPLTVLCDMDTDGGGWTVFQRRMDGS VDFYRDWAAYKQGFGSQLGEFWLGNDNIHALTAQGSSELRVDLVDFEGNHQFAKYKSFKVADEA EKYKLVLGAFVGGSAGNSLTGHNNNFFSTKDQDNDVSSSNCAEKFQGAWWYADCHASNLNGLYL MGPHESYANGINWSAAKGYKYSYKVSEMKVRPA		
	SEQ ID NO: 103	1000 bp	
NOV11p, CG56653-08 DNA Sequence	C**ATG**GAGCTGAGTGGAGCCACCATGGCCCGGGGGCTCGCTGTCCTGCTAGTCTTGTTCCTGCAT ATCAAGAACCTGCCTGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGG GCTCTGACAAGCTCACCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGG AGAGGCAGGTGTCATTGGAGAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGA CCAGTGGGGCCCAAGGGAGACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGC AGTCTCAGTCGTGTGCGACAGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCT GAGCGGCTGGCACAATATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGAC ACGGACGGAGGGGGCTGGACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGG ACTGGGCCGCATACAAGCAGGGCTTCGGCAGTCAGCTGGGGGAGTTCTGGCTGGGGAACGACAA CATCCACGCCCTGACTGCCCAGGGAAGCAGCGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGC AACCACCAGTTTGCTAAGTACAAATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGG TACTGGGAGCCTTTGTCGGGGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTT CTCCACCAAAGACCAAGACAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAGGGAGCC TGGTGGTACGCCGACTGTCATGCTTCGAGCCTCAATGGTCTCTACCTCATGGGACCCCATGAGA GCTATGCCAATGGTATCAACTGGAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGA GATGAAGGTGCGGCCCGCC**TAG**ACGGGCCAGGACCCCTAA		
	ORF Start: ATG at 2		ORF Stop: TAG at 980
	SEQ ID NO: 104	326 aa	MW at 35064.1kD
NOV11p, CG56653-08 Protein Sequence	MELSGATMARGLAVLLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKG EAGVIGERGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFL SGWHNIYLPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQGFGSQLGEFWLGNDN IHALTAQGSSELRVDLVDFEGNHQFAKYKSFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFF STKDQDNDVSSSNCAEKFQGAWWYADCHASSLNGLYLMGPHESYANGINWSAAKGYKYSYKVSE MKVRPA		
	SEQ ID NO: 105	1144 bp	
NOV11q,	TTTTAGGTCTGTTTGTCGTAGGCAGATGGAGCTTGTTATAATTATGCCTCATAGGGATAGTACA		

CG56653-09 DNA Sequence	AGGAAGGGGTAGGCTATGTGTTTTGTCAGGGAGTTGAGAAACTGTGGCACAAGGCGAGAGCTGG TTTCCTCTGCCCTGTTAGAGCTGGGGGACTCTTCAGAGTCAAAGGCCAGAGAGC**ATG**GAGCTGA GTGGAGCCACCATGGCCCGGGGGCTCGCTGTCCTGCTAGTCTTGTTCCTGCATATCAAGAACCT GCCTGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGACAAG CTCACCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCAGGTG TCATTGGAGAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGACCAGTGGGGCC CAAAGGAGACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTCG TGTGCGACAGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTGGC ACACCATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAGG GGGCTGGACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGCCGCA TACAAGCAGGGCTTCGGCAGTCAGCTGGGGGGTAATTCTCTAACGGGCCACAACAACAACTTCT TCTCCACCAAAGACCAAGACAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAAGGAGC CTGGTGGTACGCCGACTGTCATGCTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAG AGCTATGCCAATGGTATCAACTGGAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAG AGATGAAGGTGCGGCCCGCC**TAG**ACGGGCCAGGACCCCTCCACATGCACCTGCTAGTGGGGAGG CCACACCCACAAGCGCTGCGTCGTGGAAGTCACCCCATTTCCCCAGCCAGACACACTCCCATGA CGCCCACAGCTGCCCCTTTGCCCCCAGCTCAGTCAAGCCGCCACATGCCCACAACC		
	ORF Start: ATG at 183		ORF Stop: TAG at 981
	SEQ ID NO: 106	266 aa	MW at 28476.8kD
NOV11q, CG56653-09 Protein Sequence	MELSGATMARGLAVLLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKG EAGVIGERGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFL SGWHTIYLPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQGFGSQLGGNSLTGHN NNFFSTKDQDNDVSSSNCAEKFQGAWWYADCHASNLNGLYLMGPHESYANGINWSAAKGYKYSY KVSEMKVRPA		
	SEQ ID NO: 107	651 bp	
NOV11r, CG56653-10 DNA Sequence	GGATCC**GGC**CCACGCAACTGCAAGGACCTGCTAGACCGAGGGCACTTCCTGAGCGGCTGGCACA CCATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAGGGGG CTGGACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGCCGCATAC AAGCAGGGCTTCGGCAGTCAGCTGGGGGAGTTCTGGCTGGGGAACGACAACATCCACGCCCTGA CTGCCCAGGGAAGCAGCGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACCAGTTTGC TAAGTACAAATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCCTTT GTCGGGGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAAGACC AAGACAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAAGGAGCCTGGTGGTACGCCGA CTGTCATGCTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAATGGT ATCAACTGGAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGCGGC CCGCC**CTC**GAG		
	ORF Start: at 7		ORF Stop: at 646
	SEQ ID NO: 108	213 aa	MW at 23830.2kD
NOV11r, CG56653-10 Protein Sequence	GPRNCKDLLDRGHFLSGWHTIYLPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQ GFGSQLGEFWLGNDNIHALTAQGSSELRVDLVDFEGNHQFAKYKSFKVADEAEKYKLVLGAFVG GSAGNSLTGHNNNFFSTKDQDNDVSSSNCAEKFQGAWWYADCHASNLNGLYLMGPHESYANGIN WSAAKGYKYSYKVSEMKVRPA		
	SEQ ID NO: 109	972 bp	
NOV11s, CG56653-12 DNA Sequence	GGATCCACC**ATG**GCCCGGGGGCTCGCTGTCCTGCTAGTCTTGTTCCTGCATATCAAGAACCTGC CTGCCCAGGCTGCGGACACATGTCCAGAGGTGAAGGTGGTGGGCCTGGAGGGCTCTGACAAGCT CACCATTCTCCGAGGCTGCCCGGGGCTGCCCGGGGCCCCAGGGCCAAAGGGAGAGGCAGGTGTC ATTGGAGAGAGAGGAGAACGCGGTCTCCCTGGAGCCCCTGGAAAGGCAGGACCAGTGGGGCCCA AAGGAGACCGAGGAGAGAAGGGGATGCGTGGAGAGAAAGGAGACGCTGGGCAGTCTCAGTCGTG TGCGACAGGCCCACGCAACTGCAAGGACCTGCTAGACCGGGGGTATTTCCTGAGCGGCTGGCAC ACCATCTACCTGCCCGACTGCCGGCCCCTGACTGTGCTCTGTGACATGGACACGGACGGAGGGG GCTGGACCGTTTTCCAGCGGAGGATGGATGGCTCTGTGGACTTCTATCGGGACTGGGCCGCATA CAAGCAGGGCTTCGGCAGTCAGCTGGGGGAGTTCTGGCTGGGGAACGACAACATCCACGCCCTG ACTGCCCAGGGAAGCAGCGAGCTCCGTGTAGACCTGGTGGACTTTGAGGGCAACCACCAGTTTG CTAAGTACAAATCATTCAAGGTGGCTGACGAGGCAGAGAAGTACAAGCTGGTACTGGGAGCCTT TGTCGGGGGCAGTGCGGGTAATTCTCTAACGGGCCACAACAACAACTTCTTCTCCACCAAAGAC CAAGACAATGATGTGAGTTCTTCGAATTGTGCTGAGAAGTTCCAGGGAGCCTGGTGGTACGCCG ACTGTCATGCTTCAAACCTCAATGGTCTCTACCTCATGGGACCCCATGAGAGCTATGCCAATGG TATCAACTGGAGTGCGGCGAAGGGGTACAAATATAGCTACAAGGTGTCAGAGATGAAGGTGCGG CCCGCC**CTC**GAG		
	ORF Start: ATG at 10		ORF Stop: at 967
	SEQ ID NO: 110	319 aa	MW at 34388.3kD

NOV11s, CG56653-12 Protein Sequence	MARGLAVLLVLFLHIKNLPAQAADTCPEVKVVGLEGSDKLTILRGCPGLPGAPGPKGEAGVIGE RGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNCKDLLDRGYFLSGWHTIY LPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQGFGSQLGEFWLGNDNIHALTAQ GSSELRVDLVDFEGNHQFAKYKSFKVADEAEKYKLVLGAFVGGSAGNSLTGHNNNFFSTKDQDN DVSSSNCAEKFQGAWWYADCHASNLNGLYLMGPHESYANGINWSAAKGYKYSYKVSEMKVRPA

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 11B.

Table 11B. Comparison of NOV11a against NOV11b through NOV11s.

Protein Sequence	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV11b	8..326 1..319	319/319 (100%) 319/319 (100%)
NOV11c	30..322 1..293	293/293 (100%) 293/293 (100%)
NOV11d	11..326 1..316	315/316 (99%) 316/316 (99%)
NOV11e	1..326 1..283	282/326 (86%) 282/326 (86%)
NOV11f	7..326 9..285	276/320 (86%) 276/320 (86%)
NOV11g	113..326 16..186	170/214 (79%) 170/214 (79%)
NOV11h	27..326 1..300	298/300 (99%) 299/300 (99%)
NOV11i	4..326 1..323	320/323 (99%) 323/323 (99%)
NOV11j	11..326 5..320	315/316 (99%) 316/316 (99%)
NOV11k	183..190 8..15	4/8 (50%) 6/8 (75%)
NOV11l	5..326 1..322	321/322 (99%) 322/322 (99%)
NOV11m	1..326 1..308	307/326 (94%) 307/326 (94%)
NOV11n	20..159 1..140	140/140 (100%) 140/140 (100%)
NOV11o	8..326 1..289	288/319 (90%) 288/319 (90%)
NOV11p	1..326 1..326	324/326 (99%) 325/326 (99%)

NOV11q	1..326 1..266	245/326 (75%) 252/326 (77%)
NOV11r	114..326 1..213	212/213 (99%) 213/213 (99%)
NOV11s	8..326 1..319	319/319 (100%) 319/319 (100%)

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 30 and 31

PSORT II analysis:	PSG: a new signal peptide prediction method
	N-region: length 10; pos.chg 1; neg.chg 1
	H-region: length 12; peak value 11.35
	PSG score: 6.95
	
	GvH: von Heijne's method for signal seq. recognition
	GvH score (threshold: -2.1): -3.79
	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: 1
	Number of TMS(s) for threshold 0.5: 1
	INTEGRAL Likelihood = -3.35 Transmembrane 6 - 22
	PERIPHERAL Likelihood = 4.51 (at 234)
	ALOM score: -3.35 (number of TMSs: 1)
	
	MTOP: Prediction of membrane topology (Hartmann et al.)
	Center position for calculation: 13
	Charge difference: -1.5 C(-0.5) - N(1.0)
	N >= C: N-terminal side will be inside
	
	>>> membrane topology: type 2 (cytoplasmic tail 1 to 6)
	
	MITDISC: discrimination of mitochondrial targeting seq
	R content: 1 Hyd Moment(75): 3.15
	Hyd Moment(95): 5.83 G content: 2
	D/E content: 2 S/T content: 2
	Score: -6.93
	
	Gavel: prediction of cleavage sites for mitochondrial preseq
	R-2 motif at 20 ARG|LA
	
	NUCDISC: discrimination of nuclear localization signals
	pat4: none
	pat7: none
	bipartite: none
	content of basic residues: 10.7%
	NLS Score: -0.47
	
	KDEL: ER retention motif in the C-terminus: none
	
	ER Membrane Retention Signals:
	KKXX-like motif in the C-terminus: KVRP

	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): 34.8 %: mitochondrial 30.4 %: cytoplasmic 8.7 %: Golgi 8.7 %: endoplasmic reticulum 4.3 %: vacuolar 4.3 %: extracellular, including cell wall 4.3 %: nuclear 4.3 %: vesicles of secretory system >> prediction for CG56653-11 is mit (k=23)

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**

AAR94183	Human 35 kDa opsonin protein P35 - Homo sapiens, 313 aa. [JP08038182-A, 13-FEB-1996]	1..326 1..313	250/326 (76%) 267/326 (81%)	e-147
AAR30971	TGF-beta-1 binding protein - Sus scrofa, 324 aa. [WO9222319-A, 23-DEC-1992]	1..324 2..322	241/324 (74%) 270/324 (82%)	e-146
AAR94179	Human 35 kDa opsonin protein P35 fragment (III) residues 26-313 - Homo sapiens, 288 aa. [JP08038182-A, 13-FEB-1996]	28..326 2..288	237/299 (79%) 253/299 (84%)	e-142
AAR94178	Human 35 kDa opsonin protein P35 fragment (II) residues 96-313 - Homo sapiens, 218 aa. [JP08038182-A, 13-FEB-1996]	109..326 1..218	182/218 (83%) 195/218 (88%)	e-111
AAB29658	Human membrane-associated protein HUMAP-15 - Homo sapiens, 299 aa. [WO200065054-A2, 02-NOV-2000]	25..324 21..299	144/300 (48%) 188/300 (62%)	6e-78

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
O00602	Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1) (Ficolin-A) (Ficolin A) (M-Ficolin) - Homo sapiens (Human), 326 aa.	1..326 1..326	326/326 (100%) 326/326 (100%)	0.0
S61517	ficolin-1 precursor - human, 326 aa.	1..326 1..326	324/326 (99%) 325/326 (99%)	0.0
Q29042	Ficolin - Sus scrofa (Pig), 326 aa.	1..324 1..324	253/324 (78%) 279/324 (86%)	e-154
P57756	Ficolin 2 precursor (Collagen/fibrinogen domain-containing protein 2) (Ficolin-B) (Ficolin B) (Serum lectin P35) (EBP-37) (Hucolin) - Rattus norvegicus (Rat), 319 aa.	8..324 1..317	244/317 (76%) 271/317 (84%)	e-151
Q29041	Ficolin - Sus scrofa (Pig), 323 aa.	1..324 1..321	242/324 (74%) 270/324 (82%)	e-146

PFam analysis predicts 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**
Collagen	50..108	27/60 (45%) 46/60 (77%)	3.1e-09
fibrinogen_C	114..326	118/272 (43%) 161/272 (59%)	1.9e-94

Example 12. NOV 12: CG59713 Van Gogh.

The NOV12 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 12A.

Table 12A. NOV12 Sequence Analysis			
	SEQ ID NO: 111	1589 bp	
NOV12a, CG59713-03 DNA Sequence	CATTGCT**ATG**GATACCGAATCCACTTATTCTGGATATTCTTACTATTCAAGTCATTCGAAAAAA TCTCACAGACAAGGGGAAAGAACTAGAGAGAGACACAAGTCACCCCGGAATAAAGACGGCAGAG GGTCAGAAAAGTCTGTCACCATTCAACCTCCCACTGGAGAGCCCCTGTTGGGAAATGATTCTAC TCGGACAGAGGAAGTTCAGGATGACAACTGGGGAGAGACCACCACGGCCATCACAGGCACCTCG GAGCACAGCATATCCCAAGAGGACATTGCCAGGATCAGCAAGGACATGGAGGACAGCGTGGGGC TGGATTGCAAACGCTACCTGGGCCTCACCGTCGCCTCTTTTCTTGGACTTCTAGTTTTCCTCAC CCCTATTGCCTTCATCCTTTTACCTCCGATCCTGTGGAGGGATGAGCTGGAGCCTTGTGGCACA ATTTGTGAGGGGCTCTTTATCTCCATGGCATTCAAACTCCTCATTCTGCTCATAGGGACCTGGG CACTTTTTTTCCGCAAGCGGAGAGCTGACATGCCACGGGTGTTTGTGTTTCGTGCCCTTTTGTT GGTCCTCATCTTTCTCTTTGTGGTTTCCTATTGGCTTTTTTACGGGGTCCGCATTTTGGACTCT CGGGACCGGAATTACCAGGGCATTGTGCAATATGCAGTCTCCCTTGTGGATGCCCTCCTCTTCA TCCATTACCTGGCCATCGTCCTGCTGGAGCTCAGGCAGCTGCAGCCCATGTTCACGCTGCAGGT GGTCCGCTCCACCGATGGCGAGTCCCGCTTCTACAGCCTGGGACACCTGAGTATCCAGCGAGCA GCATTGGTGGTCCTAGAAAATTACTACAAAGATTTCACCATCTATAACCCAAACCTCCTAACAG CCTCCAAATTCCGAGCAGCCAAGCATATGGCCGGGCTGAAAGTCTACAATGTAGATGGCCCCAG TAACAATGCCACTGGCCAGTCCCGGGCCATGATTGCTGCAGCTGCTCGGCGCAGGGACTCAAGC CACAACGAGTTGTATTATGAAGAGGCCGAACATGAACGGCGAGTAAAGAAGCGGAAAGCAAGGC TGGTGGTTGCAGTGGAAGAGGCCTTCATCCACATTCAGCGTCTCCAGGCTGAGGAGCAGCAGAA AGCCCCAGGGGAGGTGATGGACCCTAGGGAGGCCGCCCAGGCCATTTTCCCCTCCATGGCCAGG GCTCTCCAGAAGTACCTGCGCATCACCCGGCAGCAGAACTACCACAGCATGGAGAGCATCCTGC AGCACCTGGCCTTCTGCATCACCAACGGCATGACCCCCAAGGCCTTCCTAGAACGGTACCTCAG TGCGGGCCCCACCCTGCAATATGACAAGGACCGCTGGCTCTCTACACAGTGGAGGCTTGTCAGT GATGAGGCTGTGACTAATGGATTACGGGATGGAATTGTGTTCGTCCTTAAGTGCTTGGACTTCA GCCTCGTAGTCAATGTGAAGAAAATTCCATTCATCATACTCTCTGAAGAGTTCATAGACCCCAA ATCTCACAAATTTGTCCTTCGCTTACAGTCTGAGACATCCGTTT**TAA**AAGTTCT		
	ORF Start: ATG at 8		ORF Stop: TAA at 1580
	SEQ ID NO: 112	524 aa	MW at 59974.2kD
NOV12a, CG59713-03 Protein Sequence	MDTESTYSGYSYYSSHSKKSHRQGERTRERHKSPRNKDGRGSEKSVTIQPPTGEPLLGNDSTRT EEVQDDNWGETTTAITGTSEHSISQEDIARISKDMEDSVGLDCKRYLGLTVASFLGLLVFLTPI AFILLPPILWRDELEPCGTICEGLFISMAFKLLILLIGTWALFFRKRRADMPRVFVFRALLLVL IFLFVVSYWLFYGVRILDSRDRNYQGIVQYAVSLVDALLFIHYLAIVLLELRQLQPMFTLQVVR STDGESRFYSLGHLSIQRAALVVLENYYKDFTIYNPNLLTASKFRAAKHMAGLKVYNVDGPSNN ATGQSRAMIAAAARRRDSSHNELYYEEAEHERRVKKRKARLVVAVEEAFIHIQRLQAEEQQKAP GEVMDPREAAQAIFPSMARALQKYLRITRQQNYHSMESILQHLAFCITNGMTPKAFLERYLSAG PTLQYDKDRWLSTQWRLVSDEAVTNGLRDGIVFVLKCLDFSLVVNVKKIPFIILSEEFIDPKSH KFVLRLQSETSV		
	SEQ ID NO: 113	1603 bp	

NOV12b, CG59713-01 DNA Sequence	CCCTCCATTGCT**ATG**GATACCGAATCCACTTATTCTGGATATTCTTACTATTCAAGTCATTCGA AAAAATCTCACAGACAAAGGGAAAGAACTAGAGAGAGACACAAGTCACCCCGGAATAAAGACGG CAGAGGGTCAGAAAAGTCTGTCACCATTCAACCTCCCACTGGAGAGCCCCTGTTGGGAAATGAT TCTACTCGGACAGAGGAACAGGATGACAACTGGGGAGAGACCACCACGGCCATCACAGGCACCT CGGAGCACAGCATATCCCAAGAGGACATTGCCAGGATCAGCAAGGACATGGAGGACAGCGTGGG GCTGGATTGCAAACGCTACCTGGGCCTCACCGTCGCCTCTTTTCTTGGACTTCTAGTTTTCCTC ACCCCTATTGCCTTCATCCTTTTACCTCCGATCCTGTGGAGGGATGAGCTGGAGCCTTGTGGCA CAATTTGTGAGGGGCTCTTTATCTCCATGGCATTCAAACTCCTCATTCTGCTCATAGGGACCTG GGCACTTTTTTTCCGCAAGCGGAGAGCTGACATGCCACGGGTGTTTGTGTTTCGTGCCCTTTTG TTGGTCCTCATCTTTCTCTTTGTGGTTTCCTATTGGCTTTTTTTACGGGGTCCGCATTTTGGACT CTCGGGACCGGAATTACCAGGGCATTGTGCAATATGCAGTCTCCCTTGTGGATGCCCTCCTCTT CATCCATTACCTGGCCATCGTCCTGCTGGAGCTCAGGCAGCTGCAGCCCATGTTCACGCTGCAG GTGGTCCGCTCCACCGATGGCGAGTCCCGCTTCTACAGCCTGGGACACCTGAGTATCCAGCGAG CAGCATTGGTGGTCCTAGAAAATTACTACAAAGATTTCACCATCTATAACCCAAACCTCCTAAC AGCCTCCAAATTCCGAGCAGCCAAGCATATGGCCGGGCTGAAAGTCTACAATGTATTCCCCAAA GGCCCCAGTAACAATGCCACTGGCCAGTCCCGGGCCATGATTGCTGCAGCTGCTCGGCGCAGGG ACTCAAGCCACAACGAGTTGTATTATGAAGAGGCCGAACATGAACGGCGAGTAAAGAAGCGGAA AGCAAGGCTGGTGGTTGCAGTGGAAGAGGCCTTCATCCACATTCAGCGTCTCCAGGCTGAGGAG CAGCAGAAAGCCCCAGGGGAGGTGATGGACCCTAGGGAGGCCGCCCAGGCCATTTTCCCCTCCA TGGCCAGGGCTCTCCAGAAGTACCTGCGCATCACCCGGCAGCAGAACTACCACAGCATGGAGAG CATCCTGCAGCACCTGGCCTTCTGCATCACCAACGGCATGACCCCCAAGGCCTTCCTAGAACGG TACCTCAGTGCGGGCCCCACCCTGCAATATGACAAGGACCGCTGGCTCTCTACACAGTGGAGGC TTGTCAGTGATGAGGCTGTGACTAATGGATTACGGGATGGAATTGTGTTCGTCCTTAAGTGCTT GGACTTCAGCCTCGTAGTCAATGTGAAGAAAATTCCATTCATCATACTCTCTGAAGAGTTCATA GACCCCAAATCTCACAAATTTGTCCTTCGCTTACAGTCTGAGACATCCGTT**TAA**AAGTTCTATA TTT
	ORF Start: ATG at 13 | | ORF Stop: TAA at 1588
	SEQ ID NO: 114 | 525 aa | MW at 60231.6kD
NOV12b, CG59713-01 Protein Sequence	MDTESTYSGYSYYSSHSKKSHRQRERTRERHKSPRNKDGRGSEKSVTIQPPTGEPLLGNDSTRT EEQDDNWGETTTAITGTSEHSISQEDIARISKDMEDSVGLDCKRYLGLTVASFLGLLVFLTPIA FILLPPILWRDELEPCGTICEGLFISMAFKLLILLIGTWALFFRKRRADMPRVFVFRALLLVLI FLFVVSYWLFYGVRILDSRDRNYQGIVQYAVSLVDALLFIHYLAIVLLELRQLQPMFTLQVVRS TDGESRFYSLGHLSIQRAALVVLENYYKDFTIYNPNLLTASKFRAAKHMAGLKVYNVFPKGPSN NATGQSRAMIAAAARRRDSSHNELYYEEAEHERRVKKRKARLVVAVEEAFIHIQRLQAEEQQKA PGEVMDPREAAQAIFPSMARALQKYLRITRQQNYHSMESILQHLAFCITNGMTPKAFLERYLSA GPTLQYDKDRWLSTQWRLVSDEAVTNGLRDGIVFVLKCLDFSLVVNVKKIPFIILSEEFIDPKS HKFVLRLQSETSV
	SEQ ID NO: 115 | 342 bp |
NOV12c, CG59713-02 DNA Sequence	GGATCC**ATG**GATACCGAATCCACTTATTCTGGATATTCTTACTATTCAAGTCATTCGAAAAAAT CTCACAGACAAGGGGAAAGAACTAGAGAGAGACACAAGTCACCCCGGAATAAAGACGGCAGAGG GTCAGAAAAGTCTGTCACCATTCAACCTCCCACTGGAGGGCCCCTGTTGGGAAATGATTCTACT CGGACAGAGGAAGTTCAGGATGACAACTGGGGAGAGACCACCACGGCCATCACAGGCACCTCGG AGCACAGCATATCCCAAGAGGACATTGCCAGGATCAGCAAGGACATGGAGGACAGCGTGGGGCT GGATTGCAAACGCTAC**CTC**GAG
	ORF Start: ATG at 7 | | ORF Stop: at 337
	SEQ ID NO: 116 | 110 aa | MW at 12329.1kD
NOV12c, CG59713-02 Protein Sequence	MDTESTYSGYSYYSSHSKKSHRQGERTRERHKSPRNKDGRGSEKSVTIQPPTGGPLLGNDSTRT EEVQDDNWGETTTAITGTSEHSISQEDIARISKDMEDSVGLDCKRY

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 12B.

Table 12B. Comparison of NOV12a against NOV12b and NOV12c.

Protein Sequence	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region

NOV12b	1..524 1..525	521/526 (99%) 521/526 (99%)
NOV12c	1..110 1..110	109/110 (99%) 109/110 (99%)

Further analysis of the NOV12a protein yielded the following properties shown in Table 12C.

Table 12C. Protein Sequence Properties NOV12a

SignalP analysis:	No Known Signal Sequence Predicted

PSORT II analysis:	PSG: a new signal peptide prediction method
	N-region: length 4; pos.chg 0; neg.chg 2
	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): -13.05
	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: 5
	INTEGRAL Likelihood = -8.39 Transmembrane 117 - 133
	INTEGRAL Likelihood = -4.83 Transmembrane 156 - 172
	INTEGRAL Likelihood =-13.85 Transmembrane 182 - 198
	INTEGRAL Likelihood = -6.00 Transmembrane 224 - 240
	INTEGRAL Likelihood = -3.19 Transmembrane 478 - 494
	PERIPHERAL Likelihood = 3.66 (at 264)
	ALOM score: -13.85 (number of TMSs: 5)
	
	MTOP: Prediction of membrane topology (Hartmann et al.)
	Center position for calculation: 124
	Charge difference: -2.0 C(-3.0) - 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): 3.29
	Hyd Moment(95): 0.89 G content: 0
	D/E content: 2 S/T content: 1
	Score: -7.75
	
	Gavel: prediction of cleavage sites for mitochondrial preseq
	cleavage site motif not found
	
	NUCDISC: discrimination of nuclear localization signals
	pat4: RKRR (5) at 173
	pat4: KKRK (5) at 355
	pat7: none
	bipartite: none
	content of basic residues: 12.6%
	NLS Score: 0.15
	
	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 ***
         LLLVLIFLFVVSYWLFYGVRIL at 188
 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):

        44.4 %: endoplasmic reticulum
        33.3 %: mitochondrial
        11.1 %: nuclear
        11.1 %: vesicles of secretory system

>> prediction for CG59713-03 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

ABP70133	Human NOV40a - Homo sapiens, 525 aa. [WO200272771-A2, 19-SEP-2002]	1..524 1..525	521/526 (99%) 521/526 (99%)	0.0
AAY73493	Human secreted protein clone yk224_1 protein sequence SEQ ID NO:208 - Homo sapiens, 426 aa. [WO9958642-A2, 18-NOV-1999]	99..524 1..426	426/426 (100%) 426/426 (100%)	0.0
AAG81268	Human AFP protein sequence SEQ ID NO:54 - Homo sapiens, 346 aa. [WO200129221-A2, 26-APR-2001]	179..524 1..346	344/346 (99%) 344/346 (99%)	0.0
ABB89231	Human polypeptide SEQ ID NO 1607 - Homo sapiens, 303 aa. [WO200190304-A2, 29-NOV-2001]	256..524 35..303	268/269 (99%) 269/269 (99%)	e-151
AAY86275	Human secreted protein HSRGW16, SEQ ID NO:190 - Homo sapiens, 295 aa. [WO9966041-A1, 23-DEC-1999]	270..524 41..295	255/255 (100%) 255/255 (100%)	e-143

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
Q8TAA9	Four-transmembrane protein associating with dishevelled (VANGL1) - Homo sapiens (Human), 524 aa.	1..524 1..524	524/524 (100%) 524/524 (100%)	0.0
Q8N559	Similar to vang-like 1 (van gogh, Drosophila) - Homo sapiens (Human), 503 aa (fragment).	25..524 4..503	500/500 (100%) 500/500 (100%)	0.0
Q91ZD4	Lpp1 (Ltap) - Mus musculus (Mouse), 521 aa.	1..524 1..521	381/526 (72%) 454/526 (85%)	0.0
Q9ULK5	Hypothetical protein KIAA1215 - Homo sapiens (Human), 575 aa (fragment).	1..524 55..575	383/526 (72%) 452/526 (85%)	0.0
Q90X64	Strabismus - Xenopus laevis (African clawed frog), 521 aa.	1..524 1..521	382/526 (72%) 447/526 (84%)	0.0

PFam analysis predicts 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**

Example B: Sequencing Methodology and Identification of NOVX Clones

1. **GeneCalling™ Technology:** This is a proprietary method of performing differential gene expression profiling between two or more samples developed at CuraGen and described by Shimkets, et al., "Gene expression analysis by transcript profiling coupled to a gene database query" Nature Biotechnology 17:198-803 (1999). cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then digested with up to as many as 120 pairs of restriction enzymes and pairs of linker-adaptors specific for each pair of restriction enzymes were ligated to the appropriate end. The restriction digestion generates a mixture of unique cDNA gene fragments. Limited PCR amplification is performed with primers homologous to the linker adapter sequence where one primer is biotinylated and the other is fluorescently labeled. The doubly labeled material is isolated and the fluorescently labeled single strand is resolved by capillary gel electrophoresis. A computer algorithm compares the electropherograms from an experimental and control group for each of the restriction digestions. This and additional sequence-derived information is used to predict the identity of each differentially expressed gene fragment using a variety of genetic databases. The identity of the gene fragment is confirmed by additional, gene-specific competitive PCR or by isolation and sequencing of the gene fragment.

2. **SeqCalling™ Technology:** cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example,

growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

3. PathCalling™ Technology: The NOVX nucleic acid sequences are derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, are sequenced. In silico prediction was based on sequences available in CuraGen Corporation's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods summarized below:

cDNA libraries were derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then directionally cloned into the appropriate two-hybrid vector (Gal4-activation domain (Gal4-AD) fusion). Such cDNA libraries as well as commercially available cDNA libraries from Clontech (Palo Alto, CA) were then transferred from E.coli into a CuraGen Corporation proprietary yeast strain (disclosed in U. S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corportion proprietary library of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR

product was sequenced; sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: the cDNA fragment derived by the screening procedure, covering the entire open reading frame is, as a recombinant DNA, cloned into pACT2 plasmid (Clontech) used to make the cDNA library. The recombinant plasmid is inserted into the host and selected by the yeast hybrid diploid generated during the screening procedure by the mating of both CuraGen Corporation proprietary yeast strains N106' and YULH (U. S. Patents 6,057,101 and 6,083,693).

4. **RACE:** Techniques based on the polymerase chain reaction such as rapid amplification of cDNA ends (RACE), were used to isolate or complete the predicted sequence of the cDNA of the invention. Usually multiple clones were sequenced from one or more human samples to derive the sequences for fragments. Various human tissue samples from different donors were used for the RACE reaction. The sequences derived from these procedures were included in the SeqCalling Assembly process described in preceding paragraphs.

5. **Exon Linking:** The NOVX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of

human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product derived from exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

6. **Physical Clone:** Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clones used for expression and screening purposes.

Example C. Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples

derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on metabolic diseases), AI_comprehensive_panel (containing normal tissue and samples from autoinflammatory diseases), Panel CNSD.01 (containing samples from normal and diseased brains) and CNS_neurodegeneration_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example, β-actin and GAPDH). Normalized RNA (5 ul) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10 μg of total RNA were performed in a volume of 20 μl and incubated for 60 minutes at 42°C. This reaction can be scaled up to 50 μg of total RNA in a final volume of 100 μl. sscDNA samples are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (Tm) range = 58°-60°C, primer optimal Tm = 59°C, maximum primer difference = 2°C, probe does not have 5'G, probe Tm must be 10°C greater than primer Tm, amplicon size 75bp to 100bp. The probes and

primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate (Applied Biosystems). PCR cocktails included either a single gene specific probe and primers set, or two multiplexed probe and primers sets (a set specific for the target clone and another gene-specific set multiplexed with the target probe). PCR reactions were set up using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at 48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were analyzed and processed as described previously.

Panels 1, 1.1, 1.2, and 1.3D

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through

the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var = small cell variant,

non-s = non-sm = non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

General_screening_panel_v1.4, v1.5, v1.6 and 1.7

The plates for Panels 1.4, 1.5, 1.6 and 1.7 include 2 control wells (genomic DNA control and chemistry control) and 88 to 94 wells containing cDNA from various samples. The samples in Panels 1.4, 1.5, 1.6 and 1.7 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in Panels 1.4, 1.5, 1.6 and 1.7 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on Panels 1.4, 1.5, 1.6 and 1.7 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal

skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose. Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

Panels 2D, 2.2, 2.3 and 2.4

The plates for Panels 2D, 2.2, 2.3 and 2.4 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI) or from Ardais or Clinomics). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI/ CHTN/Ardais/Clinomics). Unmatched RNA samples from tissues without malignancy (normal tissues) were also obtained from Ardais or Clinomics. This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen. These panels may alternatively be refered to as an Oncology_cell_line_screening_panel_v # Summary.

HASS Panel v 1.0

The HASS panel v 1.0 plates are comprised of 93 cDNA samples and two controls. Specifically, 81 of these samples are derived from cultured human cancer cell lines that had been subjected to serum starvation, acidosis and anoxia for different time periods as well as controls for these treatments, 3 samples of human primary cells, 9 samples of malignant brain cancer (4 medulloblastomas and 5 glioblastomas) and 2 controls. The human cancer

cell lines are obtained from ATCC (American Type Culture Collection) and fall into the following tissue groups: breast cancer, prostate cancer, bladder carcinomas, pancreatic cancers and CNS cancer cell lines. These cancer cells are all cultured under standard recommended conditions. The treatments used (serum starvation, acidosis and anoxia) have been previously published in the scientific literature. The primary human cells were obtained from Clonetics (Walkersville, MD) and were grown in the media and conditions recommended by Clonetics. The malignant brain cancer samples are obtained as part of a collaboration (Henry Ford Cancer Center) and are evaluated by a pathologist prior to CuraGen receiving the samples . RNA was prepared from these samples using the standard procedures. The genomic and chemistry control wells have been described previously.

ARDAIS Panel v 1.0

The plates for ARDAIS panel v 1.0 generally include 2 control wells and 22 test samples composed of RNA isolated from human tissue procured by surgeons working in close cooperation with Ardais Corporation. The tissues are derived from human lung malignancies (lung adenocarcinoma or lung squamous cell carcinoma) and in cases where indicated many malignant samples have "matched margins" obtained from noncancerous lung tissue just adjacent to the tumor. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue) in the results below. The tumor tissue and the "matched margins" are evaluated by independent pathologists (the surgical pathologists and again by a pathologist at Ardais). Unmatched malignant and non-malignant RNA samples from lungs were also obtained from Ardais. Additional information from Ardais provides a gross histopathological assessment of tumor differentiation grade and stage. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical state of the patient.

ARDAIS Prostate v 1.0

The plates for ARDAIS prostate 1.0 generally include 2 control wells and 68 test samples composed of RNA isolated from human tissue procured by surgeons working in close cooperation with Ardais Corporation. The tissues are derived from human prostate malignancies and in cases where indicated malignant samples have "matched margins" obtained from noncancerous prostate tissue just adjacent to the tumor. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue) in the results below. The tumor

tissue and the "matched margins" are evaluated by independent pathologists (the surgical pathologists and again by a pathologist at Ardais). RNA from unmatched malignant and non-malignant prostate samples were also obtained from Ardais. Additional information from Ardais provides a gross histopathological assessment of tumor differentiation grade and stage. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical state of the patient.

Panel 3D, 3.1 and 3.2

The plates of Panel 3D, 3.1, and 3.2 are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D, 3.1, 3.2, 1, 1.1., 1.2, 1.3D, 1.4, 1.5, and 1.6 are of the most common cell lines used in the scientific literature.

Panels 4D, 4R, and 4.1D

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells,

human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2μg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5μg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately $2x10^{6}$cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol ($5.5x10^{-5}$M) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM

Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10μg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and plated at 10^6cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5μg/ml anti-CD28 (Pharmingen) and 3ug/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resupended at 10^6cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5μg/ml or anti-CD40 (Pharmingen) at approximately 10μg/ml and IL-4 at 5-10ng/ml. Cells were harvested for RNA preparation at 24,48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10μg/ml anti-CD28 (Pharmingen) and 2μg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10^5-10^6cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1μg/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1μg/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1μg/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at $5x10^5$cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to $5x10^5$cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1μg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were

activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at -20°C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300µl of RNAse-free water and 35µl buffer (Promega) 5µl DTT, 7µl RNAsin and 8µl DNAse were added. The tube was incubated at 37°C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNAse free water. RNA was stored at -80°C.

AI_comprehensive panel_v1.0

The plates for AI_comprehensive panel_v1.0 include two control wells and 89 test samples comprised of cDNA isolated from surgical and postmortem human tissues obtained from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

Joint tissues including synovial fluid, synovium, bone and cartilage were obtained from patients undergoing total knee or hip replacement surgery at the Backus Hospital. Tissue samples were immediately snap frozen in liquid nitrogen to ensure that isolated RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues were supplied by Clinomics and were obtained during autopsy of trauma victims.

Surgical specimens of psoriatic tissues and adjacent matched tissues were provided as total RNA by Clinomics. Two male and two female patients were selected between the ages of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

Surgical specimens of diseased colon from patients with ulcerative colitis and Crohns disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three female and three male Crohn's patients between the ages of 41-69 were used.

Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female patients. Four of the patients were taking lebvid and two were on phenobarbital.

Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with alpha-1anti-trypsin deficiencies. Asthma patients ranged in age from 36-75, and excluded smokers to prevent those patients that could also have COPD. COPD patients ranged in age from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids, and bronchodilators.

In the labels employed to identify tissues in the AI_comprehensive panel_v1.0 panel, the following abbreviations are used:

AI = Autoimmunity

Syn = Synovial

Normal = No apparent disease

Rep22 /Rep20 = individual patients

RA = Rheumatoid arthritis

Backus = From Backus Hospital

OA = Osteoarthritis

(SS) (BA) (MF) = Individual patients

Adj = Adjacent tissue

Match control = adjacent tissues

-M = Male

-F = Female

COPD = Chronic obstructive pulmonary disease

AI.05 chondrosarcoma

The AI.05 chondrosarcoma plates are comprised of SW1353 cells that had been subjected to serum starvation and treatment with cytokines that are known to induce MMP (1, 3 and 13) synthesis (eg. IL1beta). These treatments include: IL-1beta (10 ng/ml), IL-1beta + TNF-alpha (50 ng/ml), IL-1beta + Oncostatin (50 ng/ml) and PMA (100 ng/ml). The SW1353 cells were obtained from the ATCC (American Type Culture Collection) and

were all cultured under standard recommended conditions. The SW1353 cells were plated at 3 x10^5 cells/ml (in DMEM medium-10 % FBS) in 6-well plates. The treatment was done in triplicate, for 6 and 18 h. The supernatants were collected for analysis of MMP 1, 3 and 13 production and for RNA extraction. RNA was prepared from these samples using the standard procedures.

Panels 5D and 5I

The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being repaired/closed, the obstetrician removed a small sample (less than 1 cc) of the exposed metabolic tissues during the closure of each surgical level. The biopsy material was rinsed in sterile saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus) and subcutaneous adipose. Patient descriptions are as follows:

Patient 2: Diabetic Hispanic, overweight, not on insulin

Patient 7-9: Nondiabetic Caucasian and obese (BMI>30)

Patient 10: Diabetic Hispanic, overweight, on insulin

Patient 11: Nondiabetic African American and overweight

Patient 12: Diabetic Hispanic on insulin

Adiocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable

for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose

Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated

Donor 2 and 3 AD: Adipose, Adipose Differentiated

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

GO Adipose = Greater Omentum Adipose

SK = Skeletal Muscle

UT = Uterus

PL = Placenta

AD = Adipose Differentiated

AM = Adipose Midway Differentiated

U = Undifferentiated Stem Cells

Human Metabolic RTQ-PCR Panel

The plates for the Human Metabolic RTQ-PCR Panel include two control wells (genomic DNA control and chemistry control) and 211 cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. This panel is useful for establishing the tissue and cellular expression profiles for genes believed to play a role in the etiology and pathogenesis of obesity and/or diabetes and to confirm differential expression of such genes derived from other methods. Metabolic tissues were obtained from patients enrolled in the CuraGen Gestational Diabetes study and from autopsy tissues from Type II diabetics

and age, sex and race-matched control patients. One or more of the following were used to characterize the patients: body mass index [BMI = wt (kg) / ht (m^2)], serum glucose, HgbA1c. Cell lines used in this panel are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines. RNA from human Pancreatic Islets was also obtained.

In the Gestational Diabetes study, subjects are young (18-40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarian section. After delivery of the infant, when the surgical incisions were being repaired/closed, the obstetrician removed a small sample (less than 1cc) of the exposed metabolic tissues during the closure of each surgical level. The biopsy material was rinsed in sterile saline, blotted, and then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus), and subcutaneous adipose. Patient descriptions are as follows:

Patient 7 - Non-diabetic Caucasian and obese
Patient 8 - Non-diabetic Caucasian and obese
Patient 12 - Diabetic Caucasian with unknown BMI and on insulin
Patient 13 - Diabetic Caucasian, overweight, not on insulin
Patient 15 - Diabetic Caucasian, obese, not on insulin
Patient 17 - Diabetic Caucasian, normal weight, not on insulin
Patient 18 - Diabetic Hispanic, obese, not on insulin
Patient 19 - Non-diabetic Caucasian and normal weight
Patient 20 - Diabetic Caucasian, overweight, and on insulin
Patient 21 - Non-diabetic Caucasian and overweight
Patient 22 - Diabetic Caucasian, normal weight, on insulin
Patient 23 - Non-diabetic Caucasian and overweight
Patient 25 - Diabetic Caucasian, normal weight, not on insulin
Patient 26 - Diabetic Caucasian, obese, on insulin
Patient 27 - Diabetic Caucasian, obese, on insulin

Total RNA was isolated from metabolic tissues of 12 Type II diabetic patients and 12 matched control patients included hypothalamus, liver, pancreas, small intestine, psoas muscle, diaphragm muscle, visceral adipose, and subcutaneous adipose. The diabetics and non-diabetics were matched for age, sex, ethnicity, and BMI where possible.

The panel also contains pancreatic islets from a 22 year old male patient (with a BMI of 35) obtained from the Diabetes Research Institute at the University of Miami School of Medicine. Islet tissue was processed to total RNA at CuraGen.

Cell lines used in this panel are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured at an outside facility. The RNA was extracted at CuraGen according to CuraGen protocols. All samples were then processed at CuraGen to produce single stranded cDNA.

In the labels used to identify tissues in the Human Metabolic panel, the following abbreviations are used:

Pl = placenta

Go = greater omentum

Sk = skeletal muscle

Ut = uterus

CC = Caucasian

HI = Hispanic

AA = African American

AS = Asian

Diab = Type II diabetic

Norm = Non-diabetic

Overwt = Overweight; med BMI

Obese = Hi BMI

Low BM = 20-25

Med BM = 26-30

Hi BMI = Greater than 30

M = Male

= Patient identifier

Vis. = Visceral

SubQ = Subcutaneous

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in

liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supernuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus palladus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

Glob Palladus= Globus palladus

Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

Panel CNS_Neurodegeneration_V1.0

The plates for Panel CNS_Neurodegeneration_V1.0 include two control wells and 47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains six brains from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no evidence of dementia prior to death. The eight normal control brains are divided

into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the following regions are represented: hippocampus, temporal cortex (Brodman Area 21), parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus; the parietal cortex shows moderate neuronal death in the late stages of the disease; the occipital cortex is spared in AD and therefore acts as a "control" region within AD patients. Not all brain regions are represented in all cases. These panels may alternatively be refered to as an CNS_1 Summary.

In the labels employed to identify tissues in the CNS_Neurodegeneration_V1.0 panel, the following abbreviations are used:

AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology upon autopsy

Control = Control brains; patient not demented, showing no neuropathology

Control (Path) = Control brains; pateint not demented but showing sever AD-like pathology

SupTemporal Ctx = Superior Temporal Cortex

Inf Temporal Ctx = Inferior Temporal Cortex

Panel CNS_Neurodegeneration_V2.0

The plates for Panel CNS_Neurodegeneration_V2.0 include two control wells and 47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains sixteen brains from Alzheimer's disease (AD) patients, and twenty-nine brains from "Normal controls" who showed no evidence of dementia prior to death. The twenty-nine normal control brains

are divided into two categories: Fourteen controls with no dementia and no Alzheimer's like pathology (Controls) and fifteen controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD senile plaque load). Tissue from the temporal cotex (Broddmann Area 21) was selected for all samples from the Harvard Brain Tissue Resource Center; from the two sample from the Human Brain and Spinal Fluid Resource Center (samples 1 and 2) tissue from the inferior and superior temporal cortex was used; each sample on the panel represents a pool of inferior and superior temporal cortex from an individual patient. The temporal cortex was chosen as it shows a loss of neurons in the intermediate stages of the disease. Selection of a region which is affected in the early stages of Alzheimer's disease (e.g., hippocampus or entorhinal cortex) could potentially result in the examination of gene expression after vulnerable neurons are lost, and missing genes involved in the actual neurodegeneration process.

In the labels employed to identify tissues in the CNS_Neurodegeneration_V2.0 panel, the following abbreviations are used:

AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology upon autopsy

Control = Control brains; patient not demented, showing no neuropathology

AH3 = Control brains; pateint not demented but showing sever AD-like pathology

Inf & Sup Temp Ctx Pool = Pool of inferior and superior temporal cortex for a given individual

A. NOV 1: CG110853, Testis Specific Protein TPX-1 like.

Expression of NOV 1b, gene CG110853-01 was assessed using the primer-probe set Ag6875, described in Table AA. Results of the RTQ-PCR runs are shown in Table AB. CG110853-01 represents a full length physical clone.

Table AA. Probe Name Ag6875

Primers	Sequnces	Length	Start Position	SEQ ID No
Forward	5'-ggcaaacatagtagtattttagactatcttg-3'	31	259	117
Probe	TET-5'-attgggactctttggtcctacaccat-3'-TAMRA	26	290	118
Reverse	5'-gtatgacgagatcctagattttgtct-3'	26	316	119

Table AB. General screening panel v1.6

Column A - Rel. Exp.(%) Ag6875, Run 278388050			
Tissue Name	**A**	**Tissue Name**	**A**
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.1
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	**100.0**	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.1	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.4	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.1	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.1
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.1	Thymus Pool	0.0
Trachea	2.9	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	2.9	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.1	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	0.1
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.3
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0

Fetal Liver	0.4	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.1
Renal ca. 786-0	0.0	Salivary Gland	2.1
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.2

General_screening_panel_v1.6 Summary: Ag6875 High expression of this gene is mainly seen in testis (CT=25.4). Therefore, expression of this gene may be used to distinguish testis from other samples used in panel. Furthermore, therapeutic modulation of this gene may be useful in the treatment of testis related diseases including fertility and hypogonadism.

Moderate to low expression of this gene is also seen in a ovarian cancer cell line, ovary, trachea, fetal lung, fetal liver, salivary gland, cerebral cortex, pituitary gland, salivary gland and pancreas. Interestingly, this gene is expressed at much higher levels in fetal (CTs=30-33) when compared to adult lung and liver (CT=40). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung and liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance lung and liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung and liver related diseases.

B. NOV 2: CG113367: Serotransferrin precursor.

Expression of NOV 2b, gene CG113367-03 was assessed using the primer-probe set Ag6392, described in Table BA. Results of the RTQ-PCR runs are shown in Tables BB, BC and BD.

Table BA. Probe Name Ag6392

Primers	Seuences	Length	Start Position	SEQ ID No
Forward	5'-gccagaccccgaatcac-3'	17	307	120
Probe	TET-5'-aagcttgcgtccacaagatattacgtcaac-3'-TAMRA	30	349	121
Reverse	5'-ccgagcagtcagttacgttg-3'	20	398	122

Table BB. CNS neurodegeneration v1.0

Column A - Rel. Exp.(% Ag6392, Run 269253953			
Tissue Name	**A**	**Tissue Name**	**A**
AD 1 Hippo	17.2	Control (Path) 3 Temporal Ctx	1.0
AD 2 Hippo	22.4	Control (Path) 4 Temporal Ctx	8.8
AD 3 Hippo	6.5	AD 1 Occipital Ctx	21.0
AD 4 Hippo	5.9	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	42.0	AD 3 Occipital Ctx	12.9
AD 6 Hippo	29.3	AD 4 Occipital Ctx	20.0
Control 2 Hippo	37.4	AD 5 Occipital Ctx	20.0
Control 4 Hippo	7.5	AD 6 Occipital Ctx	41.8
Control (Path) 3 Hippo	3.5	Control 1 Occipital Ctx	3.6
AD 1 Temporal Ctx	20.4	Control 2 Occipital Ctx	51.1
AD 2 Temporal Ctx	21.9	Control 3 Occipital Ctx	10.2
AD 3 Temporal Ctx	2.8	Control 4 Occipital Ctx	8.5
AD 4 Temporal Ctx	17.1	Control (Path) 1 Occipital Ctx	59.9
AD 5 Inf Temporal Ctx	**100.0**	Control (Path) 2 Occipital Ctx	11.8
AD 5 SupTemporal Ctx	48.3	Control (Path) 3 Occipital Ctx	3.5
AD 6 Inf Temporal Ctx	31.2	Control (Path) 4 Occipital Ctx	7.9
AD 6 Sup Temporal Ctx	34.6	Control 1 Parietal Ctx	9.0
Control 1 Temporal Ctx	3.1	Control 2 Parietal Ctx	32.3
Control 2 Temporal Ctx	28.1	Control 3 Parietal Ctx	19.2
Control 3 Temporal Ctx	6.6	Control (Path) 1 Parietal Ctx	23.2
Control 4 Temporal Ctx	5.8	Control (Path) 2 Parietal Ctx	20.2
Control (Path) 1 Temporal Ctx	19.3	Control (Path) 3 Parietal Ctx	1.4
Control (Path) 2 Temporal Ctx	10.4	Control (Path) 4 Parietal Ctx	16.5

Table BC. General screening panel v1.6

Column A - Rel. Exp.(%) Ag6392, Run 277247700			
Tissue Name	**A**	**Tissue Name**	**A**
Adipose	0.4	Renal ca. TK-10	48.6
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.6	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.4	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	4.3
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0

Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.1
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.1
Trachea	0.8	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.8	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.1	Brain (Amygdala) Pool	12.7
Lung ca. NCI-H526	0.0	Brain (cerebellum)	6.8
Lung ca. NCI-H23	0.0	Brain (fetal)	0.2
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	11.4
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	11.3
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	9.9
Liver	27.5	Brain (Thalamus) Pool	22.4
Fetal Liver	74.2	Brain (whole)	7.3
Liver ca. HepG2	**100.0**	Spinal Cord Pool	32.8
Kidney Pool	0.0	Adrenal Gland	0.1
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.2
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table BD. Panel 4.1D

Colun A - Rel. Exp.(%) Ag6392, Run 269097110			
Tissue Name	**A**	**Tissue Name**	**A**
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.1
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0

Secondary Th2 rest	0.1	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.1	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.1
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	**100.0**
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.1
LAK cells PMA/ionomycin	0.1	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.1	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.1
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.1
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.1
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.1
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.1
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.2	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.1
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.1
Macrophages rest	0.0	Lung	0.1
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	0.1
HUVEC starved	0.1		

CNS_neurodegeneration_v1.0 Summary: Ag6392 This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. This gene is found to be slightly upregulated in the temporal cortex of Alzheimer's disease patients. Blockade of this receptor may be of use in the treatment of this disease and to decrease neuronal death.

General_screening_panel_v1.6 Summary: Ag6392 Highest expression of this gene is detected in a liver cancer HepG2 cell line (CT=21.2). In addition, high expression of this gene is mainly seen in adult and fetal liver. Therefore, expression of this gene may be used to identify liver and liver related tissues. Furthermore, therapeutic modulation of this gene may be useful in the treatment of liver related diseases including liver cancer.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate to low levels in pancreas, adipose, adrenal gland, pituitary gland, fetal skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

This gene is expressed at much higher levels in fetal (CT=31.5) when compared to adult skeletal muscle (CT=40) therefore expression of this gene can be used to distinguish fetal from adult skeletal muscle. In addition, the relative overexpression of this gene in fetal skeletal muscle suggests that the protein product may enhance muscular growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of muscle related diseases. More specifically, treatment of weak or dystrophic muscle with the protein encoded by this gene could restore muscle mass or function.

High expression of this gene is also seen in a colon cancer CaCo-2 and a renal cancer TK-10 cell lines. Therefore, expression of this gene may be used as marker to detect the presence of these cancers and therapeutic modulation of this gene or its protein product may be useful in the treatement of colon and renal cancer.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag6392 High expression of this gene is seen exclusively in liver cirrhosis sample (CT=26.4). Therefore, gene, protein, antibodies or small molecule therapeutics targeting this gene or it's protein product could reduce or inhibit fibrosis that occurs in liver cirrhosis. In addition, gene or protein expression levels could be used for the diagnosis of liver cirrhosis.

C. NOV 3: CG132364: Programed cell death protein 1 precursor (Protein PD-1).

Expression of NOV 3b, gene CG132364-02 was assessed using the primer-probe set Ag6158 described in Table CA. Results of the RTQ-PCR runs are shown in Tables CB, CC, CD and CE.

Table CA. Probe Name Ag6158

Primers	Sequnces	Length	Start Position	SEQ ID No
Forward	5'-cattccgctaggaaagacaat-3'	21	620	123
Probe	TET-5'-catactccgtctgctcagggacacag-3'-TAMRA	26	589	124
Reverse	5'-atttccagtggcgagagaag-3'	20	549	125

Table CB. AI comprehensive panel v1.0

Column A - Rel. Exp.(%) Ag6158, Run 254398404			
Tissue Name	**A**	**Tissue Name**	**A**
110967 COPD-F	3.1	112427 Match Control Psoriasis-F	4.6
110980 COPD-F	0.0	112418 Psoriasis-M	2.2
110968 COPD-M	2.1	112723 Match Control Psoriasis-M	0.0
110977 COPD-M	3.0	112419 Psoriasis-M	0.0
110989 Emphysema-F	2.4	112424 Match Control Psoriasis-M	2.6
110992 Emphysema-F	0.6	112420 Psoriasis-M	3.0
110993 Emphysema-F	3.0	112425 Match Control Psoriasis-M	6.1
110994 Emphysema-F	0.0	104689 (MF) OA Bone-Backus	10.3
110995 Emphysema-F	5.5	104690 (MF) Adj "Normal" Bone-Backus	1.8
110996 Emphysema-F	3.5	104691 (MF) OA Synovium-Backus	2.9
110997 Asthma-M	2.1	104692 (BA) OA Cartilage-Backus	10.2
111001 Asthma-F	0.7	104694 (BA) OA Bone-Backus	14.0
111002 Asthma-F	2.9	104695 (BA) Adj "Normal" Bone-Backus	5.9
111003 Atopic Asthma-F	2.0	104696 (BA) OA Synovium-Backus	1.0
111004 Atopic Asthma-F	0.6	104700 (SS) OA Bone-Backus	4.5
111005 Atopic Asthma-F	1.8	104701 (SS) Adj "Normal" Bone-Backus	0.0
111006 Atopic Asthma-F	1.2	104702 (SS) OA Synovium-Backus	1.7
111417 Allergy-M	0.0	117093 OA Cartilage Rep7	0.0

112347 Allergy-M	0.2	112672 OA Bone5	0.4
112349 Normal Lung-F	0.0	112673 OA Synovium5	0.3
112357 Normal Lung-F	6.4	112674 OA Synovial Fluid cells5	0.0
112354 Normal Lung-M	4.6	117100 OA Cartilage Rep14	0.4
112374 Crohns-F	1.4	112756 OA Bone9	1.7
112389 Match Control Crohns-F	1.7	112757 OA Synovium9	0.8
112375 Crohns-F	1.5	112758 OA Synovial Fluid Cells9	0.0
112732 Match Control Crohns-F	71.2	117125 RA Cartilage Rep2	2.2
112725 Crohns-M	0.0	113492 Bone2 RA	4.9
112387 Match Control Crohns-M	0.0	113493 Synovium2 RA	0.5
112378 Crohns-M	0.3	113494 Syn Fluid Cells RA	3.5
112390 Match Control Crohns-M	0.9	113499 Cartilage4 RA	1.6
112726 Crohns-M	3.5	113500 Bone4 RA	4.7
112731 Match Control Crohns-M	4.0	113501 Synovium4 RA	1.4
112380 Ulcer Col-F	3.2	113502 Syn Fluid Cells4 RA	1.8
112734 Match Control Ulcer Col-F	**100.0**	113495 Cartilage3 RA	0.0
112384 Ulcer Col-F	4.9	113496 Bone3 RA	1.0
112737 Match Control Ulcer Col-F	0.0	113497 Synovium3 RA	2.3
112386 Ulcer Col-F	0.0	113498 Syn Fluid Cells3 RA	6.9
112738 Match Control Ulcer Col-F	11.3	117106 Normal Cartilage Rep20	1.5
112381 Ulcer Col-M	0.2	113663 Bone3 Normal	0.2
112735 Match Control Ulcer Col-M	1.3	113664 Synovium3 Normal	0.0
112382 Ulcer Col-M	1.0	113665 Syn Fluid Cells3 Normal	0.5
112394 Match Control Ulcer Col-M	0.9	117107 Normal Cartilage Rep22	0.9
112383 Ulcer Col-M	3.6	113667 Bone4 Normal	0.0
112736 Match Control Ulcer Col-M	0.9	113668 Synovium4 Normal	1.0
112423 Psoriasis-F	0.0	113669 Syn Fluid Cells4 Normal	0.6

Table CC. General screening panel v1.7

Column A - Rel. Exp.(%) Ag6158, Run 318350011			
Tissue Name	**A**	**Tissue Name**	**A**
Adipose	10.9	Gastric ca. (liver met.) NCI-N87	0.0
HUVEC	0.0	Stomach	0.5
Melanoma* Hs688(A).T	0.0	Colon ca. SW-948	0.0
Melanoma* Hs688(B).T	0.3	Colon ca. SW480	0.0
Melanoma (met) SK-MEL-5	0.0	Colon ca. (SW480 met) SW620	1.5
Testis	0.3	Colon ca. HT29	0.1
Prostate ca. (bone met) PC-3	0.0	Colon ca. HCT-116	1.1
Prostate ca. DU145	0.4	Colon cancer tissue	0.4
Prostate pool	0.9	Colon ca. SW1116	1.5
Uterus pool	0.3	Colon ca. Colo-205	0.0
Ovarian ca. OVCAR-3	1.0	Colon ca. SW-48	0.0
Ovarian ca. (ascites) SK-OV-3	0.0	Colon	3.3

Ovarian ca. OVCAR-4	0.0	Small Intestine	0.8
Ovarian ca. OVCAR-5	1.5	Fetal Heart	3.7
Ovarian ca. IGROV-1	0.2	Heart	0.4
Ovarian ca. OVCAR-8	1.8	Lymph Node pool 1	1.6
Ovary	2.2	Lymph Node pool 2	**100.0**
Breast ca. MCF-7	0.4	Fetal Skeletal Muscle	0.4
Breast ca. MDA-MB-231	0.3	Skeletal Muscle pool	0.2
Breast ca. BT-549	0.0	Skeletal Muscle	1.5
Breast ca. T47D	0.2	Spleen	16.6
Breast pool	1.4	Thymus	17.0
Trachea	12.6	CNS cancer (glio/astro) SF-268	0.0
Lung	15.7	CNS cancer (glio/astro) T98G	0.0
Fetal Lung	4.6	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.4
Lung ca. LX-1	1.4	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.7
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. NCI-H23	2.0	Brain (Amygdala)	0.0
Lung ca. NCI-H460	0.0	Brain (Cerebellum)	0.4
Lung ca. HOP-62	2.9	Brain (Fetal)	0.4
Lung ca. NCI-H522	0.6	Brain (Hippocampus)	0.0
Lung ca. DMS-114	2.1	Cerebral Cortex pool	0.6
Liver	0.3	Brain (Substantia nigra)	0.3
Fetal Liver	2.0	Brain (Thalamus)	1.1
Kidney pool	4.4	Brain (Whole)	1.6
Fetal Kidney	0.5	Spinal Cord	0.7
Renal ca. 786-0	0.0	Adrenal Gland	3.0
Renal ca. A498	0.3	Pituitary Gland	4.4
Renal ca. ACHN	0.0	Salivary Gland	3.6
Renal ca. UO-31	0.0	Thyroid	3.3
Renal ca. TK-10	1.8	Pancreatic ca. PANC-1	0.9
Bladder	3.7	Pancreas pool	0.6

Table CD. Oncology cell line screening panel v3.2

Column A - Rel. Exp.(%) Ag6158, Rn 311303313			
Tissue Name	**A**	**Tissue Name**	**A**
94905_Daoy_Medulloblastoma/Cerebellum_sscDNA	0.0	94954_Ca Ski_Cervical epidermoid carcinoma (metastasis)_sscDNA	0.0
94906_TE671_Medulloblastom/Cerebellum_sscDNA	0.0	94955_ES-2_Ovarian clear cell carcinoma_sscDNA	0.0
94907_D283 Med_Medulloblastoma/Cerebellum_sscDNA	0.0	94957_Ramos/6h stim_ Stimulated with PMA/ionomycin 6h_sscDNA	0.0
94908_PFSK-1_Primitive Neuroectodermal/Cerebellum_sscDNA	12.4	94958_Ramos/14h stim_ Stimulated with PMA/ionomycin 14h_sscDNA	8.5

94909_XF-498_CNS_sscDNA	0.0	94962_MEG-01_Chronic myelogenous leukemia (megokaryoblast)_sscDNA	0.0
94910_SNB-78_CNS/glioma_sscDNA	0.0	94963_Raji_Burkitt's lymphoma_sscDNA	0.0
94911_SF-268_CNS/glioblastoma_sscDNA	0.0	94964_Daudi_Burkitt's lymphoma_sscDNA	0.0
94912_T98G_Glioblastoma_sscDNA	0.0	94965_U266_B-cell plasmacytoma/myeloma_sscDNA	0.0
96776_SK-N-SH_Neuroblastoma (metastasis)_sscDNA	0.0	94968_CA46_Burkitt's lymphoma_sscDNA	20.6
94913_SF-295_CNS/glioblastoma_sscDNA	0.0	94970_RL_non-Hodgkin's B-cell lymphoma_sscDNA	0.0
132565_NT2 pool_sscDNA	41.5	94972_JM1_pre-B-cell lymphoma/leukemia_sscDNA	3.6
94914_Cerebellum_sscDNA	9.9	94973_Jurkat_T cell leukemia_sscDNA	30.1
96777_Cerebellum_sscDNA	6.8	94974_TF-1_Erythroleukemia_sscDNA	0.0
94916_NCI-H292_Mucoepidermoid lung carcinoma_sscDNA	0.0	94975_HUT 78_T-cell lymphoma_sscDNA	18.9
94917_DMS-114_Small cell lung cancer_sscDNA	15.5	94977_U937_Histiocytic lymphoma_sscDNA	0.0
94918_DMS-79_Small cell lung cancer/neuroendocrine_sscDNA	0.0	94980_KU-812_Myelogenous leukemia_sscDNA	0.0
94919_NCI-H146_Small cell lung cancer/neuroendocrine_sscDNA	0.0	94981_769-P_Clear cell renal carcinoma_sscDNA	0.0
94920_NCI-H526_Small cell lung cancer/neuroendocrine_sscDNA	15.3	94983_Caki-2_Clear cell renal carcinoma_sscDNA	0.0
94921_NCI-N417_Small cell lung cancer/neuroendocrine_sscDNA	0.0	94984_SW 839_Clear cell renal carcinoma_sscDNA	0.0
94923_NCI-H82_Small cell lung cancer/neuroendocrine_sscDNA	0.0	94986_G401_Wilms' tumor_sscDNA	0.0
94924_NCI-H157_Squamous cell lung cancer (metastasis)_sscDNA	0.0	126768_293 cells_sscDNA	2.5
94925_NCI-H1155_Large cell lung cancer/neuroendocrine_sscDNA	15.6	94987_Hs766T_Pancreatic carcinoma (LN metastasis)_sscDNA	0.0
94926_NCI-H1299_Large cell lung cancer/neuroendocrine_sscDNA	0.0	94988_CAPAN-1_Pancreatic adenocarcinoma (liver metastasis)_sscDNA	0.0
94927_NCI-H727_Lung carcinoid_sscDNA	0.0	94989_SU86.86_Pancreatic carcinoma (liver metastasis)_sscDNA	0.0
94928_NCI-UMC-11_Lung carcinoid_sscDNA	7.5	94990_BxPC-3_Pancreatic adenocarcinoma_sscDNA	0.0
94929_LX-1_Small cell lung cancer_sscDNA	19.6	94991_HPAC_Pancreatic adenocarcinoma_sscDNA	0.0
94930_Colo-205_Colon cancer_sscDNA	0.0	94992_MIA PaCa-2_Pancreatic carcinoma_sscDNA	9.2
94931_KM12_Colon cancer_sscDNA	0.0	94993_CFPAC-1_Pancreatic ductal adenocarcinoma_sscDNA	0.0

94932_KM20L2_Colon cancer_sscDNA	0.0	94994_PANC-1_Pancreatic epithelioid ductal carcinoma_sscDNA	18.2
94933_NCI-H716_Colon cancer_sscDNA	0.0	94996_T24_Bladder carcinma (transitional cell)_sscDNA	0.0
94935_SW-48_Colon adenocarcinoma_sscDNA	0.0	94997_5637_Bladder carcinoma_sscDNA	0.0
94936_SW1116_Colon adenocarcinoma_sscDNA	5.8	94998_HT-1197_Bladder carcinoma_sscDNA	0.0
94937_LS 174T_Colon adenocarcinoma_sscDNA	0.0	94999_UM-UC-3_Bladder carcinma (transitional cell)_sscDNA	0.0
94938_SW-948_Colon adenocarcinoma_sscDNA	0.0	95000_A204_Rhabdomyosarcoma_sscDNA	0.0
94939_SW-480_Colon adenocarcinoma_sscDNA	0.0	95001_HT-1080_Fibrosarcoma_sscDNA	0.0
94940_NCI-SNU-5_Gastric carcinoma_sscDNA	44.4	95002_MG-63_Osteosarcoma (bone)_sscDNA	0.0
112197_KATO III_Stomach_sscDNA	0.0	95003_SK-LMS-1_Leiomyosarcoma (vulva)_sscDNA	0.0
94943_NCI-SNU-16_Gastric carcinoma_sscDNA	0.0	95004_SJRH30_Rhabdomyosarcoma (met to bone marrow)_sscDNA	15.2
94944_NCI-SNU-1_Gastric carcinoma_sscDNA	0.0	95005_A431_Epidermoid carcinoma_sscDNA	0.0
94946_RF-1_Gastric adenocarcinoma_sscDNA	**100.0**	95007_WM266-4_Melanoma_sscDNA	0.0
94947_RF-48_Gastric adenocarcinoma_sscDNA	46.7	112195_DU 145_Prostate_sscDNA	0.0
96778_MKN-45_Gastric carcinoma_sscDNA	0.0	95012_MDA-MB-468_Breast adenocarcinoma_sscDNA	0.0
94949_NCI-N87_Gastric carcinoma_sscDNA	0.0	112196_SSC-4_Tongue_sscDNA	0.0
94951_OVCAR-5_Ovarian carcinoma_sscDNA	19.6	112194_SSC-9_Tongue_sscDNA	0.0
94952_RL95-2_Uterine carcinoma_sscDNA	0.0	112191_SSC-15_Tongue_sscDNA	11.7
94953_HelaS3_Cervical adenocarcinoma_sscDNA	0.0	95017_CAL 27_Squamous cell carcinoma of tongue_sscDNA	0.0

Table CE. Panel 4.1D

Colun A - Rel. Exp. (%) Ag6158, Run 254398392			
Tissue Name	**A**	**Tissue Name**	**A**
Secondary Th1 act	72.7	HUVEC IL-1beta	0.0
Secondary Th2 act	**100.0**	HUVEC IFN gamma	0.0
Secondary Tr1 act	10.7	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.5	HUVEC IL-11	0.0
Secondary Tr1 rest	0.3	Lung Microvascular EC none	0.0
Primary Th1 act	2.2	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	66.4	Microvascular Dermal EC none	0.0

Primary Tr1 act	80.7	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.1	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	1.4	Small airway epithelium none	0.0
Primary Tr1 rest	0.5	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	9.7	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	24.5	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	1.2	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	35.6	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	1.9	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	1.1
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.5	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	2.4	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	1.2	Liver cirrhosis	3.9
LAK cells IL-2+IL-12	2.5	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	3.5	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	1.5	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	33.7	NCI-H292 IL-13	0.2
NK Cells IL-2 rest	3.2	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	3.7	HPAEC none	0.0
Two Way MLR 5 day	4.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	9.3	Lung fibroblast none	0.0
PBMC rest	0.2	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	12.5	Lung fibroblast IL-4	0.0
PBMC PHA-L	8.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.6
B lymphocytes PWM	59.9	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	25.2	Dermal fibroblast CCD1070 TNF alpha	0.5
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	9.4	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	1.7	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.6
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	5.5	Colon	0.6
Macrophages rest	1.5	Lung	0.4
Macrophages LPS	3.3	Thymus	2.0
HUVEC none	0.0	Kidney	0.6
HUVEC starved	0.0		

AI_comprehensive panel_v1.0 Summary: Ag6158 Moderate expression of this gene is seen mainly in control colon sample from Crohn's and ulcerative colitis patient (CTs=30-33.5). Low expression of this gene is also seen in normal lung, bone and synovial

fluid cells from rheumotoid arthritis patient and emphysema sample. Therefore, therapeutic modulation of this gene or its protein product may be useful in the treatment of Crohn's disease, ulcerativ colitis, asthma, allergy, emphysema and rheumatoid arthritis.

General_screening_panel_v1.7 Summary: Ag6158 Highest expression of this gene is seen in lymphoid organ LN pool (CT=28.2). In addition, moderate expression of this gene is also seen in lung and trachea. Therefore, therefore, therapeutic modulation of this gene or its protein product may be useful in the treatment of immunological disorders including autoimmune diseases and asthma.

This gene shows moderate to low expression in normal tissues. Among tissues with metabolic or endocrine function, this gene is expressed at low to moderate levels in adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, fetal heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

This gene is expressed at much higher levels in fetal (CTs=33) when compared to adult heart, liver (CTs=36). This observation suggests that expression of this gene can be used to distinguish fetal from adult heart and liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance heart and liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of heart and liver related diseases.

In addition, low expression of this gene is also seen in whole brain and thalamus

Oncology_cell_line_screening_panel_v3.2 Summary: Ag6158 Low expression of this gene is exclusively seen in two gastric cancer cell lines (CTs=33.8-34.9). Therefore, expression of this gene may be used as marker to detect the presence of gastric cancer and also, therapeutic modulation of this gene may be useful in the treatment of gastric cancer.

Panel 4.1D Summary: Ag6158 Highest expression of this gene is seen in activated secondary Th2 cells (CT=29.7). This gene is upregulated in activated primary and secondary Th1, Th2, and Tr1 cells, PMA/ionomycin treated LAK cells and eosinophils, activated PBMC cells, activated B lymphocytes, activated monocytes and macrophages. Moderate expression of this gene is also seen in resting secondary CD8 lymphocyte. Therefore, the gene, encoded protein, antibodies to the protein or small molecule drugs that antagonzie the function of this gene product may reduce or eliminate the symptoms in

patients with several types of autoimmune and inflammatory diseases, such as lupus erythematosus, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, or psoriasis.

D. NOV 6: CG176697: Novel Plasma Membrane Protein.

Expression of NOV 6b gene CG176697-01 was assessed using the primer-probe set Ag6406, described in Table DA. Results of the RTQ-PCR runs are shown in Tables DB, and DC.

Table DA. Probe Name Ag6406

Primers	Sequnces	Length	Start Position	SEQ ID No
Forward	5'-catggaccatatagcttttatagagaatt-3'	29	1584	126
Probe	TET-5'-cacctggtgagacctcagtcctatgg-3'-TAMRA	26	1620	127
Reverse	5'-aatagaaaactgtaggaatttgcatagtag-3'	30	1648	128

Table DB. General screening panel v1.6

Column A - Rel. Exp. (%) Ag6406, Run 277248428			
Tissue Name	**A**	**Tissue Name**	**A**
Adipose	15.1	Renal ca. TK-10	47.6
Melanoma* Hs688(A).T	30.8	Bladder	29.3
Melanoma* Hs688(B).T	21.2	Gastric ca. (liver met.) NCI-N87	52.9
Melanoma* M14	22.8	Gastric ca. KATO III	38.2
Melanoma* LOXIMVI	9.9	Colon ca. SW-948	17.1
Melanoma* SK-MEL-5	45.7	Colon ca. SW480	25.9
Squamous cell carcinoma SCC-4	29.9	Colon ca.* (SW480 met) SW620	17.0
Testis Pool	17.8	Colon ca. HT29	6.3
Prostate ca.* (bone met) PC-3	26.6	Colon ca. HCT-116	38.2
Prostate Pool	20.0	Colon ca. CaCo-2	32.1
Placenta	1.3	Colon cancer tissue	26.4
Uterus Pool	5.5	Colon ca. SW1116	11.7
Ovarian ca. OVCAR-3	74.7	Colon ca. Colo-205	6.2
Ovarian ca. SK-OV-3	87.7	Colon ca. SW-48	15.5
Ovarian ca. OVCAR-4	11.7	Colon Pool	14.9
Ovarian ca. OVCAR-5	59.0	Small Intestine Pool	18.9
Ovarian ca. IGROV-1	50.3	Stomach Pool	24.0
Ovarian ca. OVCAR-8	35.8	Bone Marrow Pool	13.1
Ovary	12.5	Fetal Heart	22.4
Breast ca. MCF-7	62.0	Heart Pool	13.1
Breast ca. MDA-MB-231	16.3	Lymph Node Pool	22.1

Breast ca. BT 549	60.3	Fetal Skeletal Muscle	5.8
Breast ca. T47D	29.9	Skeletal Muscle Pool	2.1
Breast ca. MDA-N	20.0	Spleen Pool	9.7
Breast Pool	17.1	Thymus Pool	11.7
Trachea	12.2	CNS cancer (glio/astro) U87-MG	58.2
Lung	22.5	CNS cancer (glio/astro) U-118-MG	16.2
Fetal Lung	39.5	CNS cancer (neuro;met) SK-N-AS	58.2
Lung ca. NCI-N417	9.9	CNS cancer (astro) SF-539	10.4
Lung ca. LX-1	10.9	CNS cancer (astro) SNB-75	**100.0**
Lung ca. NCI-H146	8.0	CNS cancer (glio) SNB-19	38.7
Lung ca. SHP-77	58.2	CNS cancer (glio) SF-295	41.2
Lung ca. A549	25.2	Brain (Amygdala) Pool	8.5
Lung ca. NCI-H526	4.6	Brain (cerebellum)	28.9
Lung ca. NCI-H23	36.6	Brain (fetal)	15.8
Lung ca. NCI-H460	19.2	Brain (Hippocampus) Pool	19.1
Lung ca. HOP-62	24.0	Cerebral Cortex Pool	21.2
Lung ca. NCI-H522	60.3	Brain (Substantia nigra) Pool	11.1
Liver	2.7	Brain (Thalamus) Pool	25.7
Fetal Liver	21.0	Brain (whole)	13.4
Liver ca. HepG2	16.5	Spinal Cord Pool	9.8
Kidney Pool	38.4	Adrenal Gland	21.8
Fetal Kidney	30.8	Pituitary gland Pool	7.1
Renal ca. 786-0	76.3	Salivary Gland	6.1
Renal ca. A498	19.5	Thyroid (female)	13.1
Renal ca. ACHN	26.6	Pancreatic ca. CAPAN2	27.7
Renal ca. UO-31	37.6	Pancreas Pool	19.1

Table DC. Panel 4.1D

Colun A - Rel. Exp.(%) Ag6406, Run 269239991			
Tissue Name	**A**	**Tissue Name**	**A**
Secondary Th1 act	34.2	HUVEC IL-1beta	20.9
Secondary Th2 act	43.5	HUVEC IFN gamma	21.2
Secondary Tr1 act	35.8	HUVEC TNF alpha + IFN gamma	3.3
Secondary Th1 rest	2.1	HUVEC TNF alpha + IL4	9.5
Secondary Th2 rest	4.8	HUVEC IL-11	10.1
Secondary Tr1 rest	7.6	Lung Microvascular EC none	54.0
Primary Th1 act	4.0	Lung Microvascular EC TNFalpha + IL-1beta	15.7
Primary Th2 act	39.2	Microvascular Dermal EC none	7.7
Primary Tr1 act	27.9	Microsvasular Dermal EC TNFalpha + IL-1beta	4.5
Primary Th1 rest	3.9	Bronchial epithelium TNFalpha + IL1beta	31.4
Primary Th2 rest	3.1	Small airway epithelium none	19.6
Primary Tr1 rest	0.8	Small airway epithelium TNFalpha + IL-1beta	41.2
CD45RA CD4 lymphocyte act	21.2	Coronery artery SMC rest	9.5

CD45RO CD4 lymphocyte act	47.0	Coronery artery SMC TNFalpha + IL-1beta	10.4
CD8 lymphocyte act	25.7	Astrocytes rest	20.2
Secondary CD8 lymphocyte rest	2.5	Astrocytes TNFalpha + IL-1beta	16.4
Secondary CD8 lymphocyte act	10.4	KU-812 (Basophil) rest	71.2
CD4 lymphocyte none	2.2	KU-812 (Basophil) PMA/ionomycin	78.5
2ry Th1/Th2/Tr1_anti-CD95 CH11	5.0	CCD1106 (Keratinocytes) none	24.0
LAK cells rest	17.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	6.4
LAK cells IL-2	16.5	Liver cirrhosis	14.4
LAK cells IL-2+IL-12	2.2	NCI-H292 none	59.0
LAK cells IL-2+IFN gamma	8.0	NCI-H292 IL-4	88.9
LAK cells IL-2+ IL-18	3.4	NCI-H292 IL-9	32.1
LAK cells PMA/ionomycin	34.2	NCI-H292 IL-13	52.5
NK Cells IL-2 rest	47.0	NCI-H292 IFN gamma	38.7
Two Way MLR 3 day	16.8	HPAEC none	12.9
Two Way MLR 5 day	2.3	HPAEC TNF alpha + IL-1 beta	32.3
Two Way MLR 7 day	6.3	Lung fibroblast none	32.3
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	9.0
PBMC PWM	10.2	Lung fibroblast IL-4	26.1
PBMC PHA-L	12.0	Lung fibroblast IL-9	14.4
Ramos (B cell) none	34.6	Lung fibroblast IL-13	16.4
Ramos (B cell) ionomycin	44.8	Lung fibroblast IFN gamma	23.7
B lymphocytes PWM	15.2	Dermal fibroblast CCD1070 rest	18.9
B lymphocytes CD40L and IL-4	27.5	Dermal fibroblast CCD1070 TNF alpha	10.7
EOL-1 dbcAMP	46.7	Dermal fibroblast CCD1070 IL-1 beta	8.4
EOL-1 dbcAMP PMA/ionomycin	9.7	Dermal fibroblast IFN gamma	12.2
Dendritic cells none	39.8	Dermal fibroblast IL-4	27.9
Dendritic cells LPS	24.7	Dermal Fibroblasts rest	34.4
Dendritic cells anti-CD40	14.9	Neutrophils TNFa+LPS	0.0
Monocytes rest	7.7	Neutrophils rest	0.0
Monocytes LPS	14.5	Colon	15.9
Macrophages rest	27.5	Lung	3.1
Macrophages LPS	9.6	Thymus	4.2
HUVEC none	23.3	Kidney	**100.0**
HUVEC starved	19.5		

General_screening_panel_v1.6 Summary: Ag6406 Highest expression of this gene is detected in a brain cancer SNB-75 cell line (CT=28.9). Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung,

liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag6406 Highest expression of this gene is detected in kidney (CT=32.8). Low expression of this gene is seen in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, lung microvascular endothelial cell, HUVEC, macrophage, basophils, keratinocytes, activated polarized T cells, as well as, epithelial and fibroblast cell types from lung and skin. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

E. NOV 7: CG178810 POLIOVIRUS RECEPTOR.

Expression of genes CG178810-02 and CG178810-03 were assessed using the primer-probe set Ag6524, described in Table EA. Results of the RTQ-PCR runs are shown in Tables EB and EC. CG178810-02 and CG178810-03 represent full length physical clones.

Table EA. Probe Name Ag6524

Primers	Sequnces	Length	Start Position	SEQ ID No
Forward	5'-acaaaccaatcaacacaactttaatc-3'	26	187	129

Probe	TET-5'-aacgtcaccaatgccctaggagctc-3'-TAMRA	25	159	130
Reverse	5'-cactgggaggtccctctttt-3'	19	114	131

Table EB. General screening panel v1.6

Column A - Rel. Exp.(% Ag6524, Run 277240076			
Tissue Name	**A**	**Tissue Name**	**A**
Adipose	0.0	Renal ca. TK-10	52.1
Melanoma* Hs688(A).T	21.6	Bladder	10.0
Melanoma* Hs688(B).T	41.2	Gastric ca. (liver met.) NCI-N87	73.2
Melanoma* M14	17.7	Gastric ca. KATO III	62.0
Melanoma* LOXIMVI	28.5	Colon ca. SW-948	9.0
Melanoma* SK-MEL-5	21.0	Colon ca. SW480	36.3
Squamous cell carcinoma SCC-4	18.9	Colon ca.* (SW480 met) SW620	16.7
Testis Pool	3.5	Colon ca. HT29	19.2
Prostate ca.* (bone met) PC-3	26.8	Colon ca. HCT-116	34.9
Prostate Pool	4.5	Colon ca. CaCo-2	30.6
Placenta	5.7	Colon cancer tissue	13.3
Uterus Pool	1.0	Colon ca. SW1116	3.4
Ovarian ca. OVCAR-3	11.7	Colon ca. Colo-205	4.8
Ovarian ca. SK-OV-3	36.9	Colon ca. SW-48	3.6
Ovarian ca. OVCAR-4	43.5	Colon Pool	4.5
Ovarian ca. OVCAR-5	29.1	Small Intestine Pool	2.1
Ovarian ca. IGROV-1	31.6	Stomach Pool	2.9
Ovarian ca. OVCAR-8	12.9	Bone Marrow Pool	2.4
Ovary	3.8	Fetal Heart	2.5
Breast ca. MCF-7	9.3	Heart Pool	4.6
Breast ca. MDA-MB-231	39.8	Lymph Node Pool	4.8
Breast ca. BT 549	52.5	Fetal Skeletal Muscle	4.0
Breast ca. T47D	4.6	Skeletal Muscle Pool	1.1
Breast ca. MDA-N	2.2	Spleen Pool	2.5
Breast Pool	4.9	Thymus Pool	3.4
Trachea	4.0	CNS cancer (glio/astro) U87-MG	22.2
Lung	1.3	CNS cancer (glio/astro) U-118-MG	41.2
Fetal Lung	10.4	CNS cancer (neuro;met) SK-N-AS	9.9
Lung ca. NCI-N417	3.9	CNS cancer (astro) SF-539	36.3
Lung ca. LX-1	28.1	CNS cancer (astro) SNB-75	31.0
Lung ca. NCI-H146	8.5	CNS cancer (glio) SNB-19	31.0
Lung ca. SHP-77	19.3	CNS cancer (glio) SF-295	**100.0**
Lung ca. A549	52.9	Brain (Amygdala) Pool	2.6
Lung ca. NCI-H526	8.0	Brain (cerebellum)	14.2
Lung ca. NCI-H23	18.8	Brain (fetal)	8.8
Lung ca. NCI-H460	18.4	Brain (Hippocampus) Pool	4.2

Lung ca. HOP-62	33.7	Cerebral Cortex Pool	4.6
Lung ca. NCI-H522	18.7	Brain (Substantia nigra) Pool	3.5
Liver	2.1	Brain (Thalamus) Pool	4.6
Fetal Liver	8.8	Brain (whole)	4.5
Liver ca. HepG2	10.2	Spinal Cord Pool	3.6
Kidney Pool	8.1	Adrenal Gland	10.3
Fetal Kidney	4.4	Pituitary gland Pool	2.8
Renal ca. 786-0	44.8	Salivary Gland	1.4
Renal ca. A498	23.2	Thyroid (female)	1.4
Renal ca. ACHN	10.5	Pancreatic ca. CAPAN2	32.8
Renal ca. UO-31	37.4	Pancreas Pool	4.6

Table EC. Panel 4.1D

Colun A - Rel. Exp.(%) Ag6524, Run 271411897			
Tissue Name	**A**	**Tissue Name**	**A**
Secondary Th1 act	10.2	HUVEC IL-1beta	58.6
Secondary Th2 act	13.7	HUVEC IFN gamma	45.1
Secondary Tr1 act	3.2	HUVEC TNF alpha + IFN gamma	34.9
Secondary Th1 rest	0.3	HUVEC TNF alpha + IL4	47.3
Secondary Th2 rest	0.6	HUVEC IL-11	16.6
Secondary Tr1 rest	0.1	Lung Microvascular EC none	64.6
Primary Th1 act	1.9	Lung Microvascular EC TNFalpha + IL-1beta	15.7
Primary Th2 act	6.3	Microvascular Dermal EC none	10.7
Primary Tr1 act	7.9	Microsvasular Dermal EC TNFalpha + IL-1beta	15.5
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	9.0
Primary Th2 rest	0.2	Small airway epithelium none	17.1
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	40.9
CD45RA CD4 lymphocyte act	14.9	Coronery artery SMC rest	25.2
CD45RO CD4 lymphocyte act	3.4	Coronery artery SMC TNFalpha + IL-1beta	33.4
CD8 lymphocyte act	1.2	Astrocytes rest	7.6
Secondary CD8 lymphocyte rest	1.4	Astrocytes TNFalpha + IL-1beta	2.6
Secondary CD8 lymphocyte act	1.3	KU-812 (Basophil) rest	2.2
CD4 lymphocyte none	0.2	KU-812 (Basophil) PMA/ionomycin	2.6
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	20.4
LAK cells rest	1.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	12.2
LAK cells IL-2	0.8	Liver cirrhosis	2.2
LAK cells IL-2+IL-12	0.1	NCI-H292 none	9.7
LAK cells IL-2+IFN gamma	0.4	NCI-H292 IL-4	17.1
LAK cells IL-2+ IL-18	0.2	NCI-H292 IL-9	15.5
LAK cells PMA/ionomycin	3.0	NCI-H292 IL-13	18.4
NK Cells IL-2 rest	0.2	NCI-H292 IFN gamma	11.7
Two Way MLR 3 day	1.1	HPAEC none	10.4
Two Way MLR 5 day	0.3	HPAEC TNF alpha + IL-1 beta	82.4

Two Way MLR 7 day	1.3	Lung fibroblast none	**100.0**
PBMC rest	0.2	Lung fibroblast TNF alpha + IL-1 beta	9.5
PBMC PWM	2.0	Lung fibroblast IL-4	31.4
PBMC PHA-L	0.7	Lung fibroblast IL-9	21.9
Ramos (B cell) none	0.0	Lung fibroblast IL-13	20.9
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	59.5
B lymphocytes PWM	1.6	Dermal fibroblast CCD1070 rest	39.8
B lymphocytes CD40L and IL-4	0.8	Dermal fibroblast CCD1070 TNF alpha	42.9
EOL-1 dbcAMP	8.5	Dermal fibroblast CCD1070 IL-1 beta	30.1
EOL-1 dbcAMP PMA/ionomycin	2.2	Dermal fibroblast IFN gamma	4.9
Dendritic cells none	7.7	Dermal fibroblast IL-4	9.2
Dendritic cells LPS	4.6	Dermal Fibroblasts rest	11.5
Dendritic cells anti-CD40	1.8	Neutrophils TNFa+LPS	0.5
Monocytes rest	1.2	Neutrophils rest	2.1
Monocytes LPS	8.8	Colon	0.1
Macrophages rest	1.1	Lung	0.6
Macrophages LPS	2.0	Thymus	0.7
HUVEC none	36.1	Kidney	6.8
HUVEC starved	48.3		

General_screening_panel_v1.6 Summary: Ag6524 Highest expression of this gene is seen in a brain cancer SF-295 cell line. Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as

Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag6524 Highest expression of this gene is seen in lung fibroblasts (CT=27). This gene shows a wide spread expression in this panel, with moderate expression seen in endothelial cells, and fibroblasts cells. Moderate to low expression of this gene is seen in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

F. NOV 8: CG179299 Novel Membrane Protein.

Expression of gene CG179299-01 was assessed using the primer-probe set Ag6490, described in Table FA. Results of the RTQ-PCR runs are shown in Tables FB and FC.

Table FA. Probe Name Ag6490

Primers	Seuences	Length	Start Position	SEQ ID No
Forward	5'-gcgtttgctgagagtagtatctgt-3'	24	981	132
Probe	TET-5'-cattcagttaccaaaagactgtaaaactgc-3'-TAMRA	30	1014	133
Reverse	5'-atttcaccagaagtgacagaaatatc-3'	26	1047	134

Table FB. General screening panel v1.6

Column A - Rel. Exp. (%) Ag6490, Run 277240055			
Tissue Name	**A**	**Tissue Name**	**A**
Adipose	19.5	Renal ca. TK-10	27.5
Melanoma* Hs688(A).T	51.1	Bladder	43.8
Melanoma* Hs688(B).T	53.2	Gastric ca. (liver met.) NCI-N87	71.7
Melanoma* M14	24.3	Gastric ca. KATO III	44.1
Melanoma* LOXIMVI	31.0	Colon ca. SW-948	6.0
Melanoma* SK-MEL-5	39.0	Colon ca. SW480	22.7
Squamous cell carcinoma SCC-4	12.5	Colon ca.* (SW480 met) SW620	24.7
Testis Pool	17.3	Colon ca. HT29	8.6
Prostate ca.* (bone met) PC-3	68.3	Colon ca. HCT-116	43.2

Prostate Pool	23.8	Colon ca. CaCo-2	27.7
Placenta	27.9	Colon cancer tissue	33.9
Uterus Pool	12.2	Colon ca. SW1116	3.2
Ovarian ca. OVCAR-3	26.6	Colon ca. Colo-205	3.2
Ovarian ca. SK-OV-3	48.0	Colon ca. SW-48	5.3
Ovarian ca. OVCAR-4	23.0	Colon Pool	37.1
Ovarian ca. OVCAR-5	61.1	Small Intestine Pool	31.4
Ovarian ca. IGROV-1	17.3	Stomach Pool	23.3
Ovarian ca. OVCAR-8	12.1	Bone Marrow Pool	15.0
Ovary	25.2	Fetal Heart	18.2
Breast ca. MCF-7	66.0	Heart Pool	12.4
Breast ca. MDA-MB-231	29.1	Lymph Node Pool	39.5
Breast ca. BT 549	46.7	Fetal Skeletal Muscle	12.2
Breast ca. T47D	20.0	Skeletal Muscle Pool	3.6
Breast ca. MDA-N	10.3	Spleen Pool	21.9
Breast Pool	40.1	Thymus Pool	35.6
Trachea	25.2	CNS cancer (glio/astro) U87-MG	47.0
Lung	8.4	CNS cancer (glio/astro) U-118-MG	**100.0**
Fetal Lung	75.8	CNS cancer (neuro;met) SK-N-AS	45.4
Lung ca. NCI-N417	4.1	CNS cancer (astro) SF-539	31.4
Lung ca. LX-1	21.9	CNS cancer (astro) SNB-75	54.0
Lung ca. NCI-H146	9.8	CNS cancer (glio) SNB-19	16.8
Lung ca. SHP-77	35.8	CNS cancer (glio) SF-295	80.7
Lung ca. A549	72.7	Brain (Amygdala) Pool	20.0
Lung ca. NCI-H526	4.8	Brain (cerebellum)	45.7
Lung ca. NCI-H23	62.4	Brain (fetal)	42.3
Lung ca. NCI-H460	22.5	Brain (Hippocampus) Pool	22.1
Lung ca. HOP-62	23.3	Cerebral Cortex Pool	25.5
Lung ca. NCI-H522	20.9	Brain (Substantia nigra) Pool	15.8
Liver	3.4	Brain (Thalamus) Pool	34.2
Fetal Liver	42.6	Brain (whole)	31.0
Liver ca. HepG2	11.9	Spinal Cord Pool	22.2
Kidney Pool	67.4	Adrenal Gland	31.9
Fetal Kidney	42.3	Pituitary gland Pool	9.8
Renal ca. 786-0	46.3	Salivary Gland	20.4
Renal ca. A498	11.4	Thyroid (female)	11.9
Renal ca. ACHN	38.2	Pancreatic ca. CAPAN2	15.1
Renal ca. UO-31	45.4	Pancreas Pool	21.3

Table FC. Panel 4.1D

Colun A - Rel. Exp.(%) Ag6490, Run 268350100			
Tissue Name	**A**	**Tissue Name**	**A**
Secondary Th1 act	42.3	HUVEC IL-1beta	27.4

Secondary Th2 act	64.2	HUVEC IFN gamma	34.2
Secondary Tr1 act	14.5	HUVEC TNF alpha + IFN gamma	11.1
Secondary Th1 rest	3.7	HUVEC TNF alpha + IL4	9.3
Secondary Th2 rest	8.0	HUVEC IL-11	13.9
Secondary Tr1 rest	7.9	Lung Microvascular EC none	64.6
Primary Th1 act	6.4	Lung Microvascular EC TNFalpha + IL-1beta	12.0
Primary Th2 act	51.1	Microvascular Dermal EC none	9.9
Primary Tr1 act	36.3	Microsvasular Dermal EC TNFalpha + IL-1beta	8.6
Primary Th1 rest	1.5	Bronchial epithelium TNFalpha + IL1beta	10.4
Primary Th2 rest	9.5	Small airway epithelium none	10.5
Primary Tr1 rest	0.6	Small airway epithelium TNFalpha + IL-1beta	16.5
CD45RA CD4 lymphocyte act	34.9	Coronery artery SMC rest	25.5
CD45RO CD4 lymphocyte act	62.0	Coronery artery SMC TNFalpha + IL-1beta	37.1
CD8 lymphocyte act	9.7	Astrocytes rest	7.2
Secondary CD8 lymphocyte rest	14.4	Astrocytes TNFalpha + IL-1beta	5.0
Secondary CD8 lymphocyte act	7.7	KU-812 (Basophil) rest	20.6
CD4 lymphocyte none	6.3	KU-812 (Basophil) PMA/ionomycin	40.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	13.8	CCD1106 (Keratinocytes) none	17.3
LAK cells rest	16.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	8.5
LAK cells IL-2	12.0	Liver cirrhosis	9.3
LAK cells IL-2+IL-12	1.6	NCI-H292 none	21.0
LAK cells IL-2+IFN gamma	8.4	NCI-H292 IL-4	46.3
LAK cells IL-2+ IL-18	6.0	NCI-H292 IL-9	44.1
LAK cells PMA/ionomycin	26.8	NCI-H292 IL-13	38.2
NK Cells IL-2 rest	57.8	NCI-H292 IFN gamma	15.5
Two Way MLR 3 day	21.0	HPAEC none	13.2
Two Way MLR 5 day	5.3	HPAEC TNF alpha + IL-1 beta	35.4
Two Way MLR 7 day	5.7	Lung fibroblast none	30.8
PBMC rest	5.1	Lung fibroblast TNF alpha + IL-1 beta	34.9
PBMC PWM	9.0	Lung fibroblast IL-4	19.6
PBMC PHA-L	7.3	Lung fibroblast IL-9	17.0
Ramos (B cell) none	14.8	Lung fibroblast IL-13	7.8
Ramos (B cell) ionomycin	55.1	Lung fibroblast IFN gamma	45.1
B lymphocytes PWM	13.1	Dermal fibroblast CCD1070 rest	49.7
B lymphocytes CD40L and IL-4	42.3	Dermal fibroblast CCD1070 TNF alpha	**100.0**
EOL-1 dbcAMP	37.1	Dermal fibroblast CCD1070 IL-1 beta	29.1
EOL-1 dbcAMP PMA/ionomycin	13.8	Dermal fibroblast IFN gamma	17.2
Dendritic cells none	15.4	Dermal fibroblast IL-4	34.2
Dendritic cells LPS	7.1	Dermal Fibroblasts rest	27.2
Dendritic cells anti-CD40	17.6	Neutrophils TNFa+LPS	4.2
Monocytes rest	10.4	Neutrophils rest	35.8
Monocytes LPS	32.5	Colon	3.4
Macrophages rest	7.7	Lung	2.0
Macrophages LPS	0.2	Thymus	7.4

HUVEC none	16.6	Kidney	29.5
HUVEC starved	20.9		

General_screening_panel_v1.6 Summary: Ag6490 Highest expression of this gene is detected in a brain cancer U-118-MG cell line (CT=28). Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

This gene is expressed at much higher levels in fetal (CTs=28-29) when compared to adult lung and liver (CTs=31.7-33). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung and liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance lung and liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung liver related diseases.

Panel 4.1D Summary: Ag6490 Highest expression of this gene is detected in TNF alpha treated dermal fibroblast CCD1070 (CT=29.8). This gene is expressed at moderate to low levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell,

macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General_screening_panel_v1.6 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

G. NOV 9: CG50301, humanTENM4.

Expression of gene CG50301-01 was assessed using the primer-probe sets Ag2581, Ag2910 and Ag6931, described in Tables GA, GB and GC. Results of the RTQ-PCR runs are shown in Tables GD, GE, GF, GG, GH, GI, GJ, GK, GL and GM.

Table GA. Probe Name Ag2581

Primers	Sequnces	Length	Start Position	SEQ ID No
Forward	5'-tgaccacagacatcatcagtgt-3'	22	7770	135
Probe	TET-5'-ccatcttgaaccatgcccactaccta-3'-TAMRA	26	7821	136
Reverse	5'-tcaatggtgaagtgcaggtt-3'	20	7850	137

Table GB. Probe Name Ag2910

Primers	Sequnces	Length	Start Position	SEQ ID No
Forward	5'-tgaccacagacatcatcagtgt-3'	22	7770	138
Probe	TET-5'-ccatcttgaaccatgcccactaccta-3'-TAMRA	26	7821	139
Reverse	5'-tcaatggtgaagtgcaggtt-3'	20	7850	140

Table GC. Probe Name Ag6931

Primers	Sequenes	Length	Start Position	SEQ ID No
Forward	5'-tgtatcgatgtggcctgca-3'	19	1922	141

Probe	TET-5'-accatggcacctgcatcacgggc-3'-TAMRA	23	1944	142
Reverse	5'-tccacttcctcacagctct-3'	19	1998	143

Table GD. Ardais Panel v.1.0

Column A - Rel. Exp.(%) Ag2581, Run 263147833			
Tissue Name	**A**	**Tissue Name**	**A**
136799_Lung cancer(362)	20.3	136787_lung cancer(356)	0.7
136800_Lung NAT(363)	6.3	136788_lung NAT(357)	0.9
136813_Lung cancer(372)	**100.0**	136804_Lung cancer(369)	5.6
136814_Lung NAT(373)	1.5	136805_Lung NAT(36A)	2.5
136815_Lung cancer(374)	9.5	136806_Lung cancer(36B)	16.5
136816_Lung NAT(375)	36.1	136807_Lung NAT(36C)	0.6
136791_Lung cancer(35A)	4.9	136789_lung cancer(358)	3.0
136795_Lung cancer(35E)	5.3	136802_Lung cancer(365)	18.8
136797_Lung cancer(360)	2.6	136803_Lung cancer(368)	10.7
136794_lung NAT(35D)	1.5	136811_Lung cancer(370)	5.8
136818_Lung NAT(377)	2.8	136810_Lung NAT(36F)	13.1

Table GE. General_screening_panel_v1.6

Column A - Rel. Exp.(%) Ag6931, Run 278388814			
Tissue Name	**A**	**Tissue Name**	**A**
Adipose	1.3	Renal ca. TK-10	11.3
Melanoma* Hs688(A).T	11.7	Bladder	3.1
Melanoma* Hs688(B).T	11.6	Gastric ca. (liver met.) NCI-N87	8.5
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	7.2	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.1	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	2.6	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	2.6	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	**100.0**	Colon ca. HCT-116	0.0
Prostate Pool	1.0	Colon ca. CaCo-2	1.9
Placenta	0.1	Colon cancer tissue	4.9
Uterus Pool	1.2	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	2.1	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.2	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.1	Colon Pool	4.0
Ovarian ca. OVCAR-5	11.3	Small Intestine Pool	3.0
Ovarian ca. IGROV-1	5.0	Stomach Pool	4.0
Ovarian ca. OVCAR-8	2.1	Bone Marrow Pool	3.0
Ovary	41.8	Fetal Heart	1.7
Breast ca. MCF-7	0.5	Heart Pool	1.8

Breast ca. MDA-MB-231	0.6	Lymph Node Pool	13.1
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	4.7
Breast ca. T47D	0.1	Skeletal Muscle Pool	0.3
Breast ca. MDA-N	0.2	Spleen Pool	0.7
Breast Pool	6.7	Thymus Pool	4.7
Trachea	6.3	CNS cancer (glio/astro) U87-MG	17.6
Lung	6.0	CNS cancer (glio/astro) U-118-MG	0.7
Fetal Lung	13.1	CNS cancer (neuro;met) SK-N-AS	17.4
Lung ca. NCI-N417	1.3	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	6.2
Lung ca. NCI-H146	3.5	CNS cancer (glio) SNB-19	4.3
Lung ca. SHP-77	3.1	CNS cancer (glio) SF-295	6.1
Lung ca. A549	0.5	Brain (Amygdala) Pool	10.6
Lung ca. NCI-H526	23.7	Brain (cerebellum)	10.4
Lung ca. NCI-H23	0.0	Brain (fetal)	73.2
Lung ca. NCI-H460	0.2	Brain (Hippocampus) Pool	13.0
Lung ca. HOP-62	0.6	Cerebral Cortex Pool	20.3
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	11.6
Liver	0.0	Brain (Thalamus) Pool	20.0
Fetal Liver	0.5	Brain (whole)	22.7
Liver ca. HepG2	0.0	Spinal Cord Pool	15.0
Kidney Pool	4.7	Adrenal Gland	4.4
Fetal Kidney	20.3	Pituitary gland Pool	12.5
Renal ca. 786-0	5.7	Salivary Gland	0.6
Renal ca. A498	0.6	Thyroid (female)	6.3
Renal ca. ACHN	10.1	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	47.6	Pancreas Pool	0.5

Table GF. HASS Panel v1.0

Column A Rel. Exp.(%) Ag2910, Run 264389796			
Tissue Name	**A**	**Tissue Name**	**A**
MCF-7 C1	1.7	U87-MG F1 (B)	10.0
MCF-7 C2	1.6	U87-MG F2	16.0
MCF-7 C3	1.7	U87-MG F3	20.3
MCF-7 C4	2.4	U87-MG F4	24.8
MCF-7 C5	3.4	U87-MG F5	29.1
MCF-7 C6	2.9	U87-MG F6	46.3
MCF-7 C7	2.3	U87-MG F7	43.5
MCF-7 C9	2.8	U87-MG F8	40.9
MCF-7 C10	1.1	U87-MG F9	20.3
MCF-7 C11	0.0	U87-MG F10	49.0
MCF-7 C12	0.5	U87-MG F11	48.0
MCF-7 C13	3.7	U87-MG F12	27.4

MCF-7 C15	0.4	U87-MG F13	37.6
MCF-7 C16	1.2	U87-MG F14	33.4
MCF-7 C17	0.9	U87-MG F15	36.6
T24 D1	40.3	U87-MG F16	26.8
T24 D2	19.9	U87-MG F17	39.2
T24 D3	79.0	LnCAP A1	0.0
T24 D4	68.3	LnCAP A2	0.0
T24 D5	79.6	LnCAP A3	0.0
T24 D6	2.5	LnCAP A4	0.0
T24 D7	3.2	LnCAP A5	0.3
T24 D9	1.5	LnCAP A6	0.3
T24 D10	26.2	LnCAP A7	0.2
T24 D11	38.4	LnCAP A8	0.3
T24 D12	2.7	LnCAP A9	0.2
T24 D13	1.4	LnCAP A10	0.5
T24 D15	1.2	LnCAP A11	0.0
T24 D16	1.4	LnCAP A12	0.0
T24 D17	1.4	LnCAP A13	0.0
CAPaN B1	0.0	LnCAP A14	0.0
CAPaN B2	0.0	LnCAP A15	0.0
CAPaN B3	0.0	LnCAP A16	0.0
CAPaN B4	0.0	LnCAP A17	0.0
CAPaN B5	0.0	Primary Astrocytes	82.9
CAPaN B6	0.0	Primary Renal Proximal Tubule Epithelial cell A2	89.5
CAPaN B7	0.0	Primary melanocytes A5	0.8
CAPaN B8	0.0	126443 - 341 medullo	0.0
CAPaN B9	0.0	126444 - 487 medullo	40.9
CAPaN B10	0.3	126445 - 425 medullo	20.9
CAPaN B11	0.0	126446 - 690 medullo	26.1
CAPaN B12	0.0	126447 - 54 adult glioma	6.3
CAPaN B13	0.0	126448 - 245 adult glioma	0.2
CAPaN B14	0.0	126449 - 317 adult glioma	0.0
CAPaN B15	0.0	126450 - 212 glioma	**100.0**
CAPaN B16	0.0	126451 - 456 glioma	1.2
CAPaN B17	0.0		

Table GG. Panel 1.3D

Colun A - Rel. Exp.(%) Ag2581, Run 162292620 Column B - Rel. Exp.(%) Ag2910, Run 162556486					
Tissue Name	**A**	**B**	**Tissue Name**	**A**	**B**
Liver adenocarcinoma	0.0	0.0	Kidney (fetal)	5.8	4.7
Pancreas	0.2	0.0	Renal ca. 786-0	1.7	0.1
Pancreatic ca. CAPAN 2	0.5	0.0	Renal ca. A498	0.8	0.9

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 224

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 224

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

CLAIMS

What is claimed is:

1. An isolated polypeptide comprising the mature form of an amino acid sequenced selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 58.

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 58.

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 58.

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 58.

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 activity of or latency or 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 58 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 58.

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 58.

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 58.

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 58.

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 58, 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 58.

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 58.

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.

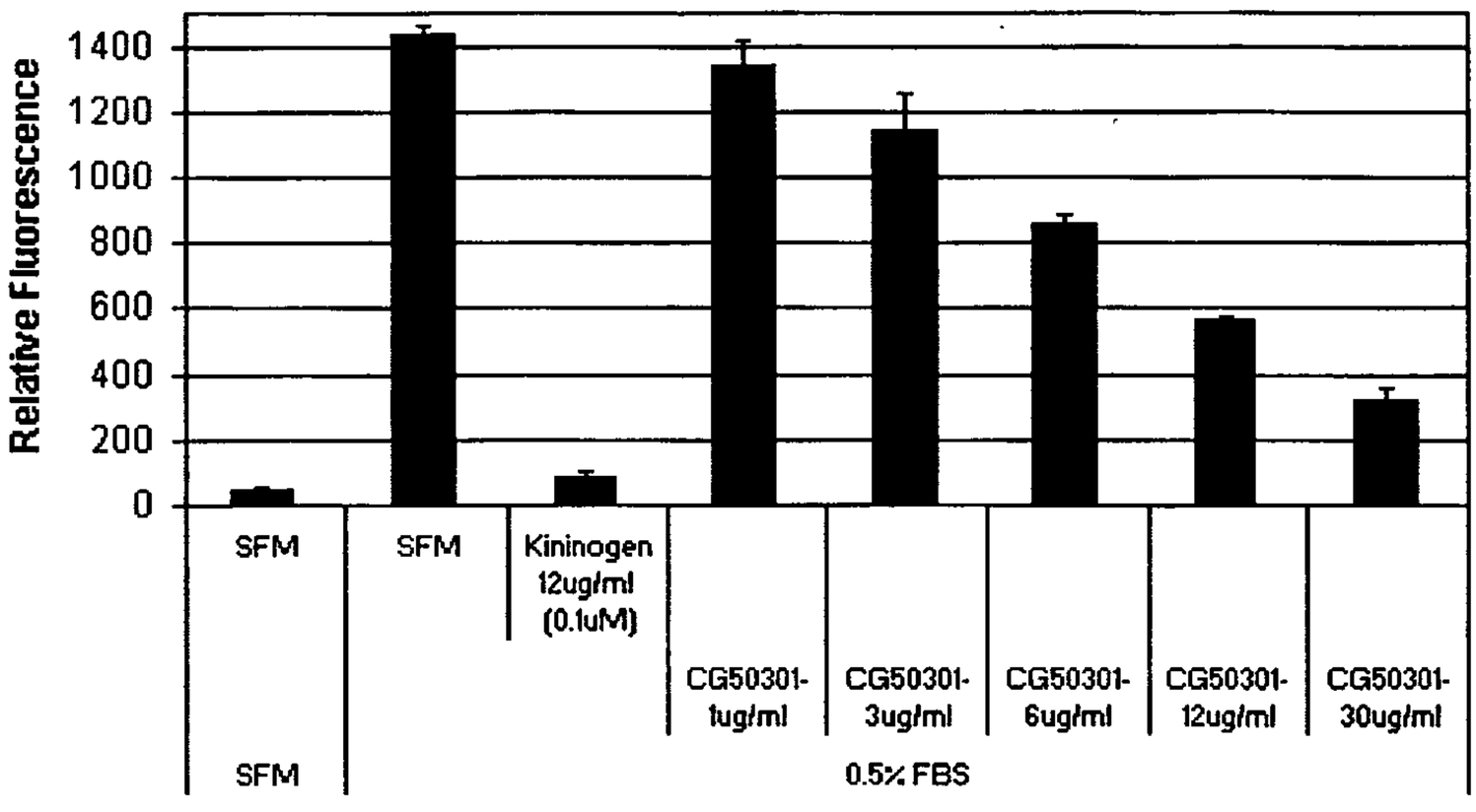

Relative Fluorescence
1400
1200
1000
800
600
400
200
0
SFM
SFM
Kininogen 12ug/ml (0.1uM)
CG50301-1ug/ml
CG50301-3ug/ml
CG50301-6ug/ml
CG50301-12ug/ml
CG50301-30ug/ml
SFM
0.5% FBS

Figure 1