



(86) Date de dépôt PCT/PCT Filing Date: 2002/03/08
(87) Date publication PCT/PCT Publication Date: 2002/10/10
(85) Entrée phase nationale/National Entry: 2003/09/05
(86) N° demande PCT/PCT Application No.: US 2002/007355
(87) N° publication PCT/PCT Publication No.: 2002/079398
(30) Priorités/Priorities: 2001/03/08 (60/274,194) US;
2001/03/08 (60/274,322) US; 2001/03/08 (60/274,281) US;
2001/03/09 (60/274,849) US; 2001/03/13 (60/275,578) US;
2001/03/13 (60/275,601) US; 2001/03/13 (60/275,579) US;
2001/03/14 (60/276,000) US; 2001/03/16 (60/276,776) US;
2001/03/19 (60/276,994) US; 2001/03/20 (60/277,239) US;
2001/03/20 (60/277,338) US; 2001/03/20 (60/277,327) US;
...

(51) Cl.Int.⁷/Int.Cl.⁷ C12N 15/63, A61K 38/16, C12N 15/12,
C07K 14/00

(71) Demandeur/Applicant:
CURAGEN CORPORATION, US

(72) Inventeurs/Inventors:
KEKUDA, RAMESH, US;
TCHERNEV, VELIZAR T., US;
LIU, XIAOHONG, US;
SPYTEK, KIMBERLY A., US;
PATTURAJAN, MEERA, US;
BURGESS, CATHERINE E., US;
...

(74) Agent: GOUDREAU GAGE DUBUC

(54) Titre : POLYPEPTIDES THERAPEUTIQUES, LES ACIDES NUCLEIQUES LES CODANT ET LEURS PROCEDES
D'UTILISATION

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

(57) **Abrégé/Abstract:**

Disclosed herein are nucleic acid sequences that encode G-coupled protein-receptor related polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies, which immunospecifically-bind to the polypeptide, as well as derivatives, variants mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

- (30) **Priorités(suite)/Priorities(continued):** 2001/03/21 (60/277,791) US; 2001/03/22 (60/277,833) US; 2001/03/23 (60/278,152) US; 2001/03/26 (60/278,894) US; 2001/03/27 (60/279,036) US; 2001/03/27 (60/278,999) US; 2001/03/30 (60/280,233) US; 2001/04/02 (60/280,802) US; 2001/05/02 (60/288,066) US; 2001/05/02 (60/288,052) US; 2001/05/02 (60/288,228) US; 2001/05/17 (60/291,766) US; 2001/06/07 (60/296,693) US; 2001/06/08 (60/296,856) US; 2001/07/05 (60/303,237) US; 2001/07/05 (60/303,230) US; 2001/08/08 (60/310,913) US; 2001/08/13 (60/311,978) US; 2001/08/14 (60/312,191) US; 2001/08/16 (60/312,916) US; 2001/08/17 (60/313,182) US; 2001/08/20 (60/313,626) US; 2001/08/21 (60/314,018) US; 2001/08/27 (60/315,227) US; 2001/09/10 (60/318,510) US; 2001/09/10 (60/318,403) US; 2001/09/14 (60/322,296) US; 2001/09/14 (60/322,360) US; 2001/09/27 (60/325,378) US; 2001/11/09 (60/332,486) US; 2001/11/09 (60/345,399) US; 2002/03/07 (10/094,886) US
- (72) **Inventeurs(suite)/Inventors(continued):** VERNET, CORINE A. M., US; LI, LI, US; GORMAN, LINDA, US; MALYANKAR, URIEL M., US; BOLDOG, FERENC L., US; GUO, XIAOJIA, US; SHENOY, SURESH G., US; PADIGARU, MURALIDHARA, US; TAUPIER, RAYMOND J., JR., US; MILLER, CHARLES E., US; CASMAN, STACIE J., US; PENA, CAROL E. A., US; GANGOLLI, ESHA A., US; GUSEV, VLADIMIR, US; SMITHSON, GLENNDA, US; ZERHUSEN, BRYAN D., US; GERLACH, VALERIE, US; POCHART, PASCALE F. J., US; FERNANDES, ELMA R., US; SHIMKETS, RICHARD A., US; RASTELLI, LUCA, US; SPADERNA, STEVEN K., US; LAROCHELLE, WILLIAM J., US; ZHONG, MEI, US; KHRAMTSOV, NIKOLAI V., US; VOSS, EDWARD Z., US; HERRMANN, JOHN L., US

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number
WO 02/079398 A3

(51) International Patent Classification ⁷ :	C12N 15/63,	60/314,018	21 August 2001 (21.08.2001)	US
	15/12, A61K 38/16, C07K 14/00	60/315,227	27 August 2001 (27.08.2001)	US
		60/318,403	10 September 2001 (10.09.2001)	US
(21) International Application Number:	PCT/US02/07355	60/318,510	10 September 2001 (10.09.2001)	US
		60/322,360	14 September 2001 (14.09.2001)	US
(22) International Filing Date:	8 March 2002 (08.03.2002)	60/322,296	14 September 2001 (14.09.2001)	US
		60/325,378	27 September 2001 (27.09.2001)	US
(25) Filing Language:	English	60/332,486	9 November 2001 (09.11.2001)	US
		60/345,399	9 November 2001 (09.11.2001)	US
(26) Publication Language:	English	10/094,886	7 March 2002 (07.03.2002)	US

(30) Priority Data:

60/274,194	8 March 2001 (08.03.2001)	US
60/274,281	8 March 2001 (08.03.2001)	US
60/274,322	8 March 2001 (08.03.2001)	US
60/274,849	9 March 2001 (09.03.2001)	US
60/275,578	13 March 2001 (13.03.2001)	US
60/275,579	13 March 2001 (13.03.2001)	US
60/275,601	13 March 2001 (13.03.2001)	US
60/276,000	14 March 2001 (14.03.2001)	US
60/276,776	16 March 2001 (16.03.2001)	US
60/276,994	19 March 2001 (19.03.2001)	US
60/277,239	20 March 2001 (20.03.2001)	US
60/277,327	20 March 2001 (20.03.2001)	US
60/277,338	20 March 2001 (20.03.2001)	US
60/277,791	21 March 2001 (21.03.2001)	US
60/277,833	22 March 2001 (22.03.2001)	US
60/278,152	23 March 2001 (23.03.2001)	US
60/278,894	26 March 2001 (26.03.2001)	US
60/278,999	27 March 2001 (27.03.2001)	US
60/279,036	27 March 2001 (27.03.2001)	US
60/280,233	30 March 2001 (30.03.2001)	US
60/280,802	2 April 2001 (02.04.2001)	US
60/288,052	2 May 2001 (02.05.2001)	US
60/288,066	2 May 2001 (02.05.2001)	US
60/288,228	2 May 2001 (02.05.2001)	US
60/291,766	17 May 2001 (17.05.2001)	US
60/296,693	7 June 2001 (07.06.2001)	US
60/296,856	8 June 2001 (08.06.2001)	US
60/303,230	5 July 2001 (05.07.2001)	US
60/303,237	5 July 2001 (05.07.2001)	US
60/310,913	8 August 2001 (08.08.2001)	US
60/311,978	13 August 2001 (13.08.2001)	US
60/312,191	14 August 2001 (14.08.2001)	US
60/312,916	16 August 2001 (16.08.2001)	US
60/313,182	17 August 2001 (17.08.2001)	US
60/313,626	20 August 2001 (20.08.2001)	US

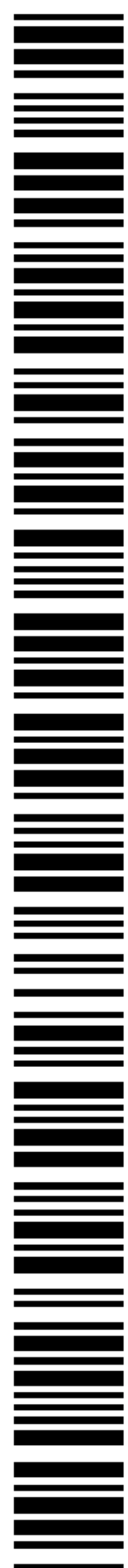
(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	10/094,886 (CIP)
Filed on	7 March 2002 (07.03.2002)
US	60/274,194 (CIP)
Filed on	8 March 2001 (08.03.2001)
US	60/274,281 (CIP)
Filed on	8 March 2001 (08.03.2001)
US	60/274,322 (CIP)
Filed on	8 March 2001 (08.03.2001)
US	60/274,849 (CIP)
Filed on	9 March 2001 (09.03.2001)
US	60/275,235 (CIP)
Filed on	12 March 2001 (12.03.2001)
US	60/275,578 (CIP)
Filed on	13 March 2001 (13.03.2001)
US	60/275,579 (CIP)
Filed on	13 March 2001 (13.03.2001)
US	60/275,601 (CIP)
Filed on	13 March 2001 (13.03.2001)
US	60/276,000 (CIP)
Filed on	14 March 2001 (14.03.2001)
US	60/276,776 (CIP)
Filed on	16 March 2001 (16.03.2001)
US	60/276,994 (CIP)
Filed on	19 March 2001 (19.03.2001)
US	60/277,239 (CIP)
Filed on	20 March 2001 (20.03.2001)
US	60/277,327 (CIP)
Filed on	20 March 2001 (20.03.2001)
US	60/277,338 (CIP)
Filed on	20 March 2001 (20.03.2001)
US	60/277,791 (CIP)
Filed on	21 March 2001 (21.03.2001)
US	60/277,833 (CIP)
Filed on	22 March 2001 (22.03.2001)

[Continued on next page]

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

(57) Abstract: Disclosed herein are nucleic acid sequences that encode G-coupled protein-receptor related polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies, which immunospecifically-bind to the polypeptide, as well as derivatives, variants mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.



WO 02/079398 A3

WO 02/079398 A3

- | | | |
|----------|--------------------------------|---|
| US | 60/278,152 (CIP) | (72) Inventors; and |
| Filed on | 23 March 2001 (23.03.2001) | (75) Inventors/Applicants (for US only): KEKUDA, Ramesh |
| US | 60/278,894 (CIP) | [IN/US]; 1213 Avalon Valley Drive, Danbury, CT 06810 |
| Filed on | 26 March 2001 (26.03.2001) | (US). TCHERNEV, Velizar, T. [BG/US]; 45 Jefferson |
| US | 60/278,999 (CIP) | Road #3-12, Branford, CT 06405 (US). LIU, Xiaohong |
| Filed on | 27 March 2001 (27.03.2001) | [CN/US]; 96 Montoya Circle, Branford, CT 06405 (US). |
| US | 60/279,036 (CIP) | SPYTEK, Kimberly, A. [US/US]; 28 Court Street #1, |
| Filed on | 27 March 2001 (27.03.2001) | New Haven, CT 06511 (US). PATTURAJAN, Meera |
| US | 60/280,233 (CIP) | [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, |
| Filed on | 30 March 2001 (30.03.2001) | CT 06405 (US). BURGESS, Catherine, E. [US/US]; |
| US | 60/280,802 (CIP) | 90 Carriage Hill Drive, Wethersfield, CT 06109 (US). |
| Filed on | 2 April 2001 (02.04.2001) | VERNET, Corine, A., M. [FR/US]; 1739 Foxon Road, |
| US | 60/288,052 (CIP) | Apartment L6, Branford, CT 06471 (US). LI, Li [CN/US]; |
| Filed on | 2 May 2001 (02.05.2001) | 56 Jerimoth Road, Branford, CT 06405 (US). GOR- |
| US | 60/288,066 (CIP) | MAN, Linda [US/US]; 329 Monticello Drive, Branford, |
| Filed on | 2 May 2001 (02.05.2001) | CT 06405 (US). MALYANKAR, Uriel, M. [IN/US]; |
| US | 60/288,228 (CIP) | 229 Branford Road, # 330, Branford, CT 06405 (US). |
| Filed on | 2 May 2001 (02.05.2001) | BOLDOG, Ferenc, L. [HU/US]; 1687 Hartford Turnpike, |
| US | 60/291,766 (CIP) | North Haven, CT 06473 (US). GUO, Xiaojia [CN/US]; |
| Filed on | 17 May 2001 (17.05.2001) | 713 Robert Frost Drive, Branford, CT 06405 (US). |
| US | 60/296,693 (CIP) | SHENOY, Suresh, G. [IN/US]; 15 Millwood Drive, |
| Filed on | 7 June 2001 (07.06.2001) | Branford, CT 06405 (US). PADIGARU, Muralidhara |
| US | 60/296,856 (CIP) | [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). |
| Filed on | 8 June 2001 (08.06.2001) | TAUPIER, Raymond, J., Jr. [US/US]; 34 Pardee Place |
| US | 60/303,230 (CIP) | Extension, East Haven, CT 06512 (US). MILLER, |
| Filed on | 5 July 2001 (05.07.2001) | Charles, E. [US/US]; 98 Saddle Hill Drive, Guilford, |
| US | 60/303,237 (CIP) | CT 06437 (US). CASMAN, Stacie, J. [US/US]; 17 Peck |
| Filed on | 5 July 2001 (05.07.2001) | Street, North Haven, CT 06473 (US). PENA, Carol, E., A. |
| US | 60/310,913 (CIP) | [US/US]; 604 Orange Street, #2, New Haven, CT 06511 |
| Filed on | 8 August 2001 (08.08.2001) | (US). GANGOLLI, Esha, A. [IN/US]; 31 Strawberry |
| US | 60/311,978 (CIP) | Hill Road, Madison, CT 06443 (US). GUSEV, Vladimir |
| Filed on | 13 August 2001 (13.08.2001) | [UA/US]; 1209 Durham Road, Madison, CT 06443 (US). |
| US | 60/312,191 (CIP) | SMITHSON, Glendda [US/US]; 125 Michael Drive, |
| Filed on | 14 August 2001 (14.08.2001) | Guilford, CT 06435 (US). ZERHUSEN, Bryan, D. |
| US | 60/312,916 (CIP) | [US/US]; 337 Monticello Drive, Branford, CT 06405 |
| Filed on | 16 August 2001 (16.08.2001) | (US). GERLACH, Valerie [US/US]; 18 Rock Pasture |
| US | 60/313,182 (CIP) | Road, Branford, CT 06405 (US). POCHART, Pascale, |
| Filed on | 17 August 2001 (17.08.2001) | F-J. [FR/US]; 99 Country Way, Madison, CT 06443 (US). |
| US | 60/313,626 (CIP) | FERNANDES, Elma, R. [IN/US]; 77 Florence Road |
| Filed on | 20 August 2001 (20.08.2001) | #2B, Branford, CT 06405 (US). SHIMKETS, Richard, |
| US | 60/314,018 (CIP) | A. [US/US]; 5 Indian Meadows Drive, Guilford, CT |
| Filed on | 21 August 2001 (21.08.2001) | 06437 (US). RASTELLI, Luca [IT/US]; 52 Pepperbush |
| US | 60/315,227 (CIP) | Lane, Guilford, CT 06437 (US). SPADERNA, Steven, |
| Filed on | 27 August 2001 (27.08.2001) | K. [US/US]; 261 Deerfield Drive, Berlin, CT 06037 (US). |
| US | 60/318,403 (CIP) | LAROCHELLE, William, J. [US/US]; 15 Devonshire |
| Filed on | 10 September 2001 (10.09.2001) | Lane, Madison, CT 06443 (US). ZHONG, Mei [CA/US]; |
| US | 60/318,510 (CIP) | Apartment 1B, 45 Harrison Avenue, Branford, CT 06405 |
| Filed on | 10 September 2001 (10.09.2001) | (US). KHRAMTSOV, Nikolai, V. [RU/US]; 38 Montoya |
| US | 60/322,296 (CIP) | Circle, Branford, CT 06405 (US). VOSS, Edward, Z. |
| Filed on | 14 September 2001 (14.09.2001) | [US/US]; 123 Knollwood Drive, Wallingford, CT 06492 |
| US | 60/322,360 (CIP) | (US). HERRMANN, John, L. [US/US]; 78 Barnshed |
| Filed on | 14 September 2001 (14.09.2001) | Lane, Guilford, CT 06437 (US). |
| US | 60/325,378 (CIP) | |
| Filed on | 27 September 2001 (27.09.2001) | |
| US | 60/332,486 (CIP) | (74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, |
| Filed on | 9 November 2001 (09.11.2001) | Glovsky, and Popeo, P., C., One Financial Center, Boston, |
| US | 60/345,399 (CIP) | MA 02111 (US). |
| Filed on | 9 November 2001 (09.11.2001) | |
- (71) **Applicant (for all designated States except US):** **CURA-
GEN CORPORATION** [US/US]; 555 Long Wharf Drive,
11th floor, New Haven, CT 06511 (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,

[Continued on next page]

WO 02/079398 A3

MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.

Published:

— *with international search report*

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

(88) Date of publication of the international search report:

21 August 2003

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 3
CONTENANT LES PAGES 1 À 215

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 3
CONTAINING PAGES 1 TO 215

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

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

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

FIELD OF THE INVENTION

5

The present invention relates to novel polypeptides 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
10 diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions.

BACKGROUND OF THE INVENTION

15

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
20 pathways include constituted of 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
25 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, such as 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
30 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 a clinical setting a subject may be suspected of suffering from a condition brought on by diminished or suppressed 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.

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 86. 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 86, 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 86. 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 86 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 86, 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 86. 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 86. 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 86 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 86 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 86 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 86 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 86, 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 86, 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 86, 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 be 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 86, 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 86, 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 86 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 86; 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 86 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 86; 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 86, 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 86 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 86, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant.

5 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 86 that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.

10 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 86, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ
15 ID NOS: 2n-1, wherein n is an integer between 1 and 86.

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 86, wherein the nucleic acid molecule
20 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 86; 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 86 is changed from that selected from the group consisting of the chosen
25 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 86; 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 86 is changed from that selected from the
30 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 86, 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 86, or a complement of the nucleotide sequence.

5 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 86, wherein the nucleic acid molecule has a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen
10 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
15 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 86. This vector can have a promoter operably linked to the nucleic acid molecule. This vector can be located within a cell.

20 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 86 in a sample, the method including providing the sample; introducing the sample to a probe that
25 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
30 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 86 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

5 predisposition to the disease.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are

10 described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following

15 detailed description and claims.

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

20 "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 1 provides a summary of the NOVX nucleic acids and their encoded polypeptides.

25

TABLE 1. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1a	CG58548-01	1	2	Neurexophilin 1 Precursor-like
1b	174307940	3	4	Neurexophilin 1 Precursor-like
1c	CG58548-02	5	6	Neurexophilin 1 Precursor-like
1d	CG58548-03	7	8	Neurexophilin 1 Precursor-like
2a	CG58542-01	9	10	Neurophilin-like
2b	169679583	11	12	Neurophilin-like
2c	169679634	13	14	Neurophilin-like
3a	CG58540-01	15	16	Cytoplasmic-Antiproteinase 3-like

4a	CG56340-03	17	18	Interferon-like
4b	174308150	19	20	Interferon-like
5a	CG58514-01	21	22	Leprecan-like
6a	CG57887-01	23	24	Tumor suppressor-like
7a	CG57885-01	25	26	Procholoecystokinin Precursor-like
8a	CG57865-01	27	28	Secreted protein-like
8b	171651532	29	30	Secreted protein-like
9a	CG54503-03	31	32	Gliacolin -like
10a	CG58600-01	33	34	Olfactomedin-like
11a	CG57572-01	35	36	CMP-N-Acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase like
12a	CG57518-01	37	38	Neural cell adhesion protein Big-2 precursor-like
12b	170108372	39	40	Neural cell adhesion protein Big-2 precursor-like
12c	170108393	41	42	Neural cell adhesion protein Big-2 precursor-like
12d	170343246	43	44	Neural cell adhesion protein Big-2 precursor-like
12e	170343692	45	46	Neural cell adhesion protein Big-2 precursor-like
12f	170684238	47	48	Neural cell adhesion protein Big-2 precursor-like
12g	170534177	49	50	Neural cell adhesion protein Big-2 precursor-like
13a	CG57409-03	51	52	Neural cell adhesion protein Big-2 precursor-like
13b	CG57409-05	53	54	MAM and Ig domain-containing protein-like
13c	CG57409-06	55	56	MAM and Ig domain-containing protein
14a	CG59262-01	57	58	Calcium binding protein S100P-like
15a	CG58635-01	59	60	S-100-like
15b	CG58635-02	61	62	Secretory carrier membrane protein-like
15c	CG58635-03	63	64	Secretory carrier membrane protein-like
16a	CG59209-01	65	66	CG3714-like
16b	174308417	67	68	CG3714-like
16c	174308429	69	70	CG3714-like
17a	CG59368-01	71	72	Preoptic regulatory factor-2-like
18a	CG58628-01	73	74	Adipophilin-like
18b	174228350	75	76	Adipophilin-like
18c	174228354	77	78	Adipophilin-like
18d	188888733	79	80	Adipophilin-like
19a	CG59342-01	81	82	FIS-like
20a	CG59486-01	83	84	Zn finger protein-like
21a	CG59446-01	85	86	Neurotransmission associated protein-like
21b	174308261	87	88	Neurotransmission associated protein-like
21c	174308266	89	90	Neurotransmission associated protein-like
21d	174308278	91	92	Neurotransmission associated protein-like
21e	174308283	93	94	Neurotransmission associated protein-like

21f	174308287	95	96	Neurotransmission associated protein-like
21g	174308293	97	98	Neurotransmission associated protein-like
21h	174308301	99	100	Neurotransmission associated protein-like
21i	174308311	101	102	Neurotransmission associated protein-like
21j	174308315	103	104	Neurotransmission associated protein-like
21k	174308321	105	106	Neurotransmission associated protein-like
21l	174308327	107	108	Neurotransmission associated protein-like
21m	174308337	109	110	Neurotransmission associated protein-like
21n	CG59446-02	111	112	Neurotransmission associated protein-like
22a	CG59375-01	113	114	Drebrin -like
23a	CG59321-01	115	116	UNC5H2-like
23b	CG59321-02	117	118	UNC5H2-like
24a	CG59591-01	119	120	Trypsin inhibitor-like
25a	CG59588-01	121	122	ISLR pecursor-like
26a	CG59584-01	123	124	Ovostatin precursor-like
26b	CG59584-02	125	126	Ovostatin precursor-like
27a	CG59417-01	127	128	Chymotrypsin precursor-like
28a	CG59415-01	129	130	Laminin type EGF-like
28b	191815704	131	132	Laminin type EGF-like
28c	191815724	133	134	Laminin type EGF-like
28d	CG59415-02	135	136	Laminin type EGF-like
29a	CG59297-01	137	138	Polycystic kidney disease 1 Protein-like
30a	CG59264-01	139	140	Polycystic kidney disease 2 Protein-like
31a	CG59623-01	141	142	Slit-like
32a	CG59247-01	143	144	Protein-tyrosine sulfotransferase-like
33a	CG59430-01	145	146	Serine Protease inhibitor-like
34a	CG59305-01	147	148	Fibronectin type III-like
34b	CG59305-02	149	150	Fibronectin type III-like
35a	CG59547-01	151	152	Adipophilin-like
36a	CG58508-01	153	154	Small inducible cytokine A4 precursor -like
36b	CG58508-02	155	156	Small inducible cytokine A4 precursor -like
36c	170072532	157	158	Small inducible cytokine A4 precursor -like
36d	170072551	159	160	Small inducible cytokine A4 precursor -like
36e	170072555	161	162	Small inducible cytokine A4 precursor -like
36f	CG58508-03	163	164	Small inducible cytokine A4 precursor -like
37a	CG59819-01	165	166	Latent transforming growth factor-like
37b	CG59819-02	167	168	Latent transforming growth factor-like
37c	CG59819-03	169	170	Latent transforming growth factor-like

38a	CG59685-01	171	172	Thrombospondin-like
38b	175070296	173	174	Thrombospondin-like
38c	175070324	175	176	Thrombospondin-like
39a	CG57167-01	177	178	Urokinase plasminogen activator surface receptor-like
40a	CG59841-01	179	180	Agrin precursor-like
41a	CG59895-01	181	182	Major urinary protein 4 precursor-like
41b	CG59895-02	183	184	Major urinary protein 4 precursor-like
42a	CG59889-01	185	186	KIAA1199-like
42b	CG59889-02	187	188	KIAA1199-like
42c	CG59889-04	189	190	KIAA1199-like
43a	CG59512-02	191	192	Small inducible cytokine A3-like
43b	CG59512-01	193	194	Small inducible cytokine A3-like
44a	CG56801-02	195	196	Thrombomodulin-like

Table 1 indicates homology of NOVX nucleic acids 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 1 will be useful in
5 therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table 1.

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
10 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 1, the NOVX polypeptides of the present invention show homology to, and contain
15 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 Examples 1-44.

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
20 molecules that modulate or inhibit diseases associated with the protein families listed in Table 1.

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 47. 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. a variety of cancers.

Additional utilities for NOVX nucleic acids and polypeptides according to the
5 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
10 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
15 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.

20 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
25 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) 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
30 acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86; (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 86, 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 86; (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 86 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 86; (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 86 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 86; (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 86, 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 86 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 86; (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 86 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 86; 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 86 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, 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 (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 post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation 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.

5 The term "probe", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), and 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
10 much slower to hybridize than shorter-length oligomer probes. Probes may be single- 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 which is separated from other nucleic acid molecules which are present in the natural source of the
15 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, 0.1 kb, or less of nucleotide sequences which naturally flank
20 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, culture medium, or of chemical precursors or other chemicals.

 A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the
25 nucleotide sequence SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 86, 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 NOS:2n-1, wherein n is an integer between 1 and 86, as a hybridization probe, NOVX molecules can be isolated using standard hybridization and
30 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,
5 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
10 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 NOS:2n-1, wherein n is an integer between 1 and 86, or a complement
15 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 NOS:2n-1, wherein n is an integer between 1 and 86, 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-
20 active portion of A NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, is one that is sufficiently complementary to the nucleotide sequence shown SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, that it can hydrogen bond with few or no mismatches to the nucleotide sequence shown SEQ ID NOS:2n-1, wherein n is an integer between 1 and
25 86, 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
30 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.

“Fragments” provided herein are defined as sequences 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 are at most some portion less than a full length sequence. Fragments may be
5 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
10 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.

“Derivatives” are nucleic acid sequences or amino acid sequences formed from the
15 native compounds either directly, by modification, or by partial substitution. “Analog” are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound, e.g. they differ 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. Homologs are
20 nucleic acid sequences or amino acid sequences of a particular gene that are 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
25 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 of the invention under
30 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

- 5 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 a human NOVX protein.
- 10 Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, 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
- 15 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
- 20 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
- 25 NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises a 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 NOS:2n-1, wherein n is an
- 30 integer between 1 and 86; or an anti-sense strand nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86; or of a naturally occurring mutant of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86.

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, 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 SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, 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 shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NOS:2n, wherein n is an integer between 1 and 86.

In addition to the human NOVX nucleotide sequences shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, 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 the human SEQ ID NOS:2n-1, wherein
5 n is an integer between 1 and 86, 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.

10 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 NOS:2n-1, wherein n is an integer between 1 and 86. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, 2000 or more nucleotides in length. In yet another
15 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
20 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
25 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 (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at
30 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 T_m , 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.

5 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
10 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 the sequences SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, corresponds to a
15 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 NOS:2n-1, wherein n is an
20 integer between 1 and 86, 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 Denhardt'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
25 well-known within the art. *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.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences SEQ ID NOS:2n-1, wherein n is an integer between 1
30 and 86, 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 SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, thereby leading to changes in the amino acid sequences of the encoded NOVX proteins, without altering the functional ability of the NOVX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence SEQ ID NOS:2n, wherein n is an integer between 1 and 86. 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 NOS:2n-1, wherein n is an integer between 1 and 86, 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 45% homologous to the amino acid sequences SEQ ID NOS:2n, wherein n is an integer between 1 and 86. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86; more preferably at least about 70% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86; still more preferably at least about 80% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86; even more preferably at least about 90% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86; and most preferably at least about 95% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86.

An isolated nucleic acid molecule encoding A NOVX protein homologous to the protein of SEQ ID NOS:2n, wherein n is an integer between 1 and 86, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, such that one or
 5 more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, 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
 10 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
 15 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
 20 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 SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

25 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
 30 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
5 regulate a specific biological function (*e.g.*, regulation of insulin release).

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
10 nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, 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
15 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 NOS:2n, wherein n is an integer between 1 and 86, or antisense nucleic acids complementary to A NOVX nucleic acid sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1
20 and 86, 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
25 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,
30 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-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (*v*), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-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
5 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. A α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the
10 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.*

15 **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
20 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
25 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 NOS:2n-1, wherein n is an integer between 1 and 86). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be
30 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 nucleobases 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 oligomers 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., S₁ 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 nucleobases, 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
5 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
10 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
20 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
25 acid sequence of NOVX polypeptides whose sequences are provided in SEQ ID NOS:2n, wherein n is an integer between 1 and 86. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:2n, wherein n is an integer between 1 and 86, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

30 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
5 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
10 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
15 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
20 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
25 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
30 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 shown in SEQ ID NOS:2n, wherein n is an integer between 1 and 86) 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 shown SEQ ID NOS:2n, wherein n is an integer between 1 and 86. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86, and retains the functional activity of the protein of SEQ ID NOS:2n, wherein n is an integer between 1 and 86, 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 SEQ ID NOS:2n, wherein n is an integer between 1 and 86, and retains the functional activity of the NOVX proteins of SEQ ID NOS:2n, wherein n is an integer between 1 and 86.

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
5 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 shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86.

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
10 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.*,
15 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
20 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-
25 linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to A NOVX protein SEQ ID NOS:2n, wherein n is an integer between 1 and 86, 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
30 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.

5 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.

10 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.

15 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
20 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
25 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
30 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
5 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
10 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
15 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
20 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
25 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
30 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.

5

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
10 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
15 stranded portions from reformed duplexes by treatment with S₁ 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
20 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
25 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.
30

NOVX Antibodies

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
5 from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, 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
10 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
15 full length protein, such as an amino acid sequence shown in SEQ ID NOs: 2n, wherein n is an integer between 1 and 86, 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
20 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
25 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
30 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.

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.

10 **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

5 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
10 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
15 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
20 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
25 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
30 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.,
5 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
10 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
15 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
20 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.

25 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
30 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
5 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
10 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
15 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); Verhoeven 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
20 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
25 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)).

30

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
5 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,
10 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
15 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
20 (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
25 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
30 complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the XenomouseTM 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

5 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
10 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
15 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
20 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-
25 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
30 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).

5 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
10 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
15 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
20 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
25 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

30 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
 5 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
 10 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
 15 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
 20 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
 25 adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), 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
 30 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 as 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.

5 **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.

10 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
15 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).

20 **Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention**

Antibodies directed against a protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of the protein (e.g., for use in measuring levels of the protein within appropriate physiological samples, for use in diagnostic
25 methods, for use in imaging the protein, and the like). In a given embodiment, antibodies against the proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antigen binding domain, are utilized as pharmacologically-active compounds (see below).

An antibody specific for a protein of the invention can be used to isolate the protein by standard techniques, such as immunoaffinity chromatography or immunoprecipitation. Such
30 an antibody can facilitate the purification of the natural protein antigen from cells and of recombinantly produced antigen expressed in host cells. Moreover, such an antibody can be used to detect the antigenic protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic protein. Antibodies directed against the 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 ^3H .

Antibody Therapeutics

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may be 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-(methacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

5 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
 10 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
 15 DEPOTTM (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

20 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
 25 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,
 30 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. Christopoulos, 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-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.

15 **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 include 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.

10 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).

15 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.,
20 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 cerevisiae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30:
25 933-943), 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
30 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.

5 **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
10 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
15 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
20 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.

25 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 SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a
30 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 SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86), 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 NOS:2n-1, wherein n is an integer between 1 and 86, 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 then 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*
5 *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
10 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
15 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
20 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
25 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 G₀ 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
30 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
5 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,
10 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
15 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
20 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,
25 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
30 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.

- 5 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
10 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.

- 15 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
20 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
25 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
30 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 be 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 ^3H , 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
5 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
10 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
15 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
20 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
25 compound, and determining the ability of the test compound to interact with A NOVX protein, wherein determining the ability of the test compound to interact with A NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX
30 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
5 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
10 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
15 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[®],
20 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
25 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
30 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
5 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
10 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,
15 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
20 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
25 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
30 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 likely to be 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.

25 **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, SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, 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.

- 5 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
10 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
15 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
20 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
25 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
30 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 predicted coding sequences, such as those in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, 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

5 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

10 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

15 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").

20 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.

25

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

30 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 NOS:2n-1, wherein n is an integer between 1 and 86, 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.

10

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
5 system (*see*, Kwok, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Q β 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.

10 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
15 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
20 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
25 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
30 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 S₁ 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
5 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,
10 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
15 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,*
20 *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
25 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.

30 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
5 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
10 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
15 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

20 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 metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's
25 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.) 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
30 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.

- 5 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
- 10 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,
- 15 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

20 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

25 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

30 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 trials 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 trials 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 cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary Osteodystrophy, and other diseases, disorders and conditions of the like.

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

DISEASE 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) 5 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, 10 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 15 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 20 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.) 25 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 30 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
5 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
10 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
15 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*
20 *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
25 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 situations* in which NOVX is abnormally
30 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 including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, 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.

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: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

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.

5

Examples

EXAMPLE 1.

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

10

Table 1A. NOV1 Sequence Analysis			
	SEQ ID NO: 1	813 bp	
NOV1a, CG58548-01 DNA Sequence	CAAAACAAATTAAAAGATGAAGGAATACTATATCCATGTAACATGTGCCAATTTAACG AACGGTGGAAAGTCAGAACTTCTGAAATCAGGAAGCAGCAAATCCACACTAAAGCAC TATGGACAGAAAGCAGCAAAGACTTGTCTATCAGCCGACTCCTGTCACAGACTTTTCG TGGCAAAGAGAATGATACAGATTGACCTGAGATATGACACCCAGAACCTTATTCT GAGCAAGACCTCTGGGACTGGCTGAGGAACTCCACAGACCTTCAAGAGCCTCGGCCA GGGCCAAGAGAAGGCCATTGTTAAAACGGGCAAGTTTAAGAAAATGTTTGGATGGGG CGATTTTCATTCCAACATCAAAACAGTGAAGCTGAACCTGTTGATAACTGGGAAAATT GTAGATCATGGCAATGGGACATTTAGTGTTTATTTAGGCATAATTCAACTGGTCAAG GGAATGTATCTGTCAGCTTGGTACCCCTACAAAAATCGTGAATTTGACTTGGCACA ACAAACCGTGATTGATGCCAAAGATTCCAAGTCTTTTAATTGTCGCATTGAATATGAA AAGGTTGACAAGGCTACCAAGAACACACTCTGCAACTATGACCCTTCAAAAACCTGTT ACCAGGAGCAAACCCAAAGTCATGTATCCTGGCTCTGCTCCAAGCCCTTTAAGGTGAT CTGTATTTACATTTCTTTTATAGTACAGATTATAAACTGGTACAGAAAGTGTGCCCT GACTACAACCTACCACAGTGACACACCTTACTTTCCCTCGGGATGAAGGTGAACATGGG G		
	ORF Start: ATG at 17	ORF Stop: TGA at 797	
	SEQ ID NO: 2	260 aa	MW at 29905.5kD
NOV1a, CG58548-01 Protein Sequence	MKEYYIHVTCANLTNGGKSELLKSGSSKSTLKHIVTESSKDLSISRLLSQTFRGKEND TDLDLRYDTPEPYSEQDLWDWLRNSTDLQEPRPRAKRRIIVKTGKFKKMFGWGDFHSN IKTVKLNLLITGKIVDHNGTFSVYFRHNSTGQGNVSVSLVPPTKIVEFDLAQQTVID AKDSKSFNCRIEYEKVDKATKNTLCNYDPSKTCYQEQTQSHVSWLCSKPFKVICIYIS FYSTDYKLVQKVC PDYNYHSDTPYFPSG		
	SEQ ID NO: 3	771 bp	
NOV1b, 174307940 DNA Sequence	GGATCCGTAACATGTGCCAATTTAACGAACGGTGGAAAGTCAGAACTTCTGAAATCAG GAAGCAGCAAATCCACACTAAAGCACATATGGACAGAAAGCAGCAAAGACTTGTCTAT CAGCCGACTCCTGTCACAGACTTTTCGTGGCAAAGAGAATGATACAGATTGACCTG AGATATGACACCCAGAACCTTATTCTGAGCAAGACCTCTGGGACTGGCTGAGGAACT CCACAGACCTTCAAGAGCCTCGGCCAGGGCCAAGAGAAGGCCATTGTTAAAACGGG CAAGTTTAAGAAAATGTTTGGATGGGGCGATTTTCATTCCAACATCAAAACAGTGAAG CTGAACCTGTTGATAACTGGGAAAATTGTAGATCATGGCAATGGGACATTTAGTGT ATTTAGGCATAATTCAACTGGTCAAGGGAATGTATCTGTCAGCTTGGTACCCCTAC AAAAATCGTGAATTTGACTTGGCACAACAAACCGTGATTGATGCCAAAGATTCCAAG TCTTTTAATTGTCGCATTGAATATGAAAAGGTTGACAAGGCTACCAAGAACACACTCT GCAACTATGACCCTTCAAAAACCTGTTACCAGGAGCAAACCCAAAGTCATGTATCCTG GCTCTGCTCCAAGCCCTTTAAGGTGATCTGTATTTACATTTCTTTTATAGTACAGAT TATAAACTGGTACAGAAAGTGTGCCCTGACTACAACCTACCACAGTGACACACCTTACT		

	TTCCCTCGGGACTCGAG		
	ORF Start: GGA at 1	ORF Stop: E at 772	
	SEQ ID NO: 4	257 aa	MW at 29326.8kD
NOV1b, 174307940 Protein Sequence	GSVTCANLTNGGKSELLKSGSSKSTLKHIEWTESSKDLSISRLLSQTFRGKENDTDLDLRYDTPEPYSEQDLWDWLRNSTDLQEP RPRAKRRPIVKTGKFKMFGWGDFHSNIKTVKLNLLITGKIVDHGNGTFSVYFRHNSTGQGNVSVSLVPPTKIVEFDLAQQTVIDAKDSKSFNCRIEYKVDKATKNTLCNYDPSKTCYQEQTQSHVSWLCSKPKVICIYISFYSTDYKLVQKVC PDYNYHSDTPYFPSGLE		
	SEQ ID NO: 5	813 bp	
NOV1c, CG58548-02 DNA Sequence	CAAAACAAATTAAAAGATGAAGGAATACTATATCCATGTAACATGTGCCAATTTAACGAACGGTGGAAAGTCAGAACTTCTGAAATCAGGAAGCAGCAAATCCACACTAAAGCACATATGGACAGAAAGCAGCAAAGACTTGTCTATCAGCCGACTCCTGTCACAGACTTTTCGTGGCAAAGAGAATGATACAGATTTGGACCTGAGATATGACACCCCAAGAACCTTATTCTGAGCAAGACCTCTGGGACTGGCTGAGGAACCTCACAGACCTTCAAGAGCCTCGGCCCAAGGCAAGAGAAGGCCCATTTGTTAAACGGGCAAGTTTAAGAAAATGTTTGGATGGGGCGATTTTTCATTCCAACATCAAAACAGTGAAGCTGAACCTGTTGATAACTGGGAAAATTGTAGATCATGGCAATGGGACATTTAGTGTTTATTTT CAGGCATAATTCAACTGGTCAAGGGAATGTATCTGT CAGCTTGGTACCCCTACAAAAATCGTGGAATTTGACTTGGCACAACAAACCGTGATTGATGCCAAAGATTCCAAGTCTTTTAATTGTCGCATTGAATATGAA AAGGTTGACAAGGCTACCAAGAACACACTCTGCAACTATGACCCTTCAAAAACCTGTTACCAGGAGCAAACCCAAAGTCATGTATCCTGGCTCTGCTCCAAGCCCTTTAAGGTGATCTGTATTTACATTTCTTTTATAGTACAGATTATAAACTGGTACAGAAAGTGTGCCCTGACTACAACTACCACAGTGACACACCTTACTTTCCCTCGGGATGAAGGTGAACATGGGG		
	ORF Start: ATG at 17	ORF Stop: TGA at 797	
	SEQ ID NO: 6	260 aa	MW at 29905.5kD
NOV1c, CG58548-02 Protein Sequence	MKEYYIHVTCANLTNGGKSELLKSGSSKSTLKHIEWTESSKDLSISRLLSQTFRGKENDTDLDLRYDTPEPYSEQDLWDWLRNSTDLQEP RPRAKRRPIVKTGKFKMFGWGDFHSN IKTVKLNLLITGKIVDHGNGTFSVYFRHNSTGQGNVSVSLVPPTKIVEFDLAQQTVIDAKDSKSFNCRIEYKVDKATKNTLCNYDPSKTCYQEQTQSHVSWLCSKPKVICIYISFYSTDYKLVQKVC PDYNYHSDTPYFPSG		
	SEQ ID NO: 7	627 bp	
NOV1d, CG58548-03 DNA Sequence	CAAAACAAATTAAAAGATGAAGGAATACTATATCCATGTAACATGTGCCAATTTAACGAACGGTGGAAAGTCAGAACTTCTGAAATCAGGAAGCAGCAAATCCACACTAAAGCACATATGGACAGAAAGCAGCAAAGACTTGTCTATCAGCCGACTCCTGTCACAGACTTTTCGTGGCAAAGAGAATGATACAGATTTGAACCTGTTGATAACTGGGAAAATTGTAGATCATGGCAATGGGACATTTAGTGTTTATTTT CAGGCATAATTCAACTGGTCAAGGGAATGTATCTGT CAGCTTGGTACCCCTACAAAAATCGTGGAATTTGACTTGGCACAACAAACCGTGATTGATGCCAAAGATTCCAAGTCTTTTAATTGTCGCATTGAATATGAAAAGGTTGACAAGGCTACCAAGAACACACTCTGCAACTATGACCCTTCAAAAACCTGTTACCAGGAGCAAACCCAAAGTCATGTATCCTGGCTCTGCTCCAAGCCCTTAAGGTGATCTGTATTTACATTTCTTTTATAGTACAGATTATAAACTGGTACAGAAAGTGTGCCCTGACTACAAC TACCACAGTGACACACCTTACTTTCCCTCGGGATGAAGGTGAACATG		
	ORF Start: ATG at 17	ORF Stop: TGA at 614	
	SEQ ID NO: 8	199 aa	MW at 22496.1kD
NOV1d, CG58548-03 Protein Sequence	MKEYYIHVTCANLTNGGKSELLKSGSSKSTLKHIEWTESSKDLSISRLLSQTFRGKENDTDLNLLITGKIVDHGNGTFSVYFRHNSTGQGNVSVSLVPPTKIVEFDLAQQTVIDAKDSKSFNCRIEYKVDKATKNTLCNYDPSKTCYQEQTQSHVSWLCSKPLKVICIYISFYSTDYKLVQKVC PDYNYHSDTPYFPSG		

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	8..260 3..255	236/253 (93%) 236/253 (93%)
NOV1c	1..260 1..260	243/260 (93%) 243/260 (93%)
NOV1d	1..260 1..199	161/260 (61%) 164/260 (62%)

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

Table 1C. Protein Sequence Properties NOV1a	
PSort analysis:	0.5297 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV1a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
5 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
ABB11858	Human neurexophilin homologue, SEQ ID NO:2228 - Homo sapiens, 305 aa. [WO200157188-A2, 09-AUG-2001]	8..260 53..305	253/253 (100%) 253/253 (100%)	e-152
AAB43066	Human ORFX ORF2830 polypeptide sequence SEQ ID NO:5660 - Homo sapiens, 253 aa. [WO200058473-A2, 05-OCT-2000]	8..260 1..253	253/253 (100%) 253/253 (100%)	e-152
AAM57924	Human brain expressed single exon probe encoded protein SEQ ID NO: 30029 - Homo sapiens, 235 aa. [WO200157275-A2, 09-AUG-2001]	14..248 1..235	235/235 (100%) 235/235 (100%)	e-140
AAB28778				5e-73

	fragment encoded by gene 45 - Homo sapiens, 128 aa. [WO200055198-A1, 21-SEP-2000]	1..128	128/128 (100%)	
AAB28779	Protein fragment encoded by gene 45 - Homo sapiens, 128 aa. [WO200055198-A1, 21-SEP-2000]	104..231 1..128	127/128 (99%) 127/128 (99%)	4e-72

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
P58417	Neurexophilin 1 precursor - Homo sapiens (Human), 271 aa.	8..260 19..271	253/253 (100%) 253/253 (100%)	e-151
Q61200	Neurexophilin 1 precursor - Mus musculus (Mouse), 253 aa (fragment).	8..260 1..253	253/253 (100%) 253/253 (100%)	e-151
Q63366	Neurexophilin 1 precursor (Neurophilin) - Rattus norvegicus (Rat), 271 aa.	8..260 19..271	251/253 (99%) 252/253 (99%)	e-150
O95156	Neurexophilin 2 precursor - Homo sapiens (Human), 262 aa (fragment).	72..260 74..262	153/189 (80%) 170/189 (88%)	3e-93
Q28145	Neurexophilin 2 precursor (Neurophilin) - Bos taurus (Bovine), 264 aa.	72..260 76..264	153/189 (80%) 170/189 (88%)	4e-93

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
No Significant Matches Found			

5

EXAMPLE 2.

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

Table 2A. NOV2 Sequence Analysis			
	SEQ ID NO: 9	796 bp	
NOV2a, CG58542-01 DNA Sequence	AGGAGGAAGATGCAACTGACTCGCTGCTGCTTCGTGTTCTTGGTGCAGGGTAGCCTCT ATCTGGTCATCTGTGGCCAGGATGATGGTCCTCCCGGCTCAGAGGACCCTGAGCGTGA TGACCACGAGGGCCAGCCCCGGCCCCGGGTGCCTCGGAAGCGGGGCCACATCTCACCT AAGTCCCGCCCCATGGCCAATTCCACTCTCCTAGGGCTGCTGGCCCCGCCTGGGGAGG CTTGGGGCATTCTTGGGCAGCCCCCAACCGCCCGAACCACAGCCCCCACCCTCAGC CAAGGTGAAGAAAATCTTTGGCTGGGGCGACTTCTACTCCAACATCAAGACGGTGGCC CTGAACCTGCTCGTCACAGGGAAGATTGTGGACCATGGCAATGGGACCTTCAGCGTCC ACTTCCAACACAATGCCACAGGCCAGGGAAACATCTCCATCAGCCTCGTGCCCCCAG TAAAGCTGTAGAGTTCCACCAGGAACAGCAGATCTTCATCGAAGCCAAGGCCTCCAAA ATCTTCAACTGCCGGATGGAGTGGGAGAAGGTAGAACGGGGCCGCGGACCTCGCTTT GCACCCACGACCCAGCCAAGATCTGCTCCCGAGACCACGCTCAGAGCTCAGCCACCTG GAGCTGCTCCCAGCCCTTCAAAGTCGTCTGTGTCTACATCGCCTTCTACAGCACGGAC TATCGGCTGGTCCAGAAGGTGTGCCAGATTACAACCTACCATAGTGATACCCCCTACT ACCCATCTGGGTGACCCGGGGCAGGCCACAGAGGCCAGGCCA		
	ORF Start: ATG at 10	ORF Stop: TGA at 766	
	SEQ ID NO: 10	252 aa	MW at 28126.7kD
NOV2a, CG58542-01 Protein Sequence	MQLTRCCFVFLVQGSYLVLICGQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSR PMANSTLLGLLAPPGEAWGILGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNL LVTGKIVDHNGTFSVHFQHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFN CRMWEKVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPFKVVCVYIAFYSTDYRL VQKVCPDYNHSDTPYYPSG		
	SEQ ID NO: 11	702 bp	
NOV2b, 169679583 DNA Sequence	GGATCCCAGGATGATGGTCCTCCCGGCTCAGAGGACCCTGAGCGTGATGACCACGAGG GCCAGCCCCGGCCCCGGGTGCCTCGGAAGCGGGGCCACATCTCACCTAAGTCCCGCCC CATGGCCAATTCCACTCTCCTAGGGCTGCTGGCCCCGCCTGGGGAGGCTTGGGGCATT CTTGGGCAGCCCCCAACCGCCCGAACCACAGCCCCCACCCTCAGCCAAGGTGAAGA AAATCTTTGGCTGGGGCGACTTCTACTCCAACATCAAGACGGTGGCCCTGAACCTGCT CGTCACAGGGAAGATTGTGGACCATGGCAATGGGACCTTCAGCGTCCACTTCCAACAC AATGCCACAGGCCAGGGAAACATCTCCATCAGCCTCGTGCCCCCAGTAAAGCTGTAG AGTTCCACCAGGAACAGCAGATCTTCATCGAAGCCAAGGCCTCCAAAATCTTCAACTG CCGGATGGAGTGGGAGAAGGTAGAACGGGGCCGCGGACCTCGCTCTGCACCCACGAC CCAGCCAAGATCTGCTCCCGAGACCACGCTCAGAGCTCAGCCACCTGGAGCTGCTCCC AGCCCTTCAAAGTCGTCTGTGTCTACATCGCCTTCTACAGCACGGACTATCGGCTGGT CCAGAAGGTGTGCCAGATTACAACCTACCATAGTGATACCCCCTACTACCCATCTGGG CTCGAG		
	ORF Start: GGA at 1	ORF Stop: 5 at 703	
	SEQ ID NO: 12	234 aa	MW at 26037.0kD
NOV2b, 169679583 Protein Sequence	GSQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGEAWGI LGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTGKIVDHNGTFSVHFQH NATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWEKVERGRRTSLCTHD PAKICSRDHAQSSATWSCSQPFKVVCVYIAFYSTDYRLVQKVCPDYNHSDTPYYPSG LE		
	SEQ ID NO: 13	702 bp	
NOV2c, 169679634 DNA Sequence	GGATCCCAGGATGATGGTCCTCCCGGCTCAGAGGACCCTGAGCGTGATGACCACGAGG GCCAGCCCCGGCCCCGGGTGCCTCGGAAGCGGGGCCACATCTCACCTAAGTCCCGCCC CATGGCCAATTCCACTCTCCTAGGGCTGCTGGCCCCGCCTGGGGAGGCTTGGGGCATT CTTGGGCAGCCCCCAACCGCCCGAACCACAGCCCCCACCCTCAGCCAAGGTGAAGA AAATCTTTGGCTGGGGCGACTTCTACTCCAACATCAAGACGGTGGCCCTGAACCTGCT CGTCACAGGGAAGATTGTGGACCATGGCAATGGGACCTTCAGCGTCCACTTCCAACAC AATGCCACAGGCCAGGGAAACATCTCCATCAGCCTCGTGCCCCCAGTAAAGCTGTAG AGTTCCACCAGGAACAGCAGATCTTCATCGAAGCCAAGGCCTCCAAAATCTTCAACTG CCGGATGGAGTGGGAGAAGGTAGAACGGGGCCGCGGACCTCGCTTTGCACCCACGAC		

	CCAGCCAAGATCTGCTCCCGAGACCACGCTCAGAGCTCAGCCACCTGGAGCTGCTCCC AGCCCTTCAAAGTCGTCTGTGTCTACATCGCCTTCTACAGCACGGACTATCGGCTGGT CCAGAAGGTGTGCCAGATTACAACCTACCATAGTGATACCCCTACTACCCATCTGGG CTCGAG		
	ORF Start: GGA at 1	ORF Stop: 5 at 703	
	SEQ ID NO: 14	234 aa	MW at 26037.0kD
NOV2c, 169679634 Protein Sequence	GSQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGEAWGI LGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTGKIVDHGNGTFSVHFQH NATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWEKVERGRRTSLCTHD PAKICSRDHAQSSATWSCSQPFKVVCVYIAFYSTDYRLVQKVCPDYNHSDTPYYPSG LE		

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

Table 2B. Comparison of NOV2a against NOV2b through NOV2c.		
Protein Sequence	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV2b	23..252	218/230 (94%)
	3..232	218/230 (94%)
NOV2c	23..252	218/230 (94%)
	3..232	218/230 (94%)

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

Table 2C. Protein Sequence Properties NOV2a	
PSort analysis:	0.7666 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 23 and 24

- 5 A search of the NOV2a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, 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
AAU29174	Human PRO polypeptide sequence #151 - Homo sapiens, 252 aa. [WO200168848-A2, 20-SEP-2001]	1..252 1..252	252/252 (100%) 252/252 (100%)	e-154

AAM39340	Human polypeptide SEQ ID NO 2485 - Homo sapiens, 252 aa. [WO200153312-A1, 26-JUL-2001]	1..252 1..252	252/252 (100%) 252/252 (100%)	e-154
AAB87571	Human PRO1327 - Homo sapiens, 252 aa. [WO200116318-A2, 08-MAR-2001]	1..252 1..252	252/252 (100%) 252/252 (100%)	e-154
AAB66150	Protein of the invention #62 - Unidentified, 252 aa. [WO200078961-A1, 28-DEC-2000]	1..252 1..252	252/252 (100%) 252/252 (100%)	e-154
AAV99401	Human PRO1327 (UNQ687) amino acid sequence SEQ ID NO:218 - Homo sapiens, 252 aa. [WO200012708-A2, 09-MAR-2000]	1..252 1..252	252/252 (100%) 252/252 (100%)	e-154

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
Q91VX5	SIMILAR TO NEUREXOPHILIN 3 - Mus musculus (Mouse), 252 aa.	1..252 1..252	243/252 (96%) 246/252 (97%)	e-148
Q9Z2N5	Neurexophilin 3 precursor - Rattus norvegicus (Rat), 252 aa.	1..252 1..252	242/252 (96%) 246/252 (97%)	e-148
O95157	Neurexophilin 3 - Homo sapiens (Human), 221 aa (fragment).	32..252 1..221	221/221 (100%) 221/221 (100%)	e-134
P58417	Neurexophilin 1 precursor - Homo sapiens (Human), 271 aa.	79..252 97..271	114/175 (65%) 143/175 (81%)	7e-68
Q63366	Neurexophilin 1 precursor (Neurophilin) - Rattus norvegicus (Rat), 271 aa.	79..252 97..271	114/175 (65%) 143/175 (81%)	7e-68

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
No Significant Matches Found			

EXAMPLE 3.

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

Table 3A. NOV3 Sequence Analysis			
	SEQ ID NO: 15	1173 bp	
NOV3a, CG58540-01 DNA Sequence	GCCCTGCATCATGGAACTCTTTCTAATGCAAGTGGTACTTTTGCCATACGCCTTTTA AAGATACTGTGTCAAGATAACCCTTCGCACAACGTGTTCTGTTCTCCTGTGAGCATCT CCTCTGCCCTGGCCATGGTTCTCCTAGGGGCAAAGGGAAACACCGCAACCCAGATGGC CCAGATAGAGTCTCTGCTCTGTCACCCAGGCTGGAGTGCAGACATTTCATCGGGCTTTC CAGTCGCTTCTCACTGAAGTGAACAAGGCTGGCACACAGTACCTGCTGAGAACGGCCA ACAGGCTCTTTGGAGAGAAAACCTTGTCAGTTCCTCTCAACGTTTAAAGGAATCCTGTCT TCAATTCTACCATGCTGAGCTGAAGGAGCTTTCCTTTATCAGAGCTGCAGAAGAGTCC AGGAAACACATCAACACCTGGGTCTCAAAAAGACCGAAGGTAAAATTGAAGAGTTGT TGCCGGGTAGCTCAATTGATGCAGAAACCAGGCTGGTTCTTGTGAATGCTGTCTATTT CAGAGGAAACTGGGATGAACAGTTTGACAAGGAGAACACCGAGGAGAGACTGTTTAAA GTCAGCAAGGCGAGTAAGGAGGAGAAACCTGTGCAAATGATGTTTAAAGCAATCTACTT TTAAGAAGACCTATATAGGAGAAATATTTACCCAAATCTTGGTGCTTCCATATGTTGG CAAGGAACTGAATATGATCATCATGCTTCCGGACGAGACCACTGACTTGAGAACGGTG GAAAAAAGTCTCACTTTTGAGAACTCACAGCCTGGACCAAGCCAGACTGTATGAAGA GTACTGAGGTTGAAGTTCTCCTTCCAAAATTTAACTACAAGAGGATTATGACATGGA ATCTGTGCTTCGGCATTGGAATTGTTGATGCCTTCCAACAGGGCAAGGCTGACTTG TCGGCAATGTCAGCGGAGAGAGACCTGTGTCTGTCCAAGTTCGTGCACAAGAGTTTGT TGGAGGTGAATGAAGAAGGCACCGAGGCAGCGGCAGCGTCGAGCTGCTTTGTAGTTGC AGAGTGCTGCATGGAATCTGGCCCCAGGTTCTGTGCTGACCACCCTTTCCTTTTCTTC ATCAGGCACAACAGAGCCAACAGCATTCTGTTCTGTGGCAGGTTCTCATCGCCATAAA GGGTGCACTTACC		
	ORF Start: ATG at 11	ORF Stop: TAA at 1157	
	SEQ ID NO: 16	382 aa	MW at 43163.1kD
NOV3a, CG58540-01 Protein Sequence	METLSNASGTFAIRLLKILCQDNPSHNVFCSPVSISSALAMVLLGAKGNTATQMAQIE SLLCHPGWSADIHRAFQSLLEVNKAGTQYLLRTANRLFGEKTCQFLSTFKESCLQFY HAELKELSFIRAAEESRKHINTWVSKKTEGKIEELLPGSSIDAETRLVLVNAVYFRGN WDEQFDKENTEERLKFVSKASKEEKPVQMMFKQSTFKKTYIGEIFTQILVLPYVGKEL NMIIMLPDETTDLRTVEKSLTFEKLTAWTKPDCMKSTEVEVLLPKFKLQEDYDMESVL RHLGIVDAFQQGKADLSAMSAERDLCLSKFVHKSFVEVNEEGTEAAAASSCFVVAECC MESGPRFCADHPFLFFIRHNRANSILFCGRFSSP		

Further analysis of the NOV3a protein yielded the following properties shown in Table 5 3B.

Table 3B. Protein Sequence Properties NOV3a	
PSort analysis:	0.6881 probability located in mitochondrial inner membrane; 0.6500 probability located in plasma membrane; 0.3773 probability located in mitochondrial intermembrane space; 0.3157 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 3C. Geneseq Results for NOV3a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY55841	Human cytoplasmic antiproteinase-3 protein (CAP-3) - Homo sapiens, 376 aa. [WO9957273-A2, 11-NOV-1999]	1..382 1..376	328/382 (85%) 353/382 (91%)	0.0
AAR99254	Cytoplasmic antiproteinase-3 protein - Homo sapiens, 376 aa. [WO9624650-A2, 15-AUG-1996]	1..382 1..376	328/382 (85%) 353/382 (91%)	0.0
AAU30834	Novel human secreted protein #1325 - Homo sapiens, 566 aa. [WO200179449-A2, 25-OCT-2001]	1..382 191..566	324/382 (84%) 351/382 (91%)	0.0
AAB11125	Human thrombin inhibitor protein - Homo sapiens, 376 aa. [US6133422-A, 17-OCT-2000]	1..382 1..376	279/382 (73%) 314/382 (82%)	e-153
AAB59176	Thrombin inhibitor protein - Unidentified, 376 aa. [US6156540-A, 05-DEC-2000]	1..382 1..376	279/382 (73%) 314/382 (82%)	e-153

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 3D.

Table 3D. Public BLASTP Results for NOV3a				
Protein Accession Number	Protein/Organism/Length	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P50453	Cytoplasmic antiproteinase 3 (CAP3) (CAP-3) (Protease inhibitor 9) (Serpin B9) - Homo sapiens (Human), 376 aa.	1..382 1..376	328/382 (85%) 353/382 (91%)	0.0
Q96J44	SERINE (OR CYSTEINE) PROTEINASE INHIBITOR, CLADE B (OVALBUMIN), MEMBER 6 - Homo sapiens (Human), 376 aa.	1..382 1..376	279/382 (73%) 314/382 (82%)	e-153
P35237	Placental thrombin inhibitor (Cytoplasmic antiproteinase) (CAP) (Protease inhibitor 6) - Homo sapiens (Human), 376 aa.	1..382 1..376	278/382 (72%) 312/382 (80%)	e-152
O02739	Serine proteinase inhibitor B-43 - Bos taurus (Bovine), 378 aa.	1..382 1..378	252/382 (65%) 303/382 (78%)	e-139
Q60854				e-136

	(SERINE (OR CYSTEINE) PROTEINASE INHIBITOR, CLADE B (OVALBUMIN), MEMBER 6) - Mus musculus (Mouse), 378 aa.	1..378	301/383 (78%)	
--	---	--------	---------------	--

PFam analysis predicts that the NOV3a protein contains the domains shown in the Table 3E.

Table 3E. Domain Analysis of NOV3a			
Pfam Domain	NOV3a Match Region	Identities/ Similarities for the Matched Region	Expect Value
serpin: domain 1 of 1	1..382	170/400 (42%) 314/400 (78%)	8.8e-159

EXAMPLE 4.

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

Table 4A. NOV4 Sequence Analysis			
	SEQ ID NO: 17	502 bp	
NOV4a, CG56340-03 DNA Sequence	GCAATATTGGCAACATCCCAATGGCCCTGTCCTTTTCTTTACTGATGGCCGTGCTGGT GCTCAGCTACAAATCCATCTGTTCTCTGGGCTGTGATCTGCCTCAGACCCACAGCCTG GGTAATAGGAGGGCCTTGATACTCCTGGCACAAATGGGAAGAATCTCTCCTTTCTCCT GCCTGAAGGACAGACATGACTTTGGATTCCCCCAGGAGGAGTTTGATGGCAACCAGTT CCAGAAGGCTCAAGCCATCTCTGTCCTCCATGAGATGATCCAGCAGACCTTCAATCTC TTCAGCACAAAGGACTCATCTGCTACTTGGGAACAGAGCCTCCTAGAAAAATTTTCCA CTGAACTTAACCAGCAGCTGACAGAGAAGAAATACAGCCCTTGTGCCTGGGAGGTTGT CAGAGCAGAAATCATGAGATCCTTCTCTTTATCAAAAATTTTTC AAGAAAGATTAAGG AGGAAGGAATGAAACCTGTTTCAACATGGAAATGATCT		
	ORF Start: ATG at 21	ORF Stop: TGA at 474	
	SEQ ID NO: 18	151 aa	MW at 17402.8kD
NOV4a, CG56340-03 Protein Sequence	MALSFSLMAVLVLSYKSICSLGCDLPQTHSLGNRRALILLAQMGRISPFSCCLKDRHD FGFPQEEFDGNQFQKAQAISVLHEMIQQTFFNLSTKDSSATWEQSLLEKFSTELNQQL TEKKYSPCAWEVVRAEIMRSFSLSKIFQERLRRKE		
	SEQ ID NO: 19	396 bp	
NOV4b, 174308150 DNA Sequence	GGATCCTGTGATCTGCCTCAGACCCACAGCCTGGGTAATAGGAGGGCCTTGATACTCC TGGCACAAATGGGAAGAATCTCTCCTTTCTCCTGCCTGAAGGACAGACATGACTTTGG ATTCCCCCAGGAGGAGTTTGATGGCAACCAGTTCCAGAAGGCTCAAGCCATCTCTGTC CTCCATGAGATGATCCAGCAGACCTTCAATCTCTTCAGCACAAAGGACTCATCTGCTA CTTGGGAACAGAGCCTCCTAGAAAAATTTTCCACTGAACTTAACCAGCAGCTGACAGA GAAGAAATACAGCCCTTGTGCCTGGGAGGTTGTGAGAGCAGAAATCATGAGATCCTTC TCTTTATCAAAAATTTTTC AAGAAAGATTAAGGAGGAAGGAACTCGAG		
	ORF Start: GGA at 1	ORF Stop: at 397	
	SEQ ID NO: 20	132 aa	MW at 15360.3kD
NOV4b, 174308150 Protein	GSCDLPQTHSLGNRRALILLAQMGRISPFSCCLKDRHDFGFPQEEFDGNQFQKAQAISV LHEMIQQTFFNLSTKDSSATWEQSLLEKFSTELNQQLTEKKYSPCAWEVVRAEIMRSF SLSKIFQERLRRKELE		

Sequence	
----------	--

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

Table 4B. Comparison of NOV4a against NOV4b and NOV4c.		
Protein Sequence	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV4b	24..151 3..130	128/128 (100%) 128/128 (100%)

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

Table 4C. Protein Sequence Properties NOV4a	
PSort analysis:	0.5231 probability located in outside; 0.1317 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 24 and 25

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

Table 4D. Geneseq Results for NOV4a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAP20108	Sequence encoded by leukocyte interferon LeIF F cDNA - Homo sapiens, 189 aa. [GB2079291-A, 20-JAN-1982]	1..151 1..189	151/189 (79%) 151/189 (79%)	7e-77
AAP40123	Sequence encoded by the cDNA insert of the recombinant plasmid CG-pBR 322/HLycIFN-1'b - Homo sapiens, 189 aa. [EP100561-A, 15-FEB-1984]	1..151 1..189	150/189 (79%) 150/189 (79%)	3e-76
AAP30179	Sequence of a polypeptide with human lymphoblastoid interferon activity encoded by plasmid CG-pBR 322/HL gamma cIFN-1'b - Homo sapiens, 189 aa. [EP76489-A, 13-APR-1983]	1..151 1..189	150/189 (79%) 150/189 (79%)	3e-76

AAB49780	Human interferon alpha-f amnio acid sequence - Homo sapiens, 189 aa. [WO200107608-A1, 01-FEB-2001]	1..151 1..189	141/189 (74%) 145/189 (76%)	2e-71
AAR62368	Interferon alpha consensus sequence - Synthetic, 187 aa. [WO9420122-A, 15-SEP-1994]	1..151 1..187	139/189 (73%) 144/189 (75%)	7e-69

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
E968396	ARTIFICIAL SEQUENCE FOR CDNA INSERT OF RECOMBINANT PLASMID CG-PBR 322/HLYCIFN-1'B - vectors, 189 aa.	1..151 1..189	151/189 (79%) 151/189 (79%)	3e-76
E968985	POLYPEPTIDE FOR THE USE OF IMMUNOMODULATOR, ANTI-TUMOR-AGENT - vectors, 189 aa.	1..151 1..189	151/189 (79%) 151/189 (79%)	3e-76
P01568	Interferon alpha-21 precursor (Interferon alpha-F) (LeIF F) - Homo sapiens (Human), 189 aa.	1..151 1..189	151/189 (79%) 151/189 (79%)	3e-76
CAA00629	ARTIFICIAL SEQUENCE FOR CDNA INSERT OF RECOMBINANT PLASMID CG-PBR 322/HLYCIFN-1'B - synthetic construct, 189 aa.	1..151 1..189	150/189 (79%) 150/189 (79%)	1e-75
Q14608	LEUKOCYTE INTERFERON-ALPHA - Homo sapiens (Human), 181 aa.	9..151 1..181	143/181 (79%) 143/181 (79%)	3e-72

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
interferon: domain 1 of 2	1..115	81/116 (70%) 109/116 (94%)	4.9e-71

interferon: domain 2 of 2	116..151	27/36 (75%) 33/36 (92%)	1e-19
---------------------------	----------	----------------------------	-------

EXAMPLE 5.

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

Table 5A. NOV5 Sequence Analysis			
	SEQ ID NO: 21	203 bp	
NOV5a, CG58514-01 DNA Sequence	ACCTCTTTGCCACCATA <u>CCATGA</u> AGGTATGCGTGATTGTCCTGTCTCTCCTCGTGATA ATAGCCGCCTTCTGCTCTGTAGCACTCTCAGCACCGAATTCCAAACCAAAGAGGCAA GCAAGTCTGCGCTGACCCAGTGAGTCCTGGGTCCAGGAGTACGTGTATGACCTGGAA <u>CTGA</u> ACTGAGCTGCTCAGAGACAGGAAGT		
	ORF Start: ATG at 20	ORF Stop: TGA at 176	
	SEQ ID NO: 22	52 aa	MW at 5408.4kD
NOV5a, CG58514-01 Protein Sequence	MKVCVIVLSLLVIIA AFCSVALSAPNSKPKEASKSALTPVSPGSRSTCMTWN		

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

Table 5B. Protein Sequence Properties NOV5a	
PSort analysis:	0.8200 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Likely cleavage site between residues 24 and 25

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

Table 5C. Geneseq Results for NOV5a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found				

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 5D.

Table 5D. Public BLASTP Results for NOV5a

Protein Accession Number	Protein/Organism/Length	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
B60407	monocyte adherence-induced protein 5 alpha - human, 52 aa.	1..52 1..52	43/52 (82%) 48/52 (91%)	4e-19

PFam analysis predicts that the NOV5a protein contains the domains shown in the Table 5E.

Table 5E. Domain Analysis of NOV5a			
Pfam Domain	NOV5a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

EXAMPLE 6.

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

Table 6A. NOV6 Sequence Analysis		
	SEQ ID NO: 23	2305 bp
NOV6a, CG57887-01 DNA Sequence	ATTTTTCCTCGGCTGCCGGCGGCTCCGACATCATGCTCCGGCTCCTCCGGCCGCTGCTGCTACTGCTGCTGCTGCCCTCCCCGGGGTCCCCTGAGCCCCCGGCCTGACCCAGCTGTCCCCGGGGCGCCCCCGCAGGCCCGGACTTGCTCTACGCTGACGGGCTGCGCGCCTACGCGGCCGGGGCTTGGGCGCCGGCCGTGGCGCTGCTGCGGGAGGCGCTGCGGAGCCAGGCGGCTGGGCCGGGTGCGGCTGGATTGCGGGGCGAGCTGCGCGGCCGATCCGGCGCCGCGCTCCCCGCGTGCTTCTCGGGGCCCGGAGCCCGACTCCGGGCCGGGACCCACGCAGGGGTCTCTGGGAGCGACAGCTTCTCCGTGCAGCGCTCCGCCGCGCAGACTGCCTGACCCAGTGCGCAGCACGGAGGCTGGGCCCGGGGGCGCGGCGCGGCTTCGCGTG GGGAGCGCGCTCCGGGACGCCTTCGCGCGTCGGGAGCCCTACAACCTGCAGAGGG CCTATTACCAGTTGAAGAAGCTGGATCTGGCAGCTGCGGCAGCACACACCTTCTTTGT AGCAAACCCCATGCACCTGCAGATGCGGGAGGACATGGCTAAGTACAGACGAATGTGCGGAGTTCGGCCCCAGAGCTTCCGGGACCTGGAGACGCCCCACACTGGGCAGCCTATGACACTGGCCTGGAGCTACTGGGGCGCCAGGAGGCAGGACTGGCACTGCCAGGCTAGAGGAGCTCTTCAGGGGAGCCTGGCCAGATGGAGAGCTGCCGTGCTGACTGTGAGGGG CCTGAGGAGCAGCAGGGGGCTGAAGAAGAGGAGGATGGGGCTGCGAGCCAGGGGGGCC TCTATGAGGCCATTGCAGGACACTGGATTGAGTCTGCGAGTGCCGGCAACGCTGTGTGGGGGAAGCAGCCACAGCCCTGGTCGCAGCTTCCCTGTCCAGACTTCCTTCCCAAC CAGCTGAGGCGGCTACATGAGGCCCATGCTCAGGTGGGCAATCTGTCCAGGCTATAGAAAATGTCCTGAGTGTCTGCTCTTCTACCCGGAGGATGAGGCTGCCAAGAGGGCTCTGAACCACTACCAGGCCAGCTGGGAGAGCCGAGACCTGGCCTCGGACCCAGAGAGGACATCCAGCGCTTCATCCTCCGATCCCTGGGGGAGAAGAGGCAGCTCTACTATGCCATGGAGCACCTGGGGACCAGCTTCAAGGATCCTGACCCCTGGACCCCTGCAGCTCTCATCCCTGAGGCACTTAGAGAAAAGCTCAGAGAGGATCAAGAGAAGAGGCCTTGGGACCATGAGCCCGTGAAGCCAAAGCCCTTGACCTACTGGAAGGATGTCCTTCTCCTGGAGGGTGTGACCTTGACCCAGGATTCCAGGCAGCTGAATGGGTGCGAGCGGGCGGTGTTGGATGGGCTGCTCACCCCAGCCGAGTGTGGGGTGCTGCTGCAGCTGGCTAAGGATGCAGCTGGGGCTGGAGCCAGGTCTGGCTATCGTGGTGGCGCTCCCTCACACCCCCCATGAACGCTTCGAGGGGCTCACGGTGCTTAAGGCTGCGCAGCTGGCCCGGGCTGGGACAGTGGGCAGTCA GGGTGCTAAGCTGCTTCTGGAGGTGAGCGAGCGGGTGCGGACCTTGACCCAGGCCTACTTCTCCCCGGAACGGCCCCCTGCATCTGTCTTACCCACCTGGTGTGCCGAGCGCCA	

	TAGAAGGAGAGCAAGAGCAGCGCATGGACCTGAGTCACCCAGTGCACGCAGACAACTG CGTCCTGGACCCTGACACGGGAGAGTGCTGGCGGGAGCCCCAGCCTACACCTATCGG GACTACAGCGGACTCCTCTACCTCAACGATGACTTCCAGGGTGGGGACCTGTTCTTCA CGGAGCCCAACGCCCTCACTGTCACGGCTCGGGTGCCTCGCTGTGGGCGCCTTGT GGCCTTCAGCTCCGGTGTGAGAATCCCCATGGGGTGTGGGCCGTGACTCGGGGACGG CGCTGTGCCCTGGCACTGTGGCACACGTGGGCACCTGAGCACAGGGAGCAGGAGTGGA TAGAAGCCAAAGAACTGCTGCAGGAGTCACAGGAGGAGGAGGAAGAGGAAGAGGAAGA AATGCCCAGCAAAGACCCTTCCCCAGAGCCCCCTAGCCGCAGGCACCAGAGGGTCCAA GACAAGACTGGAAGGGCACCTCGGGTTCGGGAGGAGCTGTGAGTGGCTGAGCCAGCTC CTTGAGGATGTGGCCACTTGACTTGTGGAAGGCCATCTTGATG		
	ORF Start: ATG at 36	ORF Stop: TGA at 2244	
	SEQ ID NO: 24	736 aa	MW at 81805.5kD
NOV6a, CG57887-01 Protein Sequence	MLRLLRPLLLLLLLLPPPGSPEPPGLTQLSPGAPPQAPDLLYADGLRAYAAGAWAPAVA LLREALRSQAALGRVRLDCGASCAADPGAALPAVLLGAPEPDSGPGPTQGSWERQLLR AALRRADCLTQCAARRLGPGGAARLRVGSALRDAFRREPYNYLQRAYYQLKKLDLAA AAAHTFFVANPMHLQMREDMAKYRRMSGVRPQSFRLDLETPPHWAAYDTGLELLGRQEA GLALPRLEEALQGSQAQMESCRADCEGPEEQQAEEEEEDGAASQGGLYEAIAGHWIQV LQCRQRCVGEAATRPGRSFPVPDFLPNQLRRLHEAHAQVGNLSQAIENVLSVLLFYPE DEAAKRALNQYQAQLGEPRPGLGPREDIQRFILRSLGEKRQLYAMEHLGTSFKDPDP WTPAALIPEALREKLREDQEKRPWDHEPVKPKPLTYWKDVLLLEGVTLTQDSRQLNGS ERAVLDGLLTPAECGVLLQLAKDAAGAGARSGYRGRSPHTPHERFEGLTVLKAQLA RAGTVGSQGAQLLLEVSRVRTLTQAYFSPERPLHLSFTHLVCRSAIEGEQEQRMDLS HPVHADNCVLDPDTGECWREPPAYTYRDYSGLLYLNDDFQGGDLFFTEPNALTVTARV RPRCGRLVAFSSGVENPHGVWAVTRGRRCALALWHTWAPEHREQEWEAKELLQESQE EEEEEEEEEMPSKDPSPPEPPSRRHQVRVQDKTGRAPRVREEL		

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

Table 6B. Protein Sequence Properties NOV6a	
PSort analysis:	0.4991 probability located in lysosome (lumen); 0.3700 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 21 and 22

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

Table 6C. Geneseq Results for NOV6a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB93142	Human protein sequence SEQ ID NO:12045 - Homo sapiens, 736 aa. [EP1074617-A2, 07-FEB-2001]	37..714 35..720	314/706 (44%) 421/706 (59%)	e-162
AAB93215				e-161

	NO:12194 - Homo sapiens, 736 aa. [EP1074617-A2, 07-FEB-2001]	35..720	421/706 (59%)	
AAB88373	Human membrane or secretory protein clone PSEC0109 - Homo sapiens, 736 aa. [EP1067182-A2, 10-JAN-2001]	37..714 35..720	313/706 (44%) 421/706 (59%)	e-161
AAB36392	Human tumour suppressor Gros1-S protein SEQ ID NO:4 - Homo sapiens, 736 aa. [WO200065047-A1, 02-NOV- 2000]	37..714 35..720	312/706 (44%) 419/706 (59%)	e-160
AAB36393	Mouse tumour suppressor Gros1-L protein SEQ ID NO:6 - Mus musculus, 747 aa. [WO200065047- A1, 02-NOV-2000]	24..714 22..722	308/721 (42%) 424/721 (58%)	e-159

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 6D.

Table 6D. Public BLASTP Results for NOV6a				
Protein Accession Number	Protein/Organism/Length	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q13512	PROTEIN B - Homo sapiens (Human), 551 aa.	186..736 1..551	551/551 (100%) 551/551 (100%)	0.0
Q15740	CHROMOSOME 12P13 SEQUENCE, COMPLETE SEQUENCE (HYPOTHETICAL 62.3 KDA PROTEIN) - Homo sapiens (Human), 551 aa.	186..736 1..551	550/551 (99%) 550/551 (99%)	0.0
O88836	CHROMOSOME 6 BAC-284H12 (RESEARCH GENETICS MOUSE BAC LIBRARY) COMPLETE SEQUENCE (RESEARCH GENETICS MOUSE BAC LIBRARY) (GENE RICH CLUSTER, B GENE) - Mus musculus (Mouse), 545 aa.	190..736 1..545	477/549 (86%) 508/549 (91%)	0.0
Q96SL5	CDNA FLJ14774 FIS, CLONE NT2RP4000051, WEAKLY SIMILAR TO SYNAPTONEMAL COMPLEX PROTEIN SC65 - Homo sapiens (Human), 736 aa.	37..714 35..720	314/706 (44%) 421/706 (59%)	e-161
Q96SK8	CDNA FLJ14791 FIS, CLONE NT2RP4001064, WEAKLY SIMILAR	37..714 35..720	313/706 (44%) 421/706 (59%)	e-161

	PROTEIN SC65 - Homo sapiens (Human), 736 aa.			
--	---	--	--	--

PFam analysis predicts that the NOV6a protein contains the domains shown in the Table 6E.

Table 6E. Domain Analysis of NOV6a			
Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

EXAMPLE 7.

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

Table 7A. NOV7 Sequence Analysis			
	SEQ ID NO: 25	372 bp	
NOV7a, CG57885-01 DNA Sequence	<u>CCATGAACAGCGGCGTGTGCCTGTGTGTGCTGATGGCGGTACTGGCGGCTGGCGCCCT</u> GACGCAGCCGGTGCCTCCCGCAGATCCCGCGGGCTCCGGGCTGCAGCGGGCAGAGGAG GCGCCCCGTAGGCAGCTGAGGGTATCGCAGAGAACGGATGGCGAGTCCCGAGCGCACCC TGGGCGCCCTGCTGGCAAGATACATCCAGCAGGCCCGGAAAGGTAAGAATGCTGCCTC CCCATCCCTCACTTCTGCCCTTGTTCCAGGCTCCCGATGCTGACCCTCTTCTCTAGC GCTAGCCTGATGGGGATGACCTCTCTCGGTAGGAAACAAGCAACATGATTTCTGGCGG TCCTTTGTAGCAATCTGAGAAGGG		
	ORF Start: ATG at 3	ORF Stop: TGA at 336	
	SEQ ID NO: 26	111 aa	MW at 11598.4kD
NOV7a, CG57885-01 Protein Sequence	MNSGVCLCVLMAVLAAGALTQVPPADPAGSGLQRAEEAPRRQLRVSQRTDGESRAHL GALLERYIQQARKGKNAASPSLTSAVLPRLPMLTLFSSASLMGMTSLGRKQAT		

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

Table 7B. Protein Sequence Properties NOV7a	
PSort analysis:	0.8200 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in microbody (peroxisome)
SignalP analysis:	Likely cleavage site between residues 21 and 22

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

Table 7C. Geneseq Results for NOV7a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE10339	Human cholecystokinin (CCK) - Homo sapiens, 136 aa. [WO200168828-A2, 20-SEP-2001]	1..110 22..129	82/113 (72%) 86/113 (75%)	1e-35
AAB24381	Human procholecystokinin amino acid sequence SEQ ID NO:1 - Homo sapiens, 115 aa. [WO200061192-A2, 19-OCT-2000]	1..110 1..108	82/113 (72%) 86/113 (75%)	1e-35
AAV04729	Rat brain cholecystokinin precursor amidation region - Rattus sp, 105 aa. [WO9910361-A1, 04-MAR-1999]	5..110 1..104	56/106 (52%) 62/106 (57%)	1e-19
AAB24382	Human CCK A amino acid sequence CCK-58 SEQ ID NO:2 - Homo sapiens, 58 aa. [WO200061192-A2, 19-OCT-2000]	46..91 1..41	31/46 (67%) 34/46 (73%)	1e-07

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

Table 7D. Public BLASTP Results for NOV7a				
Protein Accession Number	Protein/Organism/Length	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P06307	Procholecystokinin precursor (CCK) - Homo sapiens (Human), 115 aa.	1..110 1..108	82/113 (72%) 86/113 (75%)	4e-35
P23362	Procholecystokinin precursor (CCK) - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 115 aa.	1..110 1..108	77/113 (68%) 83/113 (73%)	3e-32
P01356	Procholecystokinin precursor (CCK) - Sus scrofa (Pig), 114 aa.	1..110 1..107	66/113 (58%) 73/113 (64%)	2e-24
Q9DCL5	ADULT MALE KIDNEY CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY, CLONE:0610025O15, FULL INSERT SEQUENCE - Mus musculus (Mouse), 115 aa.	1..110 1..108	63/113 (55%) 71/113 (62%)	1e-22
P09240	Procholecystokinin precursor (CCK) - Mus musculus (Mouse), 115 aa.	1..110 1..108	62/113 (54%) 69/113 (60%)	2e-21

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

Table 7E. Domain Analysis of NOV7a			
Pfam Domain	NOV7a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Gastrin: domain 1 of 1	2..71	37/80 (46%) 64/80 (80%)	7.5e-22

EXAMPLE 8.

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

Table 8A. NOV8 Sequence Analysis			
	SEQ ID NO: 27	479 bp	
NOV8a, CG57865-01 DNA Sequence	TGACTGTATCGCCGGAATTCATGAAGGATCGATTCAAGTGGTTGTCTGCTGGAGCTGCT CCTGCTGATAGGCGCCGAGTCGCCTTTCGGACGGCGCTCCGGCGGACACGTGCGTG AAGCAGCGGGCGAATCAGCCGAATCATGGCAAGGCCCGGAGTCAGCCGGCTCACTCGA ATCCGTACGAGGTGGTGGCCGTTGCGCAGACCTACCATCCCGGCCAGCAGATATCGGT GGTCATCTATCCGCACTCGGACCAGAGCACTGTCTTCCGGGGATTCTTCCTGCAGGCG CGCGATGCCAACTCGAACGAGTGGATCGGCGAGTGGGTGCAGAGCGAGAACACCAAGA CCATTCCAGAGTGCTCGGCCATCACGCACTCGGACAACCGGGACAAGCTGGGGGCCAA GCTCATCTGGAAGGCACCGCAAATAAGCGGGGACAAGTCTACTTCACGTA <u>ACT</u> GCAG CCAAGCTAATTCCGG		
	ORF Start: ATG at 21	ORF Stop: TAA at 456	
	SEQ ID NO: 28	145 aa	MW at 16248.2kD
NOV8a, CG57865-01 Protein Sequence	MKDRFKWLSLELLLLLIGAAVAFPDGAPADTCVKQRANQPNHGKARSQPAHSNPYEVVA VAQTYHPGQQISVVIYPHSDQSTVFRGFFLQARDANSNEWIGEWVQSENTKTIPECSA ITHSDNRDKLGAKLIWKAPQNKRGQVYFT		
	SEQ ID NO: 29	384 bp	
NOV8b, 171651532 DNA Sequence	GGATCCTTTCGGACGGCGCTCCGGCGGACACGTGCGTGAAGCAGCGGGCGAATCAGC CGAATCATGGCAAGGCCCGGAGTCAGCCGGCTCACTCGAATCCGTACGAGGTGGTGGC CGTTGCGCAGACCTACCATCCCGGCCAGCAGATATCGGTGGTCATCTATCCGCACTCG GACCAGAGCACTGTCTTCCGGGGATTCTTCCTGCAGGCGCGCGATGCCAACTCGAACG AGTGGATCGGCGAGTGGGTGTAGAGCGAGAACACCAAGACCATTCCAGAGTGCTCGGC CATCACGCACTCGGACAACCGGGACAAGCTGGGGGCCAAGCTCATCTGGAAGGCACCG CAAATAAGCGGGGACAAGTCTACTTCACGCTCGAG		
	ORF Start: GGA at 1	ORF Stop: TAG at 253	
	SEQ ID NO: 30	84 aa	MW at 9265.1kD
NOV8b, 171651532 Protein Sequence	GSFPDGAPADTCVKQRANQPNHGKARSQPAHSNPYEVVAVAQTYHPGQQISVVIYPHS DQSTVFRGFFLQARDANSNEWIGEWV		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 8B.

Table 8B. Comparison of NOV8a against NOV8b and NOV8c.		
Protein Sequence	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV8b	21..103 2..84	82/83 (98%) 83/83 (99%)

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

Table 8C. Protein Sequence Properties NOV8a	
PSort analysis:	0.6377 probability located in outside; 0.1821 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 22 and 23

A search of the NOV8a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
5 homologous proteins shown in Table 8D.

Table 8D. Geneseq Results for NOV8a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found				

In a BLAST search of public sequence databases, the NOV8a protein was found to have homology to the proteins shown in the BLASTP data in Table 8E.

Table 8E. Public BLASTP Results for NOV8a				
Protein Accession Number	Protein/Organism/Length	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9VAN1	CG14515 PROTEIN - Drosophila melanogaster (Fruit fly), 145 aa.	1..145 1..145	144/145 (99%) 144/145 (99%)	6e-82

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

Table 8F. Domain Analysis of NOV8a			
Pfam Domain	NOV8a Match Region		Expect Value

		Similarities for the Matched Region	
Reeler: domain 1 of 1	30..145	31/150 (21%) 78/150 (52%)	2.8e-05

EXAMPLE 9.

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

Table 9A. NOV9 Sequence Analysis			
	SEQ ID NO: 31	669 bp	
NOV9a, CG54503-03 DNA Sequence	TCCCGCGGGCCAGCGCACTACGAGATGCTGGGTCTGCTGCCGCATGGTGTGCGACCCGCG ATGGGCCCCGTGGCCCTGGTCCCGACGGCGCGCCTGCTTCCGTGCCCCCTTCCCGCC AGGCGCCAAGGGAGAGGTGGGCCGGCGCGGAAAGCAGGCCTGCGGGGGCCCCCTGGA CCACCAGGTCCAAGAGGGCCCCCAGGAGAACCCGGCAGGCCAGGCCCCCGGGCCCTC CCGGTCCAGGTCCGGGCGGGGTGGCGCCCGCTGCCGGCTACGTGCCTCGCATTGCTTT CTACGCGGGCCTGCGGCGGCCCCACGAGGGTTACGAGGTGCTGCGCTTCGACGACGTG GTGACCAACGTGGGCAACGCCTACGAGGCAGCCAGCGGCAAGTTTACTTGCCCCATGC CAGGCGTCTACTTCTTCGCTTACCACGTGCTCATGCGCGGCGGCGACGGCACCAGCAT GTGGGCCGACCTCATGAAGAACGGACAGGTCCGGGCCAGCGCCATTGCTCAGGACGCG GACCAGAACTACGACTACGCCAGCAACAGCGTCATTCTGCACCTGGACGTGGGCGACG AGGTCTTCATCAAGCTGGACGGCGGAAAGTGCACGGCGGCAACACCAACAAGTACAG CACCTTCTCCGGCTTCATCATCTACCCCGAC		
	ORF Start: TCC at 1	ORF Stop: th at 670	
	SEQ ID NO: 32	223 aa	MW at 23296.1kD
NOV9a, CG54503-03 Protein Sequence	SRGPAHYEMLGRCRMVCDPHGPRGPGPDGAPASVPPFPFGAKGEVGRRGKAGLRGPPG PPGPRGPPGEPGRPGPPGPPGPGGGVAPAAGYVPRIAFYAGLRRPHEGYEVLRFDDV VTNVGNAYEAASGKFTCPMPGVYFFAYHVLMRGGDGTSMWADLMKNGQVRASAIQAQDA DQNYDYASNSVILHLDVGDEVFIKLDGGKVHGGNTNKYSTFSGFIIYPD		

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

Table 9B. Protein Sequence Properties NOV9a	
PSort analysis:	0.8276 probability located in lysosome (lumen); 0.4500 probability located in cytoplasm; 0.4128 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV9a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 9C.

Table 9C. Geneseq Results for NOV9a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV9a	Identities/	Expect Value

		Match Residues	the Matched Region	
AAG64212	Murine HSP47 interacting protein, #2 - Mus sp, 255 aa. [JP2001145493- A, 29-MAY-2001]	5..223 20..255	169/236 (71%) 183/236 (76%)	1e-91
AAM40913	Human polypeptide SEQ ID NO 5844 - Homo sapiens, 755 aa. [WO200153312-A1, 26-JUL-2001]	19..222 519..754	90/242 (37%) 121/242 (49%)	2e-32
AAM39127	Human polypeptide SEQ ID NO 2272 - Homo sapiens, 744 aa. [WO200153312-A1, 26-JUL-2001]	19..222 508..743	90/242 (37%) 121/242 (49%)	2e-32
AAM40607	Human polypeptide SEQ ID NO 5538 - Homo sapiens, 255 aa. [WO200153312-A1, 26-JUL-2001]	19..223 43..252	82/218 (37%) 112/218 (50%)	3e-30
AAM38821	Human polypeptide SEQ ID NO 1966 - Homo sapiens, 253 aa. [WO200153312-A1, 26-JUL-2001]	19..223 41..250	82/218 (37%) 112/218 (50%)	3e-30

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 9D.

Table 9D. Public BLASTP Results for NOV9a				
Protein Accession Number	Protein/Organism/Length	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O88992	C1q-related factor precursor - Mus musculus (Mouse), 258 aa.	1..223 15..258	171/244 (70%) 183/244 (74%)	2e-93
O75973	C1q-related factor precursor - Homo sapiens (Human), 258 aa.	1..223 15..258	174/244 (71%) 185/244 (75%)	4e-93
Q9ESN4	Gliacolin precursor - Mus musculus (Mouse), 255 aa.	5..223 20..255	169/236 (71%) 183/236 (76%)	5e-91
Q921S8	PROCOLLAGEN, TYPE VIII, ALPHA 1 - Mus musculus (Mouse), 744 aa.	19..222 509..743	94/241 (39%) 123/241 (51%)	3e-34
Q9D2V4	PROCOLLAGEN, TYPE VIII, ALPHA 1 - Mus musculus (Mouse), 744 aa.	19..222 509..743	94/241 (39%) 123/241 (51%)	3e-34

PFam analysis predicts that the NOV9a protein contains the domains shown in the Table 9E.

Table 9E. Domain Analysis of NOV9a

Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Collagen: domain 1 of 1	35..92	36/60 (60%) 49/60 (82%)	0.0043
C1q: domain 1 of 1	96..220	43/140 (31%) 92/140 (66%)	6.4e-29

EXAMPLE 10.

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

[illegible]

	YVFSDGYLINYDLTFLTMKTRLPRPPTRRPSGAHAPPKPVKPNEASRP
--	--

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

Table 10B. Protein Sequence Properties NOV10a	
PSort analysis:	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1800 probability located in nucleus; 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV10a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
5 homologous proteins shown in Table 10C.

Table 10C. Geneseq Results for NOV10a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAV54368	Protein encoded by colon specific gene (CSG) clone 2348122 - Homo sapiens, 510 aa. [WO9960161-A1, 25-NOV-1999]	5..480 32..506	219/482 (45%) 310/482 (63%)	e-114
AAV22201	Human extracellular mucous matrix glycoprotein protein sequence - Homo sapiens, 510 aa. [US5929033-A, 27-JUL-1999]	5..480 32..506	219/482 (45%) 310/482 (63%)	e-114
AAE03653	Human extracellular matrix and cell adhesion molecule-17 (XMAD-17) - Homo sapiens, 510 aa. [WO200142285-A2, 14-JUN-2001]	10..480 35..506	217/477 (45%) 308/477 (64%)	e-114
AAB50955	Human PRO698 protein - Homo sapiens, 510 aa. [WO200073348-A2, 07-DEC-2000]	10..480 35..506	217/477 (45%) 308/477 (64%)	e-114
AAB65169	Human PRO698 (UNQ362) protein sequence SEQ ID NO:67 - Homo sapiens, 510 aa. [WO200073454-A1, 07-DEC-2000]	10..480 35..506	217/477 (45%) 308/477 (64%)	e-114

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 10D.

Table 10D. Public BLASTP Results for NOV10a				
Protein Accession Number	Protein/Organism/Length	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H1L6	BA209J19.1.1 (GW112 PROTEIN) - Homo sapiens (Human), 510 aa.	10..480 35..506	217/477 (45%) 308/477 (64%)	e-113
Q07081	Olfactomedin precursor (Olfactory mucus protein) - Rana catesbeiana (Bull frog), 464 aa.	53..477 32..460	155/441 (35%) 247/441 (55%)	2e-68
AAL66227	NOELIN-1 - Xenopus laevis (African clawed frog), 485 aa.	28..478 48..475	114/458 (24%) 194/458 (41%)	1e-32
AAL66226	NOELIN-2 - Xenopus laevis (African clawed frog), 458 aa.	32..478 25..448	113/454 (24%) 192/454 (41%)	3e-32
O95362	GW112 PROTEIN - Homo sapiens (Human), 187 aa.	139..315 2..176	77/178 (43%) 107/178 (59%)	7e-32

PFam analysis predicts that the NOV10a protein contains the domains shown in the Table 10E.

Table 10E. Domain Analysis of NOV10a			
Pfam Domain	NOV10a Match Region	Identities/ Similarities for the Matched Region	Expect Value
OLF: domain 1 of 1	224..481	93/294 (32%) 170/294 (58%)	8.1e-72

EXAMPLE 11.

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

Table 11A. NOV11 Sequence Analysis		
	SEQ ID NO: 35	1134 bp
NOV11a, CG57572-01 DNA Sequence	GCAGAAGAATAGGCTACTTTATTTCTGAAAAGGAGGGAGTTCCTGCCACCCATTGCA GGGAGGTCGCCATCAGGACAGTGAAGATGGTGACCCTGCGGAAGAGGACCCTGAAAGT GCTCACCTTCCTCGTGCTCTTCATCTTCCTCACCTCCTTCCTGAACTACTCCCACGCC ATGGTGGCCACCACCTGGTTCCCCAAGAAGATGGCCCTGGAGCTCTTGAGAACCTGA AGAGACTGATCAAGCACAGGCCCTGCACTTGCACCCACTGCATCAGGCAGCATGGGCT CTCAGCCTGGTTCGATGAGAGGTTCAACCAGATAGTGCAGCTGCTGCTGACTGCCAG AACGCGCTCTTGAGGACAACACCTACCAATGGTGGCTGAGGCTCCAGCAGGAGAAGA AGCCCAATATCATCAACAATACCATCAAGGAATTCAGAGCAGTACCTGGGAATGTGGA CCAATGCTGGAGAAGAGGTCGGTGGGCTGCTGGCACTGTGCTGTGCTGGGCAACTCG GGCAACCTGAGGCAATTGTCATATCACAATTTTATGCTCAGGATGAACAAGGCACCCA CGGCAGGGTTTGAAGCTGCTGCCGGGAGCAAACCGCCACCATCTGGTGTACCCTGA GAGCTTCGGGAGCTGGGGGACAATGTCAGCATGGTCCTGGTGCCCTTAAAGACCATG AACTTGGAGTGGGTGGTGAGCACCACCACCGGTGCCATTTCCACACCTACACCC	

	CGGTCCTCGTGAAGATCAGAGTGAAACAGGATAAGATCCTGATCTACCACCCAGCCTT CATCAAGTATGTCTTCGACAACTGGCTGCAGAGCCACAGGCGGTACCCACTCACCAGC ATCCTCTCGGTCATCTTCTCAATGCATGTCTGCGATAAGGTAGACTTGTATAGCTTCG GAGCAGATAGCAAAGGGAAGTGGCACCCTACTGGGAGAACCACTGTCTGCGGGGTC TTTTCACAAGACGGGGGTGCACGATGCAGGCTTTGAGTCTAACGTGACGGCCACCTTG GCTTCATCAATAAAATCCCGATCTTCAAGGGGAGATGACACAGTGAAGGGGTGAGGAT GGATGCCCCATCATGCCTCTGCGTTTCAAGCC		
	ORF Start: ATG at 85	ORF Stop: TGA at 1096	
	SEQ ID NO: 36	337 aa	MW at 38559.2kD
NOV11a, CG57572-01 Protein Sequence	MVTLRKRTLKVLTFVLFIFLTSLNYSHAMVATTWFPKKMALELLENLKRLIKHRPC TCTHCIRQHGLSAWFDERFNQIVQLLLTAQNALLEDNTYQWWLRLQQEKKPNIINNTI KEFRAVPGNVDPMLEKRSVGCWHCAVVGNSGNLRQLSYHNFMLRMNKAPTAGFEAAAG SKTAHHLVYPESFRELGDNVSMVLVPLKTMNLEWVSTTTTGAISHTYTPVLVKIRVK QDKILIIYHPAFIKYVFDNWLQSHRRYPLTSILSVIFSMHVC DKVDLYSFGADSKGNWH HYWENNLSAGSFHKTGVDAGFESNVTATLASSIKSRSSRGDDTVKG		

Further analysis of the NOV11a protein yielded the following properties shown in Table 11B.

Table 11B. Protein Sequence Properties NOV11a	
PSort analysis:	0.8200 probability located in outside; 0.5054 probability located in lysosome (lumen); 0.1565 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Likely cleavage site between residues 31 and 32

A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several

5 homologous proteins shown in Table 11C.

Table 11C. Geneseq Results for NOV11a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR65244	Human ST30 sialyltransferase - Homo sapiens, 340 aa. [WO9504816-A, 16-FEB-1995]	1..331 1..339	274/339 (80%) 292/339 (85%)	e-158
AAR65240	Porcine ST30 sialyltransferase - Sus scrofa, 343 aa. [WO9504816-A, 16-FEB-1995]	5..331 6..342	234/337 (69%) 272/337 (80%)	e-137
AAR41670	Porcine sialyltransferase - Sus scrofa, 343 aa. [WO9318157-A, 16-SEP-1993]	5..331 6..342	234/337 (69%) 272/337 (80%)	e-137
AAR75198	Rat Gal-beta-1,3GalNAc, alpha-2,3-	12..332 12..350	149/341 (43%) 203/341 (58%)	6e-78

	norvegicus, 350 aa. [JP07236477-A, 12-SEP-1995]			
AAR75200	Rat P-F4M active fragment, SF-314R - Rattus norvegicus, 314 aa. [JP07236477-A, 12-SEP-1995]	50..332 26..314	135/290 (46%) 183/290 (62%)	3e-76

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 11D.

Table 11D. Public BLASTP Results for NOV11a				
Protein Accession Number	Protein/Organism/Length	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q11201	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2, 3-sialyltransferase (EC 2.4.99.4) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha 2,3-ST) (Gal-NAC6S) (Gal-beta-1,3-GalNAc-alpha-2,3-sialyltransferase) (ST3GALIA) (ST3O) (ST3GALA.1) (SIAT4-A) - Homo sapiens (Human), 340 aa.	1..331 1..339	278/339 (82%) 295/339 (87%)	e-160
Q9UN51	ALPHA-2,3-SIALYLTRANSFERASE - Homo sapiens (Human), 340 aa.	1..331 1..339	276/339 (81%) 294/339 (86%)	e-159
P54751	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2, 3-sialyltransferase (EC 2.4.99.4) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha 2,3-ST) (GAL-NAC6S) (GAL-beta-1,3-GALNAC-alpha-2,3-sialyltransferase) (ST3GALIA) (ST3O) (ST3GALA.1) (SIAT4-A) - Mus musculus (Mouse), 337 aa.	4..331 1..336	230/336 (68%) 273/336 (80%)	e-137
A45073	Gal beta 1,3GalNAc alpha 2,3-sialyltransferase - pig, 343 aa.	5..331 6..342	234/337 (69%) 272/337 (80%)	e-137
Q02745	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2, 3-sialyltransferase (EC 2.4.99.4) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha 2,3-ST) (GAL-NAC6S) (GAL-beta-1,3-GALNAC-alpha-2,3-sialyltransferase) (ST3GALIA) (ST3O) (ST3GALA.1) (SIAT4-A) - Sus scrofa (Pig), 343 aa.	5..331 6..342	234/337 (69%) 272/337 (80%)	e-136

PFam analysis predicts that the NOV11a protein contains the domains shown in the Table 11E.

Table 11E. Domain Analysis of NOV11a			
Pfam Domain	NOV11a Match Region	Identities/ Similarities for the Matched Region	Expect Value
IF3: domain 1 of 1	193..202	6/10 (60%) 9/10 (90%)	6.3
Glyco_transf_29: domain 1 of 1	60..331	97/324 (30%) 223/324 (69%)	4.7e-73

EXAMPLE 12.

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

Table 12A. NOV12 Sequence Analysis		
	SEQ ID NO: 37	4295 bp
NOV12a, CG57518-01 DNA Sequence	TCTTCGTCGCCGCTCTCTCTCACCTCTCAGGGAAAGGGGGGACATAGGGGCGTCG CGGGGCCCCGCGAATGCGCCCCCGCGCCTCTCGGGCTGCGCCGCTCGCGGGGAT GAAGCACCGGCCGTGAAGATGGAGGTGACCTGCCTTCTACTTCTGGCGCTGATCCCCT TCCACTGCCGGGGACAAGGAGTCTACGCTCCAGCCCAGGCGCAGATCGTGATGCGGG CCAGGCATGTGTGGTGAAGAGGACAATATCAGCGAGCGTGTCTACACCATCCGGGAG GGGGACACCCTCATGCTGCAGTGCCTTGTAACAGGGCACCTCGACCCCAGGTACGGT GGACCAAGACGGCAGGTAGCGCCTCGGACAAGTTCCAGGAGACATCGGTGTTCAACGA GACGCTGCGCATCGAGCGTATTGCACGCACGCAGGGCGGCCGCTACTACTGCAAGGCT GAGAACGGCGTGGGGGTGCCGGCCATCAAGTCCATCCGCGTGGACGTGCAGTACCTGG ATGAGCCAATGCTGACGGTGCACCAGACGGTGAGCGATGTGCGAGGCAACTTCTACCA GGAGAAGACGGTGTTCCTGCGCTGTACTGTCAACTCCAACCCGCTGCCGCTTCATC TGGAAGCGGGTTCCGATACCCTATCCCACAGCCAGGACAATGGGGTTGACATCTATG AGCCCCCTTACACTCAGGGGGAGACCAAGGTCTGAAGCTGAAGAACCCTGCGGCCCA GGACTATGCCAGCTACACCTGCCAGGTGTCTGTGCGTAACGTGTGCGGCATCCCAGAC AAGGCCATCACCTTCCGGCTCACCAACACCACGGCACCCAGCCCTGAAGCTGTCTG TGAACGAACTCTGGTGGTGAACCCTGGGGAGAATGTGACGGTGCAGTGTCTGCTGAC AGGCGGTGATCCCCTCCCCAGCTGCAGTGGTCCCATGGGCCTGGCCCACTGCCCTG GGTGCTCTGGCCAGGGTGGCACCTCAGCATCCCTTCAGTGCAGGCCCGGGACTCTG GCTACTACAACCTGCACAGCCACCAACAATGTGGGCAACCCTGCCAAGAAGACTGTCAA CCTGCTGGTGGATCCATGAAGAACGCTACATTCCAGATCACTCCTGACGTGATCAAA GAGAGTGAGAACATCCAGCTGGGCCAGGACCTGAAGCTATCGTGCCACGTGGATGCAG TGCCCCAGGAGAAGGTGACCTACCAGTGGTTCAAGAATGGCAAGCCGGCACGCATGTC CAAGCGGCTGCTGGTGACCCGCAATGATCCTGAGCTGCCCCGAGTCACCAGCAGCCTA GAGCTCATTGACCTGCACTTCAGTGACTATGGCACCTACCTGTGCATGGCTTCTTTCC CAGGGGCACCCGTGCCCGACCTCAGCGTCGAGGTCAACATCTCCTCTGAGACAGTGCC GCCCACCATCAGTGTGCCAAGGGTAGGGCCGTGGTGACCGTGCGCGAGGGATCGCCT GCCGAGCTGCAATGCGAGGTGCGGGGCAAGCCGCGGCCGAGTGCTCTGGTCCCGCG TGGACAAGGAGGCTGCACTGCTGCCCTCGGGGCTGCCCTGGAGGAGACTCCGGACGG GAAGCTGCGGCTGGAGCGAGTGAGCCGAGACATGAGCGGGACCTACCGCTGCCAGACG GCCCCGTATAATGGCTTCAACGTGCGCCCCCGTGAGGCCAGGTGCAGCTGAACGTGC AGTTCCCGCCGGAGGTGGAGCCAGTTCCCAGGACGTGCGCCAGGCGCTGGGCCGGCC CGTGCTCCTGCGCTGCTCGCTGCTGCGAGGCAGCCCCAGCGCATCGCCTCGGCTGTG TGGCGTTTCAAAGGGCAGCTGCTGCCGCCGCCGCTGTTGTTCCCGCCGCCGCGAGG CGCCGGATCACGCGGAGCTGCGCCTCGACGCCGTAACCTCGCGACAGCAGCGGCAGCTA CGAGTGACGCGTCTCCAACGATGTGGGCTCGGCTGCCTGCCTCTTCCAGGTCTCCGCC AAAGCCTACAGCCCGGAGTATTACTTCGACACCCCCAACCCACCCGAGCCACAAGC TGTCCAAGAACTACTCCTACGTGCTGCAGTGGACTCAGAGGGAGCCCGACGCTGTCGA	

	CCCTGTGCTCAACTACAGACTCAGCATCCGCCAGTTGAACCAGCACAATGCGGTGGTC AAGGCCATCCCGGTCCGGCGTGTGGAGAAGGGGCAGCTGCTGGAGTACATCCTGACCG ATCTCCGTGTGCCCCACAGCTATGAGGTCCGCCTCACACCCTATACCACCTTCGGGGC TGGTGACATGGCCTCCCGCATCATCCACTACACAGAGCGCCAGATCCGCTGGCCCCCA GTCCTGGCTCTGAGGACCCTGTCTTCTGGTCCCAAGCAGGGTATCCTCTGCAGAGCCC CACACCTCAGTTCTGACTTGGTTTCCCCGCTTGCTTTCTCAGCCATCAACTCTCCGAA CCTTTCAGACAACACCTGCCACTTTGAGGATGAGAAGATCTGTGGCTATACCCAGGAC CTGACAGACAACCTTTGACTGGACGCGGCAGAATGCCCTCACCCAGAACCCCAAACGCT CCCCCAACACTGGTCCCCCACCACATAAGTGGCACCCCTGAGGGCTACTACATGTT CATCGAGACATCGAGGCCTCGGGAGCTGGGGGACCGTGCAAGGTTAGTGAGTCCCCTC TACAATGCCAGCGCCAAGTTCTACTGTGTCTCCTTCTTCTACCACATGTACGGGAAAC ACATCGGCTCCCTCAACCTCCTGGTGCGGTCCCGGAACAAAGGGGCTCTGGACACGCA CGCCTGGTCTCTCAGTGGCAATAAGGGCAATGTGTGGCAGCAGGCCCATGTGCCCATC AGCCCCAGTGGGCCCTTCCAGATTATTTTTGAGGGGGTTCGAGGCCCGGGCTACCTGG GGGATATTGCCATAGATGACGTCACACTGAAGAAGGGGGAGTGTCCCCGGAAGCAGAC GGATCCCAATAAAGGTGCAAGACGGGAAGGAGCTGCCTGCGATGGCCTGAAATTCCAC CTTTCATCCCCTATGGATGACGGAGAGCTTACAGATGACCCTATTGAATGCAAGCACC TTTGGATCCATAGAGTGGACAGTAAAGGTGCTCAGTACATGTTGGCTGAGCTGAACTG CATACTGTGGCCCCCAGGTTCTTGGTCTTTATGGACGAAGGGCACAAGGTTGGTGAA AAGGACTCCGGGGGCCAGCCCTTCCAAGTTTACACTGATTTCTCCTTTTACCCTCATG CTATCCCTGAGAAGATGTCAATAATGCCACGTTACAGGTGGGAAAACCTGAGGCTTAG AGAGGAGGAGGAATCTGCCTACGGTCACACAGCTGCAAAGGCTAGAGCTGGGACCAGG AGCTGGTCTCTTAACCGACCACCTGAGCTCAAGAGCTTTTCTCTCTGGACCAACATGA CCCAAAGTGTGCGCGAGCCTATCACAGGTCCCCTGCAATGCCAAACATACACGCACAG CAATACACAACACCTGGGGACATGGATGAAGCTGGAAACCATCATTCTCAGCAAACCTG ACACAAGAACAGAAAACCAAACACCACATGTTCTCACTACCACCCAGTCTGCCCCGC CCTCTCTCTCTCACCTGAACTTCCCCTCTCCTCAAACCTCTCGAGGCCACGCCTCTAT GTCCTTGGATGATGATGATGACGACGACGACGATGATGATGATGATGATGACGACGAT GACAATGATGATGATGATGGAAGGAAGACCTACAGAATCCCTCCAGGCTCTGACCTCA GTGCTTGTGGGTGGGTGAATGACCACATGTGCGCAGGGAGACTCCACAGGTCCTCCCGA TGAGAAGCACTCTTATGCCAAAGAGGAGACTCAGGCCAAACTGACAGGACCAGGAATT AGCTACCCTGGTAAACCCAGCTATCGACTGCACCCGAGCGGCTACACACCACTGGAGC AGTTCAGGGAGAAAGCCACCGGCATGCTCACCCCGTATGTCTCTGGCTCTGTTTCCTC TTTCTGCTTCCCCTTCCCCACCTCTGAGTCTCTGTGTTCTGCTCATGCCAATTCCCCT TCTGCCTGTCTCTGCCCCGTTCTCTCTCTGGGCTGGTCTCTCCGAGACTCTGTTCCCT TGGCTGGCATGCCCTCCACCTCCCCTGATGCTGGAGCAGTTCAGGGAGAAAGCCACCG GCATGCTCACCGTATGTCTCTGGCTCTGTTTCCTCTTTCTGCTTCCCCTTCCCCACCT TGA		
	ORF Start: ATG at 135	ORF Stop: TGA at 4293	
	SEQ ID NO: 38	1386 aa	MW at 153195.2kD
NOV12a, CG57518-01 Protein Sequence	MEVTCLLLLLALIPFHCRCGQGVYAPAAQAQIVHAGQACVVKEDNISERVYTIREDTLML QCLVTGHPRPQVRWTKTAGSASDKFQETSVFNETLRIERIARTQGGRYCYKAENGVG PAIKSIRVDVQYLDEPMLTVHQTVSDVRGNFYQEKTVFLRCTVNSNPPARFIWKRGS TLSHSQDNGVDIYEPLYTQGETKVLKLNLRPDYASYTCQVSVRNVCGIPDKAITFR LTNTTAPPALKLSVNETLVVNPGENVTVQCLLTGGDPLPQLQWSHGPGLPLGALAQG GTLISIPSVQARDSGYNCTATNNVGNPAKKTVNLLVRSMKNATFQITPDVKESENIQ LGQDLKLSCHVDAVPQEKVTYQWFKNGKPARMSKRLLVTRNDPELPAVTSSLELIDLH FSDYGTLYCMASFPGAPVPDLSVEVNISSETVPPTISVPKGRAVVTVREGSPAELQCE VRGKPRPPVLWSRVDKEAALLPSGLPLEETPDGKLRLERVSRDMSGTYRCQTARYNGF NVRPREAQVQLNVQFPPEVEPSSQDVRQALGRPVLLRCSLLRGSPQRIASAVWRFKGQ LLPPPVPVPAAEAPDHAELRLDAVTRDSSGSYECVSNDVGSAACLFQVSAKAYSPE IYFDTNPTRSHKLSKNYSYVLQWTQREPDVDPVLNYRLSIRQLNQHNAVVKAI PVR RVEKGQLLEYILTLDRVPHSYEVRLTPYTTFGAGDMASRIIHYTERQIRWPPVLALRT LSSGPKQGILCRAPHLSSDLVSPLAFSAINSPNLSDNTHFEDEKICGYTQDLTDNFD WTRQNALTQNPKRSPNTGPPTDISGTPEGYMFIEISRPRELGDRARLVSPLYNASAK FYCVSFFYHMYGKHIGSLNLLVRSRNRKALDTHAWSLSGNKGNVWQQAHVPISPSGPF QII FEGVRGPGYLGDI AIDVTLKKGECPRKQTDPNKGARREGAACDGLKFHLSSPMD DGELTDDPIECKHLWIHRVDSKGAQYMLAELNCIHVAPRFLVFMDEGHKVGKDSGGQ PFQVYTDFSFPYPAIPEKMSIMPTLQVGKLRRLREEEESAYGHTAAKARAGTRSWSLNR		

	PPELKSFSLWTNMTQSVREPITGPLQCQTYTHSNTQHLGTWMKLETIILSKLTQEOKT KHHMFSLTTQSAPPSLFSPELPLSSNSRGHASMSLDDDDDDDDDDDDDDDDDDDDNDDDD GRKTYRIPPGSDLACGWVNDHMSQGDSTGPPDEKHSYAKEETQAKLTGPGISYPGKP SYRLHPSGYTPLEQFREKATGMLTPYVSGSVSSFCFPFPTSESLCSAHANSPSACLCP LLSLGWSLRDSVPLAGMPSTSPDAGAVQGESHRHAHRMSLALFPLSASPSP		
	SEQ ID NO: 39	906 bp	
NOV12b, 170108372 DNA Sequence	GGTACCCACACAGCCCTGAAGCTGTCTGTGAACGAACTCTGGTGGTGAACCCTGGGG AGAATGTGACGGTGACGTGTCTGCTGACAGGCGGTGATCCCCTCCCCAGCTGCAGTG GTCCCATGGGCCTGGCCCACTGCCCCTGGGTGCTCTGGCCCAGGGTGGCACCCTCAGC ATCCCTTCAGTGCAGGCCCCGGGACTCTGGCTACTACAACCTGCACAGCCACCAACAATG TGGGCAACCCTGCCAAGAAGACTGTCAACCTGCTGGTGCGATCCATGAAGAACGCTAC ATTCCAGATCACTCCTGACGTGATCAAAGAGAGTGAGAACATCCAGCTGGGCCAGGAC CTGAAGCTATCGTGCCACGTGGATGCAGTGCCCCAGGAGAAGGTGACCTACCAGTGGT TCAAGAATGGCAAGCCGGCACGCATGTCCAAGCGGCTGCTGGTGACCCGCAATGATCC TGAGCTGCCCCGAGTCACCAGCAGCCTAGAGCTCATTGACCTGCACTTCAGTGACTAT GGCACCTACCTGTGCATGGCTTCTTTCCAGGGGCACCCGTGCCCCGACCTCAGCGTCG AGGTCAACATCTCCTCTGAGACAGTGCCGCCCCACCATCAGTGTGCCCAAGGGTAGGGC CGTGGTGACCGTGCGCGAGGGATCGCCTGCCGAGCTGCAATGCGAGGTGCGGGGCAAG CCGCGGCCCGCCAGTGCCCTGGTCCCGCGTGGACAAGGAGGCTGCACTGCTGCCCTCGG GGCTGCCCCCTGGAGGAGACTCCGGACGGGAAGCTACGGCTGGAGCGAGTGAGCCGAGA CATGAGCGGGACCTACCGCTGCCAGACGGCCCCGCTATAATGGCTTCAACGTGCGCCCC CGTGAGGCCACAGGTGCAGCTGAACGTGCAGGAATTC		
	ORF Start: GGT at 1	ORF Stop:	
	SEQ ID NO: 40	302 aa	MW at 32832.1kD
NOV12b, 170108372 Protein Sequence	GTPPALKLSVNETLVVNPGENVTVQCLLTGGDPLPQLQWSHGPGPLPLGALAQQGGLS IPSVQARDSGYYNCTATNNVGNPAKKTVNLLVRSMKNATFQITPDVIKESENIQLGQD LKLSCHVDAVPQEKVTYQWFKNGKPARMSKRLLVTRNDPELPAVTSSLELIDLHFSY GTYLCMASFPGAPVPDLSEVNISSETVPPTISVPKGRAVVTVREGSPAELQCEVRGK PRPPVPWSRVDKEAALLPSGLPLEETPDGKLRLERVSRDMSGTYRCQTARYNGFNVRP REAQVQLNVQEF		
	SEQ ID NO: 41	906 bp	
NOV12c, 170108393 DNA Sequence	GGTACCCACACAGCCCTGAAGCTGTCTGTGAACGAACTCTGGTGGTGAACCCTGGGG AGAATGTGACGGTGACGTGTCTGCTGACAGGCGGTGATCCCCTCCCCAGCTGCAGTG GTCCCATGGGCCTGGCCCACTGCCCCTGGGTGCTCTGGCCCAGGGTGGCACCCTCAGC ATCCCTTCAGTGCAGGCCCCGGGACTCTGGCTACTACAACCTGCACAGCCACCAACAATG TGGGCAACCCTGCCAAGAAGACTGTCAACCTGCTGGTGCGATCCATGAAGAACGCTAC ATTCCAGATCACTCCTGACGTGATCAAAGAGAGTGAGAACATCCAGCTGGGCCAGGAC CTGAAGCTATCGTGCCACGTGGATGCAGTGCCCCAGGAGAAGGTGACCTACCAGTGGT TCAAGAATGGCAAGCCGGCACGCATGTCCAAGCGGCTGCTGGTGACCCGCAATGATCC TGAGCTGCCCCGAGTCACCAGCAGCCTAGAGCTCATTGACCTGCACTTCAGTGACTAT GGCACCTACCTGTGCATGGCTTCTTTCCAGGGGCACCCGTGCCCCGACCTCAGCGTCG AGGTCAACATCTCCTCTGAGACAGTGCCGCCCCACCATCAGTGTGCCCAAGGGTAGGGC CGTGGTGACCGTGCGCGAGGGATCGCCTGCCGAGCTGCAATGCGAGGTGCGGGGCAAG CCGCGGCCCGCCAGTGCTCTGGTCCCGCGTGGACAAGGAGGCTGCACTGCTGCCCTCGG GGCTGCCCCCTGGAGGAGACTCCGGACGGGAAGCTGCGGCTGGAGCGAGTGAGCCGAGA CATGAGCGGGACCTACCGCTGCCAGACGGCCCCGCTATAATGGCTTCAACGTGCGCCCC CGTGAGGCCACAGGTGCAGCTGAACGTGCAGGAATTC		
	ORF Start: GGT at 1	ORF Stop:	
	SEQ ID NO: 42	302 aa	MW at 32848.2kD
NOV12c, 170108393 Protein Sequence	GTPPALKLSVNETLVVNPGENVTVQCLLTGGDPLPQLQWSHGPGPLPLGALAQQGGLS IPSVQARDSGYYNCTATNNVGNPAKKTVNLLVRSMKNATFQITPDVIKESENIQLGQD LKLSCHVDAVPQEKVTYQWFKNGKPARMSKRLLVTRNDPELPAVTSSLELIDLHFSY GTYLCMASFPGAPVPDLSEVNISSETVPPTISVPKGRAVVTVREGSPAELQCEVRGK PRPPVLWSRVDKEAALLPSGLPLEETPDGKLRLERVSRDMSGTYRCQTARYNGFNVRP REAQVQLNVQEF		

	SEQ ID NO: 43	720 bp
NOV12d, 170343246 DNA Sequence	GGTACCTTGAACCAGCACAAATGCGGTGGTCAAGGCCATCCCGGTCCGGCGTGTGGAGA AGGGGCAGCTGCTGGAGTACATCCTGACCGATCTCCGTGTGCCCCACAGCTATGAGGT CCGCCTCACACCCTATACCACCTTCGGGGCTGGTGACATGGCCTCCCGCATCATCCAC TACACAGAGCCCATCAACTCTCCGAACCCTTCAGACAACACCTGCCACTTTGAGGATG AGAAGATCTGTGGCTATAACCCAGGACCTGACAGACAACTTTGACTGGACGCGGCAGAA TGCCCTCAGCCAGAACCCCAAACGCTCCCCCAACACTGGTCCCCCACCACGACATAAGT GGCACCCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCTCGGGAGCTGGGGG ACCGTGCAAGGTTAGTGAGTCCCCTCTACAATGCCAGCGCCAAGTTCTACTGTGTCTC CTTCTTCTACCACATGTACGGGAAACACATCGGCTCCCTCAACCTCCTGGTGCGGTCC CGGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCAATAAGGGCAATG TGTGGCAGCAGGCCCATGTGCCCATCAGCCCCAGTGGGCCCTTCCAGATTATTTTGA GGGGGTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGACGTCACACTGAAG AAGGGGGAGTGTCCCCGGGAATTC	
	ORF Start: GGT at 1	ORF Stop: at 721
	SEQ ID NO: 44	240 aa MW at 26966.1kD
NOV12d, 170343246 Protein Sequence	GTLNQHNAVVKAI PVR RVEKGQLLEYILTDLRVPHSYEVRLTPYTTFGAGDMASRI IH YTEPINSNPNSDNTCHFEDKICGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDIS GTPEGYYMFIETSRPRELGDRARLVSPLYNASAKFYCVSFFYHMYGKHIGSLNLLVRS RNKGALDTHAWSLSGNKGNVWQQAHPVISPSPGFQIIFEGVRGPGYLGDIADDDVTLK KGECPPREF	
	SEQ ID NO: 45	720 bp
NOV12e, 170343692 DNA Sequence	GGTACCTTGAACCAGCACAAATGCGGTGGTCAAGGCCATCCCGGTCCGGCGTGTGGAGA AGGGGCAGCTGCTGGAGTACATCCTGACCGATCTCCGTGTGCCCCACAGCTATGAGGT CCGCCTCACACCCTATACCACCTTCGGGGCTGGTGACATGGCCTCCCGCATCATCCAC TACACAGAGCCCATCAACTCTCCGAACCCTTCAGACAACACCTGCCACTTTGAGGATG AGAAGATCTGTGGCTATAACCCAGGACCTGACAGACAACTTTGACTGGACGCGGCAGAA TGCCCTCAGCCAGAACCCCAAACGCTCCCCCAACACTGGTCCCCCACCACGACATAAGT GGCACCCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCTCGGGAGCTGGGGG ACCGTGCAAGGTTAGTGAGTCCCCTCTACAATGCCAGCGCCAAGTTCTACTGTGTCTC CTTCTTCTACCACATGTACGGGAAACACATCGGCTCCCTCAACCTCCTGGTGCGGTCC CGGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCAATAAGGGCAATG TGTGGCAGCAGGCCCATGTGCCCATCAGCCCCAGTGGGCCCTTCCAGATTATTTTGA GGGGGTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGACGTCACACTGAAG AAGGGGGAGTGTCCCCGGGAATTC	
	ORF Start: GGT at 1	ORF Stop: at 721
	SEQ ID NO: 46	240 aa MW at 26998.2kD
NOV12e, 170343692 Protein Sequence	GTLNQHNAVVKAI PVR RVEKGQLLEYILTDLRVPHSYEVRLTPYTTFGAGDMASRI IH YTEPINSNPNSDNTCHFEDKICGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDIS GTPEGYYMFIETSRPRELGDRARLVSPLYNACAKFYCVSFFYHMYGKHIGSLNLLVRS RNKGALDTHAWSLSGNKGNVWQQAHPVISPSPGFQIIFEGVRGPGYLGDIADDDVTLK KGECPPREF	
	SEQ ID NO: 47	720 bp
NOV12f, 170684238 DNA Sequence	GGTACCTTGAACCAGCACAAATGCGGTGGTCAAGGCCATCCCGGTCCGGCGTGTGGAGA AGGGGCAGCTGCTGGAGTACATCCTGACCGATCTCCGTGTGCCCCACAGCTATGAGGT CCGCCTCACACCCTATACCACCTTCGGGGCTGGTGACATGGCCTCCCGCATCATCCAC TACACAGAGCCCATCAACTCTCCGAACCCTTCAGACAACACCTGCCACTTTGAGGATG AGAAGATCTGTGGCTATAACCCAGGACCTGACAGACAACTTTGACTGGACGCGGCAGAA TGCCCTCAGCCAGAACCCCAAACGCTCCCCCAACACTGGTCCCCCACCACGACATAAGT GGCACCCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCTCGGGAGCTGGGGG ACCGTGCAAGGTTAGTGAGTCCCCTCTACAATGCCAGCGCCAAGTTCTACCGTGTCTC CTTCTTCTACCACATGTACGGGAAACACATCGGCTCCCTCAACCTCCTGGTGCGGTCC CGGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCAATAAGGGCAATG TGTGGCAGCAGGCCCATGTGCCCATCAGCCCCAGTGGGCCCTTCCAGATTATTTTGA GGGGGTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGACGTCACACTGAAG	

	AAGGGGGAGTGTCCCCGGGAATTC		
	ORF Start: GGT at 1	ORF Stop: at 721	
	SEQ ID NO: 48	240 aa	MW at 27035.1kD
NOV12f, 170684238 Protein Sequence	GTLNQHNNAVVKAI PVRRVEKGQLLEYILTDLRVPHSYEVRLTPYTTFGAGDMASRIIH YTEPINSPNLS DNTCHFEDEKICGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDIS GTPEGYYMFIETSRPRELGDRARLVSPLYNASAKFYRVSFFYHMYGKHIGSLNLLVRS RNKGALDTHAWSLSGNKGNVWQQAHPVISPSPGPFQII FEGVRGPGYLGDIAIDDVTLK KGECPRF		
	SEQ ID NO: 49	496 bp	
NOV12g, 170534177 DNA Sequence	GGGTACCTGTGGCTATACCCAGGACCTGACAGACAACCTTTGACTGGACGCGGCAGAAT GCCCTCACCAGAACCCCAAACGCTCCCCAACACTGGTCCCCCACCACATAAGTG GCACCCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCTCGGGAGCTGGGGGA CCGTGCAAGGTTAGTGAGTCCCCTCTACAATGCCAGCGCCAAGTTCTACTGTGTCTCC TTCTTCTATCACATGTACGGGAAACACATCGGCTCCCTCAACCTCCTGGTGCGGTCCC GGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCAATAAGGGCAATGT GTGGCAGCAGGCCCATGTGCCCATCAGTCCCAGTGGGCCCTTCCAGATTATTTTGTAG GGGGTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGACGTCACACTGAAGA AGGGGGAGTGTCCCCGGAAGCAGACGGAATTC		
	ORF Start: GGT at 2	ORF Stop: at 497	
	SEQ ID NO: 50	165 aa	MW at 18420.5kD
NOV12g, 170534177 Protein Sequence	GTCGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDISGTPEGYYMFIETSRPRELGD RARLVSPLYNASAKFYCVSFFYHMYGKHIGSLNLLVRSRNKGALDTHAWSLSGNKGNV WQQAHPVISPSPGPFQII FEGVRGPGYLGDIAIDDVTLKKGECPRKQTEF		

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

Table 12B. Comparison of NOV12a against NOV12b through NOV12g.		
Protein Sequence	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV12b	239..536 3..300	283/298 (94%) 283/298 (94%)
NOV12c	239..536 3..300	284/298 (95%) 284/298 (95%)
NOV12d	683..959 3..239	234/277 (84%) 235/277 (84%)
NOV12e	683..959 3..239	234/277 (84%) 235/277 (84%)
NOV12f	683..959 3..239	234/277 (84%) 235/277 (84%)
NOV12g	801..962 3..164	161/162 (99%) 162/162 (99%)

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

Table 12C. Protein Sequence Properties NOV12a	
PSort analysis:	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 19 and 20

A search of the NOV12a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, 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
AAE00582	Human nuclear cell adhesion molecule homologue, NCAM_d_1 protein - Homo sapiens, 946 aa. [WO200129215-A2, 26-APR-2001]	23..959 15..912	487/946 (51%) 656/946 (68%)	0.0
AAE00581	Human cell adhesion molecule homologue (CAM-H) protein #1 - Homo sapiens, 1018 aa. [WO200129215-A2, 26-APR-2001]	23..959 15..912	487/946 (51%) 656/946 (68%)	0.0
AAE00586	Human nuclear cell adhesion molecule homologue, NCAM_d_2 protein - Homo sapiens, 891 aa. [WO200129215-A2, 26-APR-2001]	71..959 8..857	455/898 (50%) 617/898 (68%)	0.0
AAY72717	HBXDJ03 clone human attractin-like protein #2 - Homo sapiens, 448 aa. [WO200116156-A1, 08-MAR-2001]	508..965 1..418	416/458 (90%) 417/458 (90%)	0.0
AAY72714	HBXDJ03 clone human attractin-like protein #1 - Homo sapiens, 448 aa. [WO200116156-A1, 08-MAR-2001]	508..965 1..418	406/458 (88%) 407/458 (88%)	0.0

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

CAB86654	DJ402N21.3 (NOVEL PROTEIN WITH IMMUNOGLOBULIN DOMAINS) - Homo sapiens (Human), 299 aa (fragment).	239..536 1..299	298/299 (99%) 298/299 (99%)	e-172
CAB86653	DJ402N21.2 (NOVEL PROTEIN WITH MAM DOMAIN) - Homo sapiens (Human), 273 aa (fragment).	683..965 1..243	242/283 (85%) 242/283 (85%)	e-138
Q9DBX0	1200011I03RIK PROTEIN - Mus musculus (Mouse), 267 aa.	689..965 1..237	227/277 (81%) 232/277 (82%)	e-129
Q9GMT4	HYPOTHETICAL 51.2 KDA PROTEIN - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 448 aa.	508..959 1..414	205/461 (44%) 281/461 (60%)	e-109
CAB86655	DJ402N21.1 (NOVEL PROTEIN) - Homo sapiens (Human), 127 aa (fragment).	1..127 1..127	127/127 (100%) 127/127 (100%)	3e-68

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
ig: domain 1 of 7	53..110	14/61 (23%) 42/61 (69%)	2.5e-08
ig: domain 2 of 7	150..216	14/70 (20%) 51/70 (73%)	3.7e-09
ig: domain 3 of 7	255..310	18/58 (31%) 38/58 (66%)	2.4e-08
PKD: domain 1 of 1	239..327	22/100 (22%) 56/100 (56%)	7.3
ig: domain 4 of 7	350..417	15/69 (22%) 49/69 (71%)	6.3e-11
ig: domain 5 of 7	456..516	18/64 (28%) 46/64 (72%)	1.7e-08
ig: domain 6 of 7	553..617	16/66 (24%) 39/66 (59%)	0.00011
fn3: domain 1 of 1	643..733	20/93 (22%) 53/93 (57%)	0.98
ig: domain 7 of 7	801..875	7/78 (9%) 54/78 (69%)	37

MAM: domain 1 of 1	793..958	65/180 (36%) 132/180 (73%)	1.3e-52
--------------------	----------	-------------------------------	---------

EXAMPLE 13.

The NOV13 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 13A.

Table 13A. NOV13 Sequence Analysis		
	SEQ ID NO: 51	4169 bp
NOV13a, CG57409-03 DNA Sequence	TCTTCGTCGCCGCTCTCTCTCTCACCTCTCAGGGAAAGGGGGGGACATAGGGGCGTCG CGGGGCCCCGCGAATGCGCCCCCGCCGCTCTCGGGCTGCGCCGCTCGCGGGGAT GAAGCACC GGCCGTGAAGATGGAGGTGACCTGCCTTCTACTTCTGGCGCTGATCCCCCT TCCACTGCCGGGGACAAGGAGTCTACGCTCCAGCCCAGGCGCAGATCGTGCATGCGGG CCAGGCATGTGTGGTGAAAGAGGACAATATCAGCGAGCGTGTCTACACCATCCGGGAG GGGGACACCCTCATGCTGCAGTGCCTTGTAACAGGGCACCCCTCGACCCCAGGTACGGT GGACCAAGACGGCAGGTAGCGCCTCGGACAAGTTCCAGGAGACATCGGTGTTCAACGA GACGCTGCGCATCGAGCGTATTGCACGCACGCAGGGCGGCCGCTACTACTGCAAGGCT GAGAACGGCGTGCGGGTGCCGGCCATCAAGTCCATCCGCGTGGACGTGCAGTACCTGG ATGAGCCAATGCTGACGGTGACCCAGACGGTGAGCGATGTGCGAGGCAACTTCTACCA GGAGAAGACGGTGTTCCTGCGCTGTACTGTCAACTCCAACCCGCTGCCGCTTCATC TGGAAGCGGGGTTCCGATACCCTATCCCACAGCCAGGACAATGGGGTTGACATCTATG AGCCCCCTCTACACTCAGGGGGAGACCAAGGTCCTGAAGCTGAAGAACCTGCGGCCCCA GGACTATGCCAGCTACACCTGCCAGGTGTCTGTGCGTAACGTGTGCGGCATCCCAGAC AAGGCCATCACCTTCCGGCTCACCAACACCACGGCACCCAGCCCTGAAGCTGTCTG TGAACGAAACTCTGGTGGTGAACCCTGGGGAGAATGTGACGGTGCAGTGTCTGCTGAC AGGCGGTGATCCCCCTCCCCAGCTGCAGTGGTCCCATGGGCCTGGCCCCACTGCCCTG GGTGCTCTGGCCAGGGTGGCACCTCAGCATCCCTTCAGTGCAGGCCCGGGACTCTG GCTACTACAAGTGCACAGCCACCAACAATGTGGGCAACCCTGCCAAGAAGACTGTCAA CCTGCTGGTGGGATCCATGAAGAACGCTACATTCCAGATCACTCCTGACGTGATCAAA GAGAGTGAGAACATCCAGCTGGGCCAGGACCTGAAGCTATCGTGCCACGTGGATGCAG TGCCCCAGGAGAAGGTGACCTACCAGTGGTTCAAGAATGGCAAGCCGGCACGCATGTC CAAGCGGCTGCTGGTGACCCGCAATGATCCTGAGCTGCCCCGAGTCACCAGCAGCCTA GAGCTCATTGACCTGCACTTCAGTGACTATGGCACCTACCTGTGCATGGCTTCTTTCC CAGGGGCACCCGTGCCGACCTCAGCGTCGAGGTCAACATCTCCTCTGAGACAGTGCC GCCCACCATCAGTGTGCCAAGGGTAGGGCCGTGGTGACCGTGCGCGAGGGATCGCCT GCCGAGCTGCAATGCGAGGTGCGGGGCAAGCCGCGGCCGAGTGCTCTGGTCCCGCG TGGACAAGGAGGCTGCACTGCTGCCCTCGGGGCTGCCCCCTGGAGGAGACTCCGGACGG GAAGCTGCGGCTGGAGCGAGTGAGCCGAGACATGAGCGGGACCTACCGCTGCCAGACG GCCCCGTATAATGGCTTCAACGTGCGCCCCCGTGAGGCCAGGTGCAGCTGAACGTGC AGTTCCCGCCGGAGGTGGAGCCCAGTTCCCAGGACGTGCGCCAGGCGCTGGGCCGGCC CGTGCTCCTGCGCTGCTCGCTGCTGCGAGGCAGCCCCAGCGCATCGCCTCGGCTGTG TGGCGTTTCAAAGGGCAGCTGCTGCCGCCGCCGCTGTTGTTCCCGCCGCCGCCGAGG CGCCGGATCACGCGGAGCTGCGCCTCGACGCCGTAACCTCGCGACAGCAGCGGCAGCTA CGAGTGCAGCGTCTCCAACGATGTGGGCTCGGCTGCCTGCCTCTTCCAGGTCTCCGCC AAAGCCTACAGCCCCGAGTTTTACTTCGACACCCCCAACCCACCCGACGCCACAAGC TGTCCAAGAACTACTCCTACGTGCTGCAGTGGACTCAGAGGGAGCCCGACGCTGTCTGA CCCTGTGCTCAACTACAGACTCAGCATCCGCCAGTTGAACCAGCACAAATGCGGTGGTC AAGGCCATCCCGGTCCGGCGTGTGGAGAAGGGGCAGCTGCTGGAGTACATCCTGACCG ATCTCCGTGTGCCCCACAGCTATGAGGTCCGCCTCACACCCTATACCACCTTCGGGGC TGGTGACATGGCCTCCCGCATCATCCACTACACAGAGCCCATCAACTCTCCGAACCTT TCAGACAACACCTGCCACTTTGAGGATGAGAAGATCTGTGGCTATAACCAGGACCTGA CAGACAACCTTTGACTGGACGCGGCAGAATGCCCTCACCCAGAACCCAAACGCTCCCC CAACACTGGTCCCCCACCAGACATAAGTGGCACCCCTGAGGGCTACTACATGTTTCATC GAGACATCGAGGCCTCGGGAGCTGGGGGACCGTGCAAGGTTAGTGAGTCCCCTCTACA ATGCCAGCGCCAAGTTCTACTGTGTCTCCTTCTTCTACCACATGTACGGGAAACACAT CGGCTCCCTCAACCTCCTGGTGCGGTCCCGGAACAAAGGGGCTCTGGACACGCACGCC TGGTCTCTCAGTGGCAATAAGGGCAATGTGTGGCAGCAGGCCCATGTGCCCATCAGCC	

	CCAGTGGGCCCTTCCAGATTATTTTTGAGGGGGTTCGAGGCCCGGGCTACCTGGGGGA TATTGCCATAGATGACGTCACACTGAAGAAGGGGGAGTGTCCCCGGAAGCAGACGGAT CCCAATAAAGGTGCAAGACGGGAAGGAGCTGCCTGCGATGGCCTGAAATTCCACCTTT CATCCCCTATGGATGACGGAGAGCTTACAGATGACCCTATTGAATGCAAGCACCTTTG GATCCATAGAGTGGACAGTAAAGGTGCTCAGTACATGTTGGCTGAGCTGAACTGCATA CATGTGGCCCCCAGGTTCCCTGGTCTTTATGGACGAAGGGCACAAGGTTGGTGAAAAGG ACTCCGGGGGCCAGCCCTTCCAAGTTTACACTGATTTCTCCTTTTACCCTCATGCTAT CCCTGAGAAGATGTCAATAATGCCACGTTACAGGTGGGAAAAGTCTGAGGCTTAGAGAG GAGGAGGAATCTGCCTACGGTCACACAGCTGCAAAGGCTAGAGCTGGGACCAGGAGCT GGTCTCTTAACCGACCACCTGAGCTCAAGAGCTTTTCTCTCTGGACCAACATGACCCA AAGTGTGCGCGAGCCTATCACAGGTCCCCTGCAATGCCAAACATACACGCACAGCAAT ACACAACACCTGGGGACATGGATGAAGCTGGAAACCATCATTCTCAGCAAACCTGACAC AAGAACAGAAAACCAAACACCACATGTTCTCACTCACCACCCAGTCTGCCCCGCCCTC TCTCTTCTCACCTGAACTTCCCCTCTCCTCAAACCTCTCGAGGCCACGCCTCTATGTCC TTGGATGATGATGATGACGACGACGACGATGATGATGATGATGATGACGACGATGACA ATGATGATGATGATGGAAGGAAGACCTACAGAATCCCTCCAGGCTCTGACCTCAGTGC TTGTGGGTGGGTGAATGACCACATGTGCGAGGGAGACTCCACAGGTCCTCCCGATGAG AAGCACTCTTATGCCAAAGAGGAGACTCAGGCCAAACTGACAGGACCAGGAATTAGCT ACCCTGGTAAACCCAGCTATCGACTGCACCCGAGCGGCTACACACCACTGGAGCAGTT CAGGGAGAAAGCCACCGGCATGCTCACCCCGTATGTCTCTGGCTCTGTTTCTCTTTT TGCTTCCCCTTCCCCACCTCTGAGTCTCTGTGTTCTGCTCATGCCAATTCCCCTTCTG CCTGTCTCTGCCCGCTTCTCTCTGGGCTGGTCTCTCCGAGACTCTGTTCCCTTGGCTG GCATGCCCTCCACCTCCCCTGATGGTTTACGACAGAGATGAAGCCGGCCTGGCTCATGGG TGTGGGTAATGTACTAGTGCAGGAGAGTGGTGGGGCCCAGTCTGGGTGCAG		
	ORF Start: ATG at 135	ORF Stop: TGA at 4080	
	SEQ ID NO: 52	1315 aa	MW at 145782.9kD
NOV13a, CG57409-03 Protein Sequence	MEVTCLLLLALIPFHCRCGQGVYAPAQAQIVHAGQACVVKEDNISERVYTIREGDTLML QCLVTGHPRPQVRWTKTAGSASDKFQETSVFNETLRIERIARTQGGRYYCKAENGVGV PAIKSIRVDVQYLDEPMLTVHQTVSDVRGNFYQEKTVFLRCTVNSNPAPARFIWKRGS TLSHSQDNGVDIYEPLYTQGETKVLKLNLRPDYASYTCQVSVRNVCGIPDKAITFR LTNTTAPPALKLSVNETLVNPGENVTVQCLLTGGDPLPQLQWSHGPGLPLGALAQQ GTLSPSVQARDSGYYNCTATNNVGNPAKKTVNLLVRSMKNATFQITPDVIKESENIQ LGQDLKLSCHVDAPQEKVTYQWFKNGKPARMSKRLLVTRNDPELPAVTSSLELIDLH FSDYGTLYLCMASFPGAPVPDLSVEVNISSETVPPTISVPKGRAVVTVREGSPAELQCE VRGKPRPPVLWSRVDKEAALLPSGLPLEETPDGKLRLERVSRDMSGTYRCQTARYNGF NVRPREAQVQLNVQFPPEVEPSSQDVRQALGRPVLLRCSLLRGSPQRIASAVWRFKGQ LLPPPPVVPAAAEAPDHAELRLDAVTRDSSGSYECVSNDVGSAACLFQVSAKAYSPE FYFDTPNPTRSHKLSKNYSYVLQWTQREPDVDPVLNYRLSIRQLNQHNAVVKAI PVR RVEKGQLLEYILTLRVPHSYEVRLTPYTTFGAGDMASRI IHYTEPINSPNLSDNTCH FEDEKICGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDISGTPEGYYMFIETSRPR ELGDRARLVSPLYNASAKFYCVSFFYHMYGKHIGSLNLLVRSRNKGALDTHAWSLSGN KGNVWQQAHPVISPSPGFQIIFEGVRGPGYLGDIADDDVTLKKGECPRKQTDPNKGAR REGAACDGLKFHLSSPMDDGELTDDPIECKHLWIHRVDSKGAQYMLAELNCIHVAPRF LVFMDEGHKVGEKDSGGQPFQVYTDFFSYPHAIPEKMSIMPTLQVGKLRLREEEESAY GHATAAKARAGTRSWSLNRPELKSFSLWTNMTQSVREPITGPLQCQTYTHSNTQHLGT WMKLETI ILSKLTQEQT KHHMFSLT TQSAPPSLFSPELPLSSNSRGHASMSLDDDDDD DDDDDDDDDDDDDDDDDDDDGRKTYRIPPGSGLSACGWVNDHMSQGDSTGPPDEKHSYAK EETQAKLTGPGISYPGKPSYRLHPSGYTPLEQFREKATGMLTPYVSGSVSSFCFPFPT SESLCSAHANSPSACLCPLLSGLVSPRLCSLGLWHALHLP		
	SEQ ID NO: 53	1500 bp	
NOV13b, CG57409-05 DNA Sequence	TGAGCCGAGACATGAGCGGGACCTACCGCTGCCAGACGGCCCGCTATAATGGCTTCAA CGTGCGCCCCCGTGAGGCCAGGTGCAGCTGAACGTGCAGTTCCTCGCCGGAGGTGGAG CCCAGTTCCAGGACGTGCGCCAGGCGCTGGGCCGGCCCGTGCTCCTGCGCTGCTCGC TGCTGCGAGGCAGCCCCAGCGCATCGCCTCGGCTGTGTGGCGTTTCAAAGGGCAGCT GCTGCCCGCCCGCCTGTTGTTCCCGCCGCCCGGAGGCGCCGGATCACGCGGAGCTG CGCCTCGACCGCGTAACTCGCGACAGCAGCGGCAGCTACGAGTGCAGCGTCTCCAACG ATGTGGGCTCGGCTGCCTGCCTCTTCCAGGTCTCCGCCAAAGCCTACAGCCCGGAGTT TTACTTCGACACCCCCAACCCACCCGCAGCCACAAGCTGTCCAAGAACTACTCCTAC		

	GTGCTGCAGTGGACTCAGAGGGAGCCCGACGCTGTGCGACCCTGTGCTCAACTACAGAC TCAGCATCCGCCAGTTGAACCAGCACAATGCGGTGGTCAAGGCCATCCCGGTCCGGCG TGTGGAGAAGGGGCAGCTGCTGGAGTACATCCTGACCGATCTCCGTGTGCCCCACAGC TATGAGGTCCGCCTCACACCCTATACCACCTTCGGGGCTGGTGACATGGCCTCCCGCA TCATCCACTACACAGAGCGCCAGATCCGCTGGCCCCCAGTCCTGGCTCTGAGGACCCT GTCCTCTGGTCCCAAGCAGGGTATCCTCTGCAGAGCCCCACACCTCAGTTCTGACTTG GTTTCCCCGCTTGCTTTCTCAGCCATCAACTCTCCGAACCTTTCAGACAACACCTGCC ACTTTGAGGATGAGAAGATCTGTGGCTATAACCAGGACCTGACAGACAACCTTTGACTG GACGCGGCAGAATGCCCTCACCCAGAACCCCAAACGCTCCCCAACACTGGTCCCCC ACCGACATAAGTGGCACCCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCTC GGGAGCTGGGGGACCGTGCAAGGTTAGTGAGTCCCCTCTACAATGCCAGCGCCAAGTT CTACTGTGTCTCCTTCTTCTACCACATGTACGGGAAACACATCGGCTCCCTCAACCCC CTGGTGCGGTCCCGGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCA ATAAGGGCAATGTGTGGCAGCAGGCCCATGTGCCCATCAGCCCCAGTGGGCCCTTCCA GATTATTTTTGAGGGGGTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGAC GTCACACTGAAGAAGGGGGAGTGTCCCCGGAAGCAGACGGATCCCAATAAAGTGGTGG TGATGCCGGGCAGTGGAGCCCCCTGCCAGTCCAGCCCACAGCTGTGGGGGCCCATGGC CATCTTCCTCTTGCGTTGCAGAGATGATGAGAGCTGTGTGGCCACCCCC		
	ORF Start: ATG at 12	ORF Stop: TGA at 1476	
	SEQ ID NO: 54	488 aa	MW at 54357.1kD
NOV13b, CG57409-05 Protein Sequence	MSGTYRCQTARYNGFNVRPREAQVQLNVQFPPEVEPSSQDVRQALGRPVLRLRCSLLRG SPQRIASAVWRFKGQLLPPPPVVPAAAEAPDHAELRLDAVTRDSSGSYECVSNDVGS AACLFQVSAKAYSPEFYFDTNPTRSHKLSKNYSYVLQWTQREPDAVDPVLNRYLSIR QLNQHNAVVKAI PVRRVEKGQLLEYILTDLRVPHSYEVRLTPYTTFGAGDMASRI IHY TERQIRWPPVLALRTLSSGPKQGILCRAPHLSSDLVSPLAFSAINSPNLSDNTHCFED EKICGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDISGTPEGYYMFIETSRPRELG DRARLVSPLYNASAKFYCVSFFYHMYGKHIGSLNPLVRSRNKGALDTHAWSLSGNKGN VWQQAHPVISPSPGFQIIFEGVRGPGYLGDIADDVTLKKGECPRKQTDPNKVVVMPG SGAPCQSSPQLWGPMIAIFLLALQR		
	SEQ ID NO: 55	1828 bp	
NOV13c, CG57409-06 DNA Sequence	TGAGCCGAGACATGAGCGGGACCTACCGCTGCCAGACGGCCCGCTATAATGGCTTCAA CGTGCGCCCCCGTGAGGCCAGGTGCAGCTGAACGTGCAGTTCCCGCCGGAGGTGGAG CCCAGTTCACAGGACGTGCGCCAGGCGCTGGGCGCGCCCGTGCTCCTGCGCTGCTCGC TGCTGCGAGGCAGCCCCAGCGCATCGCCTCGGCTGTGTGGCGTTTCAAAGGGCAGCT GCTGCCGCGCGCCTGTTGTTCCCGCGCGCGCGAGGCGCCGGATCACGCGGAGCTG CGCCTCGACGCCGTAACCTCGCGACAGCAGCGGCAGCTACGAGTGCAGCGTCTCCAACG ATGTGGGCTCGGCTGCCTGCCTCTTCCAGGTCTCCGCCAAAGCCTACAGCCCGGAGTT TTACTTCGACACCCCCAACCCACCCGCAGCCACAAGCTGTCCAAGAACTACTCCTAC GTGCTGCAGTGGACTCAGAGGGAGCCCGACGCTGTGCGACCCTGTGCTCAACTACAGAC TCAGCATCCGCCAGTTGAACCAGCACAATGCGGTGGTCAAGGCCATCCCGGTCCGGCG TGTGGAGAAGGGGCAGCTGCTGGAGTACATCCTGACCGATCTCCGTGTGCCCCACAGC TATGAGGTCCGCCTCACACCCTATACCACCTTCGGGGCTGGTGACATGGCCTCCCGCA TCATCCACTACACAGAGCGCCAGATCCGCTGGCCCCCAGTCCTGGCTCTGAGGACCCT GTCCTCTGGTCCCAAGCAGGGTATCCTCTGCAGAGCCCCACACCTCAGTTCTGACTTG GTTTCCCCGCTTGCTTTCTCAGCCATCAACTCTCCGAACCTTTCAGACAACACCTGCC ACTTTGAGGATGAGAAGATCTGTGGCTATAACCAGGACCTGACAGACAACCTTTGACTG GACGCGGCAGAATGCCCTCACCCAGAACCCCAAACGCTCCCCAACACTGGTCCCCC ACCGACATAAGTGGCACCCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCTC GGGAGCTGGGGGACCGTGCAAGGTTAGTGAGTCCCCTCTACAATGCCAGCGCCAAGTT CTACTGTGTCTCCTTCTTCTACCACATGTACGGGAAACACATCGGCTCCCTCAACCCC CTGGTGCGGTCCCGGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCA ATAAGGGCAATGTGTGGCAGCAGGCCCATGTGCCCATCAGCCCCAGTGGGCCCTTCCA GATTATTTTTGAGGGGGTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGAC GTCACACTGAAGAAGGGGGAGTGTCCCCGGAAGCAGACGGATCCCAATAAAGGTGCAA GACGGGAAGGAGGTGGGGGAGCTGAATCTGGAGGGAGCTGTGCGTGCGGGGGTTCTCT GTCTGTTGAGGGAGGGTGTTCGGGTCTGAATAGGGGTTTCACTGTCTGATGATGGGA ATCAGGTGGCTCTGACTGTGTTAACGTGTGCCCACTACGTGAGGCTGAGAACTG GTGTAACACCATGAGAAAGCTTGCCCCCACCATCGTGATGAGCATACCGACCTGGTC		

	ACCGGAACACAAACACCAACAACCACAGAGGGCGCCTCAGAATACCCAGAGGGCCCAA TACGCCGACCCGCTGTCACGAGCGCCACGAGCGGCAGAACACGACAGGCACACAACC AGCCGGAGCAAGACGGAGCCGAGAGCCCCGGGGACATAGACCCCAGCAAGCGACACAC AAGGACGCGCACAGAGCGCACACACTAACA		
	ORF Start: ATG at 12	ORF Stop: TGA at 1521	
	SEQ ID NO: 56	503 aa	MW at 55764.4kD
NOV13c, CG57409-06 Protein Sequence	MSGTYRCQTARYNGFNVRPREAQVQLNVQFPPEVEPSSQDVRQALGRPVLLRCSLLRG SPQRIASAVWRFGQLLPPPPVPAAAEAPDHAELRLDAVTRDSSGSYECSVSNVGS AACLFQVSAKAYSPEFYFDTNPNTRSHKLSKNYSYVLQWTQREPDAVDPVLNYRLSIR QLNQHNAVVKAI PVRRVEKGQLLEYILTDLRVPHSYEVRLTPYTTFGAGDMASRI IHY TERQIRWPPVLALRTLSSGPKQGILCRAPHLSSDLVSPLAFSAINSPNLSDNTHFED EKICGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDISGTPEGYYMFIETSRPRELG DRARLVSPLYNASAKFYCVSFFYHMYGKHIGSLNPLVRSRNKGALDTHAWSLSGNKGN VWQQAHPVISPSPGFQIIFEGVRGPGYLGDI AIDDVTLKKGECPRKQTDPNKGARREG GGGAESGGSCAWRGFLSVEGGCSGLNRGSDCLMMGIRWL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 13B.

Table 13B. Comparison of NOV13a against NOV13b through NOV13c.		
Protein Sequence	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV13b	508..925 1..458	403/458 (87%) 403/458 (87%)
NOV13c	508..925 1..458	403/458 (87%) 403/458 (87%)

Further analysis of the NOV13a protein yielded the following properties shown in Table 13C.

Table 13C. Protein Sequence Properties NOV13a	
PSort analysis:	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 19 and 20

- 5 A search of the NOV13a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 13D.

Table 13D. Geneseq Results for NOV13a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

AAE00582	Human nuclear cell adhesion molecule homologue, NCAM_d_1 protein - Homo sapiens, 946 aa. [WO200129215-A2, 26-APR-2001]	23..919 15..912	488/906 (53%) 657/906 (71%)	0.0
AAE00581	Human cell adhesion molecule homologue (CAM-H) protein #1 - Homo sapiens, 1018 aa. [WO200129215-A2, 26-APR-2001]	23..919 15..912	488/906 (53%) 657/906 (71%)	0.0
AAE00586	Human nuclear cell adhesion molecule homologue, NCAM_d_2 protein - Homo sapiens, 891 aa. [WO200129215-A2, 26-APR-2001]	71..919 8..857	456/858 (53%) 618/858 (71%)	0.0
AA Y72717	HBXDJ03 clone human attractin-like protein #2 - Homo sapiens, 448 aa. [WO200116156-A1, 08-MAR-2001]	508..925 1..418	418/418 (100%) 418/418 (100%)	0.0
AA Y72714	HBXDJ03 clone human attractin-like protein #1 - Homo sapiens, 448 aa. [WO200116156-A1, 08-MAR-2001]	508..925 1..418	408/418 (97%) 408/418 (97%)	0.0

In a BLAST search of public sequence databases, the NOV13a protein was found to have homology to the proteins shown in the BLASTP data in Table 13E.

Table 13E. Public BLASTP Results for NOV13a				
Protein Accession Number	Protein/Organism/Length	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAB86654	DJ402N21.3 (NOVEL PROTEIN WITH IMMUNOGLOBULIN DOMAINS) - Homo sapiens (Human), 299 aa (fragment).	239..536 1..299	298/299 (99%) 298/299 (99%)	e-172
CAB86653	DJ402N21.2 (NOVEL PROTEIN WITH MAM DOMAIN) - Homo sapiens (Human), 273 aa (fragment).	683..925 1..243	243/243 (100%) 243/243 (100%)	e-145
Q9DBX0	1200011I03RIK PROTEIN - Mus musculus (Mouse), 267 aa.	689..925 1..237	228/237 (96%) 233/237 (98%)	e-136
Q9GMT4	HYPOTHETICAL 51.2 KDA PROTEIN - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 448 aa.	508..919 1..414	206/421 (48%) 282/421 (66%)	e-115
CAB86655	DJ402N21.1 (NOVEL PROTEIN) - Homo sapiens (Human), 127 aa (fragment).	1..127 1..127	127/127 (100%) 127/127 (100%)	3e-68

PFam analysis predicts that the NOV13a protein contains the domains shown in the Table 13F.

Table 13F. Domain Analysis of NOV13a			
Pfam Domain	NOV13a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ig: domain 1 of 7	53..110	14/61 (23%) 42/61 (69%)	2.5e-08
ig: domain 2 of 7	150..216	14/70 (20%) 51/70 (73%)	3.7e-09
ig: domain 3 of 7	255..310	18/58 (31%) 38/58 (66%)	2.4e-08
PKD: domain 1 of 1	239..327	22/100 (22%) 56/100 (56%)	7.3
ig: domain 4 of 7	350..417	15/69 (22%) 49/69 (71%)	6.3e-11
ig: domain 5 of 7	456..516	18/64 (28%) 46/64 (72%)	1.7e-08
ig: domain 6 of 7	553..617	16/66 (24%) 39/66 (59%)	0.00011
fn3: domain 1 of 1	643..733	20/93 (22%) 53/93 (57%)	0.98
ig: domain 7 of 7	761..835	7/78 (9%) 54/78 (69%)	37
MAM: domain 1 of 1	753..918	65/180 (36%) 132/180 (73%)	1.3e-52

EXAMPLE 14.

The NOV14 clone was analyzed, and the nucleotide and predicted polypeptide
5 sequences are shown in Table 14A.

Table 14A. NOV14 Sequence Analysis			
	SEQ ID NO: 57	330 bp	
NOV14a, CG59262-01 DNA Sequence	<u>GGAGTGGTCAGTTCTGCTGCCGACACGCCACCCAGCTCGAGATGGCCATGGACACCA</u> TGATTAGAATCTTCCACCGCTATTCTGGCAAGGCAAGGAAGAGATTCAAGCTCAGCAA GGGGGAACTGAAACTGCTCCTGCAGCGAGAGCTCACGGAATTCCTCTCGTGCCAAAAG GAAACCCAGTTGGTTGATAAGATAGTGCAGGACCTGGATGCCAATAAGGACAACGAAG TGGATTTTAATGAATTCTGGTTCATGGTGGCAGCTCTGACAGTTGCTTGTAATGATTA CTTTGTAGAACAATTGAAGAAGAAAGGAAAATAAAGGTAA		
	ORF Start: ATG at 43	ORF Stop: TAA at 322	
	SEQ ID NO: 58	93 aa	MW at 10861.6kD

NOV14a, CG59262-01 Protein Sequence	MAMDTMIRIFHRYSGKARKRFKLSKGELKLLQRELTEFLSCQKETQLVDKIVQDLDA NKDNEVDFNEFVVMVAALTVACNDYFVEQLKKKGK
---	--

Further analysis of the NOV14a protein yielded the following properties shown in Table 14B.

Table 14B. Protein Sequence Properties NOV14a	
PSort analysis:	0.7000 probability located in plasma membrane; 0.5337 probability located in mitochondrial inner membrane; 0.3627 probability located in mitochondrial intermembrane space; 0.2997 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV14a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
5 homologous proteins shown in Table 14C.

Table 14C. Geneseq Results for NOV14a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM40258	Human polypeptide SEQ ID NO 3403 - Homo sapiens, 94 aa. [WO200153312-A1, 26-JUL-2001]	2..86 8..92	50/85 (58%) 66/85 (76%)	3e-23
AAB45531	Human S100A1 protein - Homo sapiens, 94 aa. [DE19915485-A1, 19-OCT-2000]	2..86 8..92	50/85 (58%) 66/85 (76%)	3e-23
ABB12007	Human Ca-binding protein S100P homologue, SEQ ID NO:2377 - Homo sapiens, 113 aa. [WO200157188-A2, 09-AUG-2001]	2..84 25..107	43/83 (51%) 59/83 (70%)	3e-18
AAB45545	Human S100P protein - Homo sapiens, 95 aa. [DE19915485-A1, 19-OCT-2000]	2..84 7..89	43/83 (51%) 59/83 (70%)	3e-18
AAB45544	Human S100B protein - Homo sapiens, 95 aa. [DE19915485-A1, 19-OCT-2000]	2..84 7..89	43/83 (51%) 59/83 (70%)	3e-18

In a BLAST search of public sequence databases, the NOV14a protein was found to have homology to the proteins shown in the BLASTP data in Table 14D.

Table 14D. Public BLASTP Results for NOV14a				
Protein Accession Number	Protein/Organism/Length	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAL30893	S100Z PROTEIN - Homo sapiens (Human), 99 aa.	1..93 7..99	93/93 (100%) 93/93 (100%)	4e-47
S35985	S-100 protein alpha chain - weatherfish, 95 aa.	2..89 7..94	52/88 (59%) 70/88 (79%)	3e-25
P35467	S-100 protein, alpha chain - Rattus norvegicus (Rat), 93 aa.	2..86 7..91	52/85 (61%) 66/85 (77%)	4e-23
BCBOIA	S-100 protein alpha chain - bovine, 94 aa.	2..86 8..92	50/85 (58%) 66/85 (76%)	1e-22
CAC16547	SEQUENCE 1 FROM PATENT WO0061742 - Homo sapiens (Human), 94 aa.	2..86 8..92	50/85 (58%) 66/85 (76%)	1e-22

PFam analysis predicts that the NOV14a protein contains the domains shown in the Table 14E.

Table 14E. Domain Analysis of NOV14a			
Pfam Domain	NOV14a Match Region	Identities/ Similarities for the Matched Region	Expect Value
S_100: domain 1 of 1	2..42	20/44 (45%) 31/44 (70%)	2.8e-09
efhand: domain 1 of 1	48..76	6/29 (21%) 25/29 (86%)	0.0012

EXAMPLE 15.

The NOV15 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 15A.

Table 15A. NOV15 Sequence Analysis		
	SEQ ID NO: 59	773 bp
NOV15a, CG58635-01 DNA Sequence	AGCCTTGGGTCTGAAGGGATGAGGTGGGGCCTCCTTCAAGAGACAAAGTCTGGTTCTGT CCGTGGGTCTCTGTCCCTACAGAAAAGGAGAACAACCTCCCGCCACTGCCCAAGTTC ATCCCTGTGAAGCCCTGCTTCTACCAGAACTTCTCCGACGAGATCCCAGTGGAGCACC AGGTCCTGGTGAAGAGGATCTACCGCTGTGGATGGTTTACTGCGCCACCCTCGGCGT CAACCTCATTGCCTGCCTGGCCTGGTGGATCGGCGGAGGCTCGGGGACCAACTTCGGC CTGGCCTTCGTGTGGCTGCTCCTGTTACGCTTTCGGCTACGTGTGCTGGTTCCGGC CTGTCTACAAGGCCTTCCGGGCCGACAGCTCCTTTAATTTTCATGGCGTTTTTCTTCAT CTTCGGAGCCCAGTTTGTCTGACCGTCATCCAGGCGATTGGCTTCTCCGGCTGGGGC GCGTGCGGCTGGCTGTGGCAATTGGATTCTTCCAGTACAGCCGGGCGCTGCCGTGG TCATGCTGCTTCCAGCCATCATGTTCTCCGTGTGGCTGCCATGATGGCCATCGCGAT	

	CATGAAGGTGCACAGGATCTACCGAGGGGCTGGCGGAAGCTTCCAGAAGGCACAGACG GAGTGGAACACGGGCACTTGGCGGAACCCACCGTCGAGGGAGGCCAGTACAACAAC TCTCAGGCAACAGCCTGCCCCGAGTACCCCACTGTGCCAGCTACCCGGGCAGTGGCCA GTGGCCTTAGAGGGAGCCT		
	ORF Start: ATG at 18	ORF Stop: TAG at 762	
	SEQ ID NO: 60	248 aa	MW at 27780.0kD
NOV15a, CG58635-01 Protein Sequence	MRWGLLQETKSGSVRGFSVPTEKENNFPPLPKFI PVKPCFYQNFSDIIPVEHQVLVKR IYRLWMVYCATLGVNLIACLAWWIGGSGTNFGLAFVWLLLFPCGYVCWFRPVYKAF RADSSFNFMAFFFI FGAQFVLTVIQAIGFSGWGACGWLSAIGFFQYSPGAAVVMLLPA IMFSVSAAMMAIAIMKVHRIYRGAGGSFQKAQTEWNTGTWRNPPSREAQYNNFSGNSL PEYPTVPSYPGSGQWP		
	SEQ ID NO: 61	773 bp	
NOV15b, CG58635-02 DNA Sequence	AGCCTTGGGTTGAAGGGATGAGGTGGGGCCTCCTTCAAGAGACAAAGTCTGGTTCTGT CCGTGGGTCTCTGTCCCTACAGAAAAGGAGAACTTCCCGCCACTGCCCAAGTTC ATCCCTGTGAAGCCCTGCTTCTACCAGAACTTCTCCGACGAGATCCCAGTGGAGCACC AGGTCCTGGTGAAGAGGATCTACCGGCTGTGGATGTTTTACTGCGCCACCCTCGGCGT CAACCTCATTGCCTGCCTGGCCTGGTGGATCGGCGGAGGCTCGGGGACCAACTTCGGC CTGGCCTTCGTGTGGCTGCTCCTGTTACGCCTTGGCGCTACGTGTGCTGGTTCGGC CTGTCTACAAGGCCTTCCGAGCCGACAGCTCCTTTAATTTATGGCGTTTTTCTTCAT CTTCGGAGCCCAAGTTTGTCTGACCGTCATCCAGGCGATTGGCTTCTCCGGCTGGGGC GCGTGCGGCTGGCTGTGCGCAATTGGATTCTTCCAGTACAGCCCGGGCGCTGCCGTGG TCATGCTGCTTCCAGCCATCATGTTCTCCGTGTGCGCTGCCATGATGGCCATCGCGAT CATGAAGGTGCACAGGATCTACCGAGGGGCTGGCGGAAGCTTCCAGAAGGCACAGACG GAGTGGAACACGGGCACTTGGCGGAACCCACCGTCGAGGGAGGCCAGTACAACAAC TCTCAGGCAACAGCCTGCCCCGAGTACCCCACTGTGCCAGCTACCCGGGCAGTGGCCA GTGGCCTTAGAGGGAGCCT		
	ORF Start: ATG at 18	ORF Stop: TAG at 762	
	SEQ ID NO: 62	248 aa	MW at 27828.1kD
NOV15b, CG58635-02 Protein Sequence	MRWGLLQETKSGSVRGFSVPTEKENNFPPLPKFI PVKPCFYQNFSDIIPVEHQVLVKR IYRLWMFYCATLGVNLIACLAWWIGGSGTNFGLAFVWLLLFPCGYVCWFRPVYKAF RADSSFNFMAFFFI FGAQFVLTVIQAIGFSGWGACGWLSAIGFFQYSPGAAVVMLLPA IMFSVSAAMMAIAIMKVHRIYRGAGGSFQKAQTEWNTGTWRNPPSREAQYNNFSGNSL PEYPTVPSYPGSGQWP		
	SEQ ID NO: 63	654 bp	
NOV15c, CG58635-03 DNA Sequence	ATGAGGTGGGGCCTCCTTCAAGAGACAAAGTCTGGTTCTGTCCGTGGGTTCGCGTCC CTACAGAAAAGGAGAACTTCCCGCCACTGCCCAAGTTCATCCCTGTGAAGCCCTG CTTCTACCAGAACTTCTCCGACGAGATCCCAGTGGAGCACCAGGTCCTGGTGAAGAGG ATCTACCGGCTGTGGATGTTTTACTGCGCCACCCTCGGCGTCAACCTCATTGCCTGCC TGGCCTGGTGGATCGGCGGAGGCTCGGGGACCAACTTCGGCCTGGCCTTCGTGTGGCT GCTCCTGTTACGCCTTGGCGCTACGTGTGCTGGTTCGGCCTGTCTACAAGGCCTTC CGCGGCTGGCTGTGCGCAATTGGATTCTTCCAGTACAGCCCGGGCGCTGCCGTGGTCA TGCTGCTTCCAGCCATCATGTTCTCCGTGTGCGCTGCCATGATGGCCATCGCGATCAT GAAGGCGCACAGGATCTACCGAGGGGCTGGCGGAAGCTTCCAGAAGGCACAGACGGAG TGGAACACGGGCACTTGGCGGAACCCACCGTCGAGGGAGGCCAGTACAACAACCTTCT CAGGCAACAGCCTGCCCCGAGTACCCCACTGTGCCAGCTACCCGGGCAGTGGCCAGTG GCCTTAGAGGGAGCCT		
	ORF Start: ATG at 1	ORF Stop: TAG at 643	
	SEQ ID NO: 64	214 aa	MW at 24129.8kD
NOV15c, CG58635-03 Protein Sequence	MRWGLLQETKSGSVRGFPVPTEKENNFPPLPKFI PVKPCFYQNFSDIIPVEHQVLVKR IYRLWMFYCATLGVNLIACLAWWIGGSGTNFGLAFVWLLLFPCGYVCWFRPVYKAF RGWLSAIGFFQYSPGAAVVMLLPAIMFSVSAAMMAIAIMKAHRIYRGAGGSFQKAQTE WNTGTWRNPPSREAQYNNFSGNSLPEYPTVPSYPGSGQWP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 15B.

Table 15B. Comparison of NOV15a against NOV15b through NOV15c.		
Protein Sequence	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV15b	1..248	217/248 (87%)
	1..248	217/248 (87%)
NOV15c	1..248	191/248 (77%)
	1..214	191/248 (77%)

Further analysis of the NOV15a protein yielded the following properties shown in Table 15C.

Table 15C. Protein Sequence Properties NOV15a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 17 and 18

- 5 A search of the NOV15a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 15D.

Table 15D. Geneseq Results for NOV15a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM93439	Human polypeptide, SEQ ID NO: 3078 - Homo sapiens, 229 aa. [EP1130094-A2, 05-SEP-2001]	21..248 2..229	226/228 (99%) 227/228 (99%)	e-138
AAM93704	Human polypeptide, SEQ ID NO: 3635 - Homo sapiens, 132 aa. [EP1130094-A2, 05-SEP-2001]	21..150 2..131	127/130 (97%) 129/130 (98%)	1e-75
AAM25225	Human protein sequence SEQ ID NO:740 - Homo sapiens, 185 aa. [WO200153455-A2, 26-JUL-2001]	21..131 35..145	109/111 (98%) 110/111 (98%)	1e-64
AAY11904	Human 5' EST secreted protein SEQ ID No: 504 - Homo sapiens, 108 aa. [WO9906550-A2, 11-FEB-1999]	21..126 2..107	102/106 (96%) 103/106 (96%)	7e-60

AAB62698	Human membrane recycling protein (HMRP)-1 - Homo sapiens, 347 aa. [US6235715-B1, 22-MAY-2001]	23..229 131..338	102/208 (49%) 140/208 (67%)	5e-56
----------	---	---------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV15a protein was found to have homology to the proteins shown in the BLASTP data in Table 15E.

Table 15E. Public BLASTP Results for NOV15a				
Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q969E2	HYPOTHETICAL 25.7 KDA PROTEIN (SIMILAR TO SECRETORY CARRIER MEMBRANE PROTEIN 4) - Homo sapiens (Human), 229 aa.	21..248 2..229	226/228 (99%) 227/228 (99%)	e-138
Q9ET20	SECRETORY CARRIER MEMBRANE PROTEIN 4 - Rattus norvegicus (Rat), 230 aa.	23..248 4..230	193/227 (85%) 208/227 (91%)	e-118
Q9JKV5	SECRETORY CARRIER MEMBRANE PROTEIN 4 - Mus musculus (Mouse), 230 aa.	23..248 4..230	190/227 (83%) 208/227 (90%)	e-117
Q9JKE3	SECRETORY CARRIER MEMBRANE PROTEIN 5 - Rattus norvegicus (Rat), 235 aa.	22..246 3..234	135/232 (58%) 167/232 (71%)	2e-81
Q9JKD3	SECRETORY CARRIER MEMBRANE PROTEIN 5 - Mus musculus (Mouse), 235 aa.	22..246 3..234	134/232 (57%) 166/232 (70%)	7e-81

PFam analysis predicts that the NOV15a protein contains the domains shown in the Table 15F.

Table 15F. Domain Analysis of NOV15a			
Pfam Domain	NOV15a Match Region	Identities/ Similarities for the Matched Region	Expect Value
TspO_MBR: domain 1 of 1	63..190	30/164 (18%) 91/164 (55%)	9.5
chloroa_b-bind: domain 1 of 1	181..195	5/15 (33%) 12/15 (80%)	3.7

EXAMPLE 16.

The NOV16 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 16A.

Table 16A. NOV16 Sequence Analysis			
	SEQ ID NO: 65	1642 bp	
NOV16a, CG59209-01 DNA Sequence	GCGGAGTCCGGACGTCGGGAGCAGGATGGCGGCGGAGCAGGACCCCGAGGCGCGCGCG GGGCGGCCGCTGCTCACTGACCTCTACCAGGCCACCATGGCGTTGGGCTATTGGCGCG CGGGCCGGGCGCGGGACGCCGCCGAGTTGAGCTCTTCTTCCGCCGCTGCCCGTTCCG CGGCGCCTTCGCCTTGGTAGCCGGCTTGCGCGACTGTGTGCGCTTCCTGCGCGCCTTC GACGTGCAGTTCTTGGCCTCGGTGCTGCCCCAGACACGGATCCTGCGTTCTTCGAGC ACCTTCGGGCCCCTCGACTGCTCCGAGGTGACGGTGCGAGCCCTGCCCGAGGCTCCCTC GCCTTCCCCGCAGGTGCCGCTCCTGCAGGTGTCCGGGCCGCTCCTGGTGGTGCAGCTG CTGGAGACACCGCTGCTCTGCCTGGTCAGCTACGCCAGCCTGGTGGCCACCAACGCAG CGCGCGTTTCGCTTGATCGCAGGGCCAGAGAAGCGGCTGCTAGAGATGGGCCTGAGGCG GGCTCAGGGCCCCGATGGGGCCTGACAGCCTCCACCTACAGCTACCTGGGGGCTTCGAC AGCAGCAGCAACGTGCTAGCGGGCCAGCTGCGAGGTGTGCCGGTGGCCGGGACCCCTGG CCCACTCCTTCGTCACTTCCTTTTCAGGCAGCGAGGTGCCCCCTGACCCGATGTTGGC GCCAGCAGCTGGTGAGGGCCCTGGGGTGGACCTGGCGGCCAAAGCCCAGGTGTGGCTG GAGCAGGTGTGTGCCACCTGGGGCTGGGGGTGCAGGAGCCGCATCCAGGCGAGCGGG CAGCCTTTGTGGCCTATGCCTTGGCTTTTCCCCGGGCCTTCCAGGGCCTCCTGGACAC CTACAGCGTGTGGAGGAGTGGTCTCCCCAACTTCCTAGCAGTCGCCTTGGCCCTGGGA GAGCTGGGCTACCGGGCAGTGGGCGTGAGGCTGGACAGTGGTGACCTGCTACAGCAGG CTCAGGAGATCCGCAAGGTCTTCCGAGCTGCTGCAGCCCAGTTCAGGTGCCCTGGCT GGAGTCAGTCCTCATCGTAGTCAGCAACAACATTGACGAGGAGGCGCTGGCCCGACTG GCCCAGGAGGGCAGTGAGGTGAATGTGATTGGCATTGGCACCAGTGTGGTCACCTGCC CCCAACAGCCTTCCCTGGGTGGTGTCTATAAGCTGGTGGCCGTGGGGGGCCAGCCACG AATGAAGCTGACCGAGGACCCCGAGAAGCAGACGTTGCCTGGGAGCAAGGCTGCTTTC CGGCTCCTGGGCTCTGACGGGTCTCCACTCATGGACATGCTGCAGTTAGCAGAAGAGC CAGTGCCACAGGCTGGGCAGGAGCTGAGGGTGTGGCCTCCAGGGGCCAGGAGCCCTG CACCGTGAGGCCAGCCAGGTGGAGCCACTACTGCGGCTCTGCCTCCAGCAGGGACAG CTGTGTGAGCCGCTCCCATCCCTGGCAGAGTCTAGAGCCTTGGCCCAGCTGTCCCTGA GCCGACTCAGCCCTGAGCACAGGCGGCTGCGGAGCCCTGCACAGTACCAGGTGGTGCT GTCCGAGAGGCTGCAGGCCCTGGTGAACAGTCTGTGTGCGGGGCAGTCCCCCTGAGAC TCGGAGCGGGGCTGACTG		
	ORF Start: ATG at 26	ORF Stop: TGA at 1619	
	SEQ ID NO: 66	531 aa	MW at 56889.8kD
NOV16a, CG59209-01 Protein Sequence	MAAEQDPEARAGRPLLTDLYQATMALGYWRAGRARDAAEFELFFRRCPFGGAFALVAG LRDCVRFLRAFVQFLASVLPDTPAFFEHLRALDCSEVTVRALPEAPSPSPQVPLL QVSGPLLVVQLLETPLLCLVSYASLVATNAARVRLIAGPEKRLLEMGLRRAQGPMGPD SLHLQLPGGFDSSSNVLQGLRGVVPVAGTLAHSFVTSFSGSEVPDPMLAPAAGEGPG VDLAAKAQVWLEQVCAHLGLGVQEPHPGERAAAFVAYALAFPRAFQGLLDTYSVWRSGL PNFLAVALALGELGYRAVGVRDLSDLLQQAQEIRKVFRAAAQFQVPWLESVLIVVS NNIDEEALARLAQEGSEVNVIGIGTSVVTCPQQPSLGGVYKLVAVGGQPRMKLTEDPE KQTLPGSKAAFRLLGSDGSPLMDMLQLAEFPVPQAGQELRVWPPGAQEPCTVRPAQVE PLLRLCLQQQLCEPLPSLAESRALAQLSLSRLSPEHRRLRSPAQYQVVLSERLQALV NSLCAGQSP		
	SEQ ID NO: 67	1179 bp	
NOV16b, 174308417 DNA Sequence	AGATCTACCAACGCAGCGCGCTTCGCTTGATCGCAGGGCCAGAGAAGCGGCTGCTAG AGATGGGCCTGAGGCGGGCTCAGGGCCCCGATGGGGGCTGACAGCCTCCACCTACAG CTACCTGGGCGGCTTCGACAGCAGCAGCAACGTGCTAGCGGGCCAGCTGCGAGGTGTG CCGGTGGCCGGGACCCTGGCCCACTCCTTCGTCACTTCCTTTTCAGGCAGCGAGGTGC CCCCTGACCCGATGTTGGCGCCAGCAGCTGGTGAAGGCCCTGGGGTGGACCTGGCGGC CAAAGCCCAGGTGTGGCTGGAGCAGGTGTGTGCCACCTGGGGCTGGGGGTGCAGGAG		

	CCGCATCCAGGCGAGCGGGCAGCCTTTGTGGCCTATGCCTTGGCTTTTCCCCGGGCCT TCCAGGGCCTCCTGGACACCTACAGCGTGTGGAGGAGTGGTCTCCCCAACTTCCTAGC AGTCGCCCTGGCCCTGGGAGAGCTGGGCTACCGGGCAGTGGGCGTGAGGCTGGACAGT GGTGACCTGCTACAGCAGGCTCAGGAGATCCGCAAGGTCTTCCGAGCTGCTGCAGCCC AGTTCCAGGTGCCCTGGCTGGAGTCAGTCCTCATCGTAGTCAGCAACAACATTGACGA GGAGGCGCTGGCCCGACTGGCCCAGGAGGGCAGTGAGGTGAATGTCATTGGCATTGGC ACCAGTGTGGTCACCTGCCCCCAACAGCCTTCCCTGGGTGGCGTCTATAAGCTGGTGG CCGTGGGGGGCCAGCCACGAATGAAGCTGACCGAGGACCCCGAGAAGCAGACGCTGCC TGGGAGCAAGGCTGCTTTCCGGCTCCTGGGCTCTGACGGGTCTCCACTCATGGACATG CTGCAGTTAGCAGAAGAGCCAGTGCCACAGGCTGGGCAGGAGCTGAGGGTGTGGCCTC CAGGGGCCCAGGAGCCCTGCACCGTGAGGCCAGCCCAGGTGGAGCCACTACTGCGGCT CTGCCTCCAGCAGGGACAGCTGTGTGAGCCGCTCCCATCCCTGGCAGAGTCTAGAGCC TTGGCCCAGCTGTCCCTGAGCCGACTCAGCCCTGAGCACAGGCGGCTGCGGAGCCCTG CACAGTACCAGGTGGTGCTGTCCGAGAGGCTGCAGGCCCTGGTGAACAGTCTGTGTGC GGGGCAGTCCCCCTCGAG		
	ORF Start: AGA at 1	ORF Stop: at 1180	
	SEQ ID NO: 68	393 aa	MW at 41797.4kD
NOV16b, 174308417 Protein Sequence	RSTNAARVRLIAGPEKRLLEMGLRRAQGPDGGLTASTYSYLGGFDSSSNVLAGQLRGV PVAGTLAHSFVTSFSGSEVPPDPMLAPAAGEGPGVDLAAKAQVWLEQVCAHLGLGVQE PHPGERAAAFVAYALAFPRAFQGLLDITYSVWRSGLPNFLAVALALGELGYRAVGVRLDS GDLLQQAQEIRKVFRAAAAQFQVPWLESVLIVSNNIDEEALARLAQEGSEVNVIGIG TSVVTCPQQPSLGGVYKLVAVGGQPRMKLTEDPEKQTLPGSKAAFRLLGSDGSPLMDM LQLAEFPVPQAGQELRVWPPGAQEPCTVRPAQVEPLLRLCLQQGQLCEPLPSLAESRA LAQLSLSRLSPEHRRLRSPAQYQVVLSERLQALVNSLCAGQSPLE		
	SEQ ID NO: 69	1179 bp	
NOV16c, 174308429 DNA Sequence	AGAT CTACCAACGCAGCGCGCTTCGCTTGATCGCAGGGCCAGAGAAGCGGCTGCTAG AGATGGGCCTGAGGCGGGCTCAGGGCCCCGATGGGGGCCTGACAGCCTCCACCTACAG CTACCTGGGCGGCTTCGACAGCAGCAGCAACGTGCTAGCGGGCCAGCTGCGAGGTGTG CCGGTGGCCGGGACCCTGGCCCACTCCTTCGTCACTTCCTTTTCAGGCAGCGAGGTGC CCCCTGACCCGATGTTGGCGCCAGCAGCTGGTGAGGGCCCTGGGGTGGACCTGGCGGC CAAAGCCCAGGTGTGGCTGGAGCAGGTGTGTGCCACCTGGGGCTGGGGTGCAGGAG CCGCATCCAGGCGAGCGGGCAGCCTTTGTGGCCTATGCCTTGGCTTTTCCCCGGGCCT TCCAGGGCCTCCTGGACACCTACAGCGTGTGGAGGAGTGGTCTCCCCAACTTCCTAGC AGTCGCCTTGGCCCTGGGAGAGCTGGGCTACCGGGCAGTGGGCGTGAGGCTGGACAGT GGTGACCTGCTACAGCAGGCTCAGGAGATCCGCAAGGTCTTCCGAGCTGCTGCAGCCC AGTTCCAGGTGCCCTGGCTGGAGTCAGTCCTCATCGTAGTCAGCAACAACATTGACGA GGAGGCGCTGGCCCGACTGGCCCAGGAGGGCAGTGAGGTGAATGTCATTGGCATTGGC ACCAGTGTGGTCACCTGCCCCCAACAGCCTTCCCTGGGTGGCGTCTATAAGCTGGTGG CCGTGGGGGGCCAGCCACGAATGAAGCTGACCGAGGACCCCGAGAAGCAGACGTTGCC TGGGAGCAAGGCTGCTTTCCGGCTCCTGGGCTCTGACGGGTCTCCACTCATGGACATG CTGCAGTTAGCAGAAGAGCCAGTGCCACAGGTTGGGCAGGAGCTGAGGGTGTGGCCTC CAGGGGCCCAGGAGCCCTGCACCGTGAGGCCAGCCCAGGTGGAGCCACTACTGCGGCT CTGCCTCCAGCAGGGACAGCTGTGTGAGCCGCTCCCATCCCTGGCAGAGTCTAGAGCC TTGGCCCAGCTGTCCCTGAGCCGACTCAGCCCTGAGCACAGGCGGCTGCGGAGCCCTG CACAGTACCAGGTGGTGCTGTCCGAGAGGCTGCAGGCCCTGGTGAACAGTCTGTGTGC GGGGCAGTCCCCCTCGAG		
	ORF Start: AGA at 1	ORF Stop: at 1180	
	SEQ ID NO: 70	393 aa	MW at 41825.4kD
NOV16c, 174308429 Protein Sequence	RSTNAARVRLIAGPEKRLLEMGLRRAQGPDGGLTASTYSYLGGFDSSSNVLAGQLRGV PVAGTLAHSFVTSFSGSEVPPDPMLAPAAGEGPGVDLAAKAQVWLEQVCAHLGLGVQE PHPGERAAAFVAYALAFPRAFQGLLDITYSVWRSGLPNFLAVALALGELGYRAVGVRLDS GDLLQQAQEIRKVFRAAAAQFQVPWLESVLIVSNNIDEEALARLAQEGSEVNVIGIG TSVVTCPQQPSLGGVYKLVAVGGQPRMKLTEDPEKQTLPGSKAAFRLLGSDGSPLMDM LQLAEFPVPQVGQELRVWPPGAQEPCTVRPAQVEPLLRLCLQQGQLCEPLPSLAESRA LAQLSLSRLSPEHRRLRSPAQYQVVLSERLQALVNSLCAGQSPLE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 16B.

Table 16B. Comparison of NOV16a against NOV16b through NOV16c.		
Protein Sequence	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV16b	143..531	351/391 (89%)
	2..391	353/391 (89%)
NOV16c	143..531	350/391 (89%)
	2..391	352/391 (89%)

Further analysis of the NOV16a protein yielded the following properties shown in Table 16C.

Table 16C. Protein Sequence Properties NOV16a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.2864 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV16a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 16D.

Table 16D. Geneseq Results for NOV16a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG33687	Arabidopsis thaliana protein fragment SEQ ID NO: 40861 - Arabidopsis thaliana, 553 aa. [EP1033405-A2, 06-SEP-2000]	14..531 4..548	231/547 (42%) 330/547 (60%)	e-113
AAG33686	Arabidopsis thaliana protein fragment SEQ ID NO: 40860 - Arabidopsis thaliana, 574 aa. [EP1033405-A2, 06-SEP-2000]	14..531 25..569	231/547 (42%) 330/547 (60%)	e-113
AAG33685	Arabidopsis thaliana protein fragment SEQ ID NO: 40859 - Arabidopsis thaliana, 591 aa. [EP1033405-A2, 06-SEP-2000]	14..531 42..586	231/547 (42%) 330/547 (60%)	e-113

AAY74114	Human prostate tumor EST fragment derived protein #301 - Homo sapiens, 223 aa. [DE19820190-A1, 04-NOV-1999]	334..531 26..223	197/198 (99%) 198/198 (99%)	e-109
AAG29216	Arabidopsis thaliana protein fragment SEQ ID NO: 34723 - Arabidopsis thaliana, 435 aa. [EP1033405-A2, 06-SEP-2000]	14..474 4..432	200/468 (42%) 278/468 (58%)	1e-95

In a BLAST search of public sequence databases, the NOV16a protein was found to have homology to the proteins shown in the BLASTP data in Table 16E.

Table 16E. Public BLASTP Results for NOV16a				
Protein Accession Number	Protein/Organism/Length	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BRG0	HYPOTHETICAL 58.1 KDA PROTEIN - Homo sapiens (Human), 542 aa (fragment).	1..531 5..542	514/539 (95%) 516/539 (95%)	0.0
Q9VQX4	CG3714 PROTEIN - Drosophila melanogaster (Fruit fly), 541 aa.	14..531 13..536	234/525 (44%) 330/525 (62%)	e-120
O80459	AT2G23420 PROTEIN - Arabidopsis thaliana (Mouse-ear cress), 574 aa.	14..531 25..569	231/547 (42%) 330/547 (60%)	e-112
AAK68525	HYPOTHETICAL 57.8 KDA PROTEIN - Caenorhabditis elegans, 511 aa.	13..445 9..449	198/443 (44%) 290/443 (64%)	e-101
Q95XX1	HYPOTHETICAL 59.9 KDA PROTEIN - Caenorhabditis elegans, 531 aa.	13..445 29..469	198/443 (44%) 290/443 (64%)	e-101

PFam analysis predicts that the NOV16a protein contains the domains shown in the Table 16F.

Table 16F. Domain Analysis of NOV16a			
Pfam Domain	NOV16a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

5

EXAMPLE 17.

The NOV17 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 17A.

Table 17A. NOV17 Sequence Analysis			
	SEQ ID NO: 71	572 bp	
NOV17a, CG59368-01 DNA Sequence	<u>CCGTGGTGCACGCGCTGCCCCGCATCAACCGCATGGTGCTGTGCTACCTCATCCGCTT</u> CCTGCAGGTCTTCGTGCAGCCGGCCAACGTGCGGGTCACCAAGATGGATGTCAGCAAC CTGGCCATGGTGATGGCGCCCAACTGCTTGCGCTGCCAGTCCGACGACCCGCGCGTCA TCTTCGAGAACACCCGCAAGGAGATGTCCTTCCTGCGGGTGCTCATCCAGCACCTGGA CACCAGCTTCATGGAGGGTGTGCTGT TAG CGGGGGCGCCCGGGGACAGGAGGGATGTCC TGCCGCCCCCAGCCAGGCCGAACCTCCGCACTCGCTCTCCCGGCAGAGGGGTGAGAATC <u>GCCCGGCCCAGCCCTGGAGCCCCCTCCACTCCCCCAGGCCCTGGCCCCGGCGCTCCC</u> CACGTCTTCTGCCTGGTCTGAGGGTGTAGCCAGGGGCACAGCAGCGGCGGGGAGGGCGC CTCTGGCCCCCACCTCACGGCCAGTTCCCGCGGGCACCGCCTCGCCCTCCGCTGGCC <u>GCGGGTCAGCTCCGAGAAAGTGCCTTCTGTAGCTTCATTTTATATTAATT</u>		
	ORF Start: ATG at 33	ORF Stop: TAG at 258	
	SEQ ID NO: 72	75 aa	MW at 8638.2kD
NOV17a, CG59368-01 Protein Sequence	MVLGYLIRFLQVFVQPANVAVTKMDVSNLAMVMAPNCLRCQSDDPRVIFENTRKEMSF LRVLIQHLDTSFMEGVL		

Further analysis of the NOV17a protein yielded the following properties shown in Table 17B.

Table 17B. Protein Sequence Properties NOV17a	
PSort analysis:	0.8134 probability located in mitochondrial intermembrane space; 0.5255 probability located in mitochondrial matrix space; 0.2672 probability located in lysosome (lumen); 0.2537 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 20 and 21

A search of the NOV17a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
5 homologous proteins shown in Table 17C.

Table 17C. Geneseq Results for NOV17a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE03048	Human preoptic regulatory factor-2 (hPORF-2) protein #1 - Homo sapiens, 75 aa. [WO200142464-A2, 14-JUN-2001]	1..75 1..75	75/75 (100%) 75/75 (100%)	1e-37

In a BLAST search of public sequence databases, the NOV17a protein was found to have homology to the proteins shown in the BLASTP data in Table 17D.

Table 17D. Public BLASTP Results for NOV17a

Protein Accession Number	Protein/Organism/Length	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9C0H5	KIAA1688 PROTEIN - Homo sapiens (Human), 1094 aa (fragment).	1..75 1020..1094	75/75 (100%) 75/75 (100%)	4e-37
P18890	Putative preoptic regulatory factor-2 precursor (PORF-2) - Rattus norvegicus (Rat), 75 aa.	1..75 1..75	74/75 (98%) 75/75 (99%)	5e-37
Q9VDE9	CG3421 PROTEIN - Drosophila melanogaster (Fruit fly), 1309 aa.	1..75 1235..1309	48/75 (64%) 58/75 (77%)	2e-21

PFam analysis predicts that the NOV17a protein contains the domains shown in the Table 17E.

Table 17E. Domain Analysis of NOV17a			
Pfam Domain	NOV17a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

EXAMPLE 18.

5 The NOV18 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 18A.

Table 18A. NOV18 Sequence Analysis		
	SEQ ID NO: 73	1452 bp
NOV18a, CG58628-01 DNA Sequence	ATCCTGCCCCGCAGGGTGACCCTGTTTGCAGCACGATGTCTGAAGAAGAGGCGGCTCAGATCCCCAGATCCAGTGTGTGGGAGCAGGACCAGCAGAACGTGGTGCAGCGTGTGGTGCTCTGCCCCCTGGTCAGGGCCACGTGCACCGCGGTCTGCGATGTTTACAGTGCAGCCAAGGACAGGCACCCGCTGCTGGGCTCCGCCTGCCGCCTGGCTGAGAACTGCGTGTGCGGCCTGACCACCCGTGCCCTGGACCACGCCAGCCGCTGCTCGAGCACCTGCAGCCCCAGGTGGCCACTATGAACAGCCTCGCCTGCAGGGGCCTGGACAAGCTGGAAGAGAAGCTTCCTTTCTCCAGCAACCTTCGGAGACGGTAGTGACCTCAGCCAAGGACGTGGTGGCCAGCAGTGTACGGGTGTGGTGGACCTGGCCCCGGAGGGGCCGGCGCTGGAGCGTGGAGCTGAAGCGCTCCGTGAGCCATGCTGTGGATGTTGTACTGGAAAAATCAGAGGAGCTGGTGGATCACTTCCTGCCCATGACGGAGGAAGAGCTCGCGGCACTGGCGGCTGAGGCTGAAGGCCCTGAAGTGGGTTCCGTGGAGGATCAGAGGAGACAGCAGGGCTACTTTGTGCGCCTCGGCTCCCTGTCAGCACGGATCCGCCACCTGGCCTACGAGCACTCTGTGGGGAACTGAGGCAGAGCAAACACCGTGCCAGGACACCCTGGCCCAGCTGCAGGAGACGCTGGAGCTGATAGACCACATGCAGTGTGGGGTGACCCCCACCGCCCCGGCCTGCCCTGGGAAGGTGCACGAGCTGTGGGGGGAATGGGGCCAGCGCCCTCCGGAGAGCCGCCGCCGGAGCCAGGTGGAGCTGGAGACGCTGGTGTGTCCCGCAGCCTGACCCAGGAGCTGCAGGGCACGGTAGAGGCTCTGGAGTCCAGCGTGCGGGGCCTGCCCGCCGGCGCCAGGAGAAGGTGGCTGAGGTGCGGCGCAGTGTGGATGCCCTGCAGACCGCCTTCGCTGATGCCCGCTGCTTCAGGGACGTGCCAGCGGCCGCTGGCCGAGGGCCGGGGTCGCGTGGCCCACGCGCACGCTGCGGTGGACGAGCTGCTGGAGCTGGTGGTGCAGGCCGTGCCGCTGCCCTGGCTGGTGGACCCTTCGCGCCCATCCTTGTGGAGCGACCCGAGCCCCTGCCGACCTGGCGGACC	

	TGGTGGACGAGGTCATCGGGGGCCCTGACCCCCGCTGGGCGCACCTGGACTGGCCGGC CCAGCAGAGAGCCTGGGAGGCAGAGCACAGGGACGGGAGTGGGAATGGGGATGGGGAC AGGATGGGTGTTGCCGGGGACATCTGCGAGCAGGAACCCGAGACCCCAGCTGCCCGG TCAAGCACACCCTGATGCCCCGAGCTGGACTTCTG <u>ACCCATGGGCCAGTGGAGGCGGG</u> <u>AG</u>		
	ORF Start: ATG at 36	ORF Stop: TGA at 1425	
	SEQ ID NO: 74	463 aa	MW at 50804.9kD
NOV18a, CG58628-01 Protein Sequence	MSEEEAAQIPRSSVWEQDQQNVVQRVVALPLVRATCTAVCDVYSAAKDRHPLLGSACR LAENCVCGLTTRALDHAQPLLEHLQPQVATMNSLACRGLDKLEEKLPFLQQPSETVVT SAKDVVASSVTGVVDLARRGRRWSVELKRSVSHAVDVVLEKSEELVDHFLPMTEEELA ALAAEAEGPEVGSVEDQRRQQGYFVRLGSLSARIRHLAYEHSVGKLRQSKHRAQDTLA QLQETLELIDHMQCGVTPTAPACPGKVHELWGEWGQRPPESSRRRSQVELETLVLSRSL TQELQGTVEALESSVRGLPAGAQEKVAEVRRSVDALQTAFADARCFRDVPAALAEGR GRVAHAHACVDELLELVVQAVPLPWLVGPFAPILVERPEPLPDLADLVDEVIGGPDPR WAHLDWPAQQRAWAEHRDGSNGDGD RMGVAGDICEQEPETPSCPVKHTLMPELDF		
	SEQ ID NO: 75	978 bp	
NOV18b, 174228350 DNA Sequence	AGATCTGACCAGCAGAACGTGGTGCAGCGTGTGGTGGCTCTGCCCCTGGTCAGGGCCA CGTGACCGCGGTCTGCGATGTTTACAGTGCAGCCAAGGACAGGCACCCGCTGCTGGG CTCCGCCTGCCGCTGGCTGAGAACTGCGTGTGCGGCCTGACCACCCGTGCCCTGGAC CACGCCAGCCGCTGCTCGAGCACCTGCAGCCCCAGCTGGCCACTATGAACAGCCTCG CCTGCAGGGGCCTGGACAAGCTGGAAGAGAAGCTTCCCTTTCTCCAGCAACCTTCGGA GACGGTGGTGACCTCAGCCAAGGACGTGGTGGCCAGCAGTGTACGGGTGTGGTGGAC CTGGCCCGGAGGGGCGGCGCTGGAGCGTGGAGCTGAAGCGCTCCGTGAGCCATGCTG TGGATGTTGTACTGGAAAAATCAGAGGAGCTGGTGGATCACTTCCTGCCCATGACGGA GGAAGAGCTCGCGGCACTGGCGGCTGAGGCTGAAGGCCCTGAAGTGGGTTTCGGTGGAG GATCAGAGGAGACAGCAGGGCTACTTTGTGCGCCTCGGCTCCCTGTCAGCACGGATCC GCCACCTGGCCTACGAGCACTCTGTGGGGAACTGAGGCAGAGCAAACACCGTGCCCA GGACACCCTGGCCCAGCTGCAGGAGACGCTGGAGCTGATAGACCACATGCAGTGTGGG GTGACCCCCACCGCCCCGGCCCCGCCCTGGGAAGGTGCACGAGCTGTGGGGGGAATGGG GCCAGCGCCCTCCGGAGAGCCGCCGCGGAGCCAGGCAGAGCTGGAGACGCTGGTGCT GTCCCGCAGCCTGACCCAGGAGCTGCAGGGCACGGTAGAGGCTCTGGAGTCCAGCGTG CGGGGCCTGCCCGCGGCGCCAGGAGAAGGTGGCTGAGGTGCGGCGCAGTGTGGATG CCCTGCAGACCGCCTTCGCTGATGCCCCGCTGCTTCAGGGACGTGGTTCGAC		
	ORF Start: AGA at 1	ORF Stop:	
	SEQ ID NO: 76	326 aa	MW at 35954.4kD
NOV18b, 174228350 Protein Sequence	RSDQQNVVQRVVALPLVRATCTAVCDVYSAAKDRHPLLGSACRLAENCVCGLTTRALD HAQPLLEHLQPQLATMNSLACRGLDKLEEKLPFLQQPSETVVTSAKDVVASSVTGVVD LARRGRRWSVELKRSVSHAVDVVLEKSEELVDHFLPMTEEELAALAAEAEGPEVGSVE DQRRQQGYFVRLGSLSARIRHLAYEHSVGKLRQSKHRAQDTLAQLQETLELIDHMQCG VTPTAPARPGKVHELWGEWGQRPPESSRRRSQAELETLVLSRSLTQELQGTVEALESSV RGLPAGAQEKVAEVRRSVDALQTAFADARCFRDVVD		
	SEQ ID NO: 77	978 bp	
NOV18c, 174228354 DNA Sequence	AGATCTGACCAGCAGAACGTGGTGCAGCGTGTGGTGGCTCTGCCCCTGGTCAGGGCCA CGTGACCGCGGTCTGCGATGTTTACAGTGCAGCCAAGGACAGGCACCCGCTGCTGGG CTCCGCCTGCCGCTGGCTGAGAACTGCGTGTGCGGCCTGACCACCCGTGCCCTGGAC CACGCCAGCCGCTGCTCGAGCACCTGCAGCCCCAGCTGGCCACTATGAACAGCCTCG CCTGCAGGGGCCTGGACAAGCTGGAAGAGAAGCTTCCCTTTCTCCAGCAACCTTCGGA GACGGTGGTGACCTCAGCCAAGGACGTGGTGGCCAGCAGTGTACGGGTGTGGTGGAC CTGGCCCGGAGGGGCGGCGCTGGAGCGTGGAGCTGAAGCGCTCCGTGAGCCATGCTG TGGATGTTGTACTGGAAAAATCAGAGGAGCTGGTGGATCACTTCCTGCCCATGACGGA GGAAGAGCTCGCGGCACTGGCGGCTGAGGCTGAAGGCCCTGAAGTGGGTTTCGGTGGAG GATCAGAGGAGACAGCAGGGCTACTTTGTGCGCCTCGGCTCCCTGTCAGCACGGATCC GCCACCTGGCCTACGAGCACTCTGTGGGGAACTGAGGCAGAGCAAACACCGTGCCCA GGACACCCTGGCCCAGCTGCAGGAGACGCTGGAGCTGATAGACCACATGCAGTGTGGG GTGACCCCCACCGCCCCGGCCCCGCCCTGGGAAGGTGCACGAGCTGTGGGGGGAATGGG		

	GCCAGCGCCCTCCGGAGAGCCGCCGCCGGAGCCAGGCAGAGCTGGAGACGCTGGTGCT GTCCCGCAGCCTGACCCAGGAGCTGCAGGGCACGGTAGAGGCTCTGGAGTCCAGCGTG TGGGGCCTGCCCGCCGGCGCCAGGAGAAGGTGGCTGAGGTGCGGCGCAGTGTGGATG CCCTGCAGACCGCCTTCGCTGATGCCCGCTGCTTCAGGGACGTGGTTCGAC		
	ORF Start: AGA at 1	ORF Stop:	
	SEQ ID NO: 78	326 aa	MW at 35984.4kD
NOV18c, 174228354 Protein Sequence	RSDQQNVVQRVVALPLVRATCTAVCDVYSAKDRHPLLGSACRLAENCVCGLTTRALD HAQPLLEHLQPQLATMNSLACRGLDKLEEKLPFLQQPSETVVVTSKDVVASSVTGVVD LARRGRRWSVELKRSVSHAVDVVLEKSEELVDHFLPMTEEEAALAAEAEGPEVGSVE DQRRQQGYFVRLGSLSARIRHLAYEHSVGLRQSKHRAQDTLAQLQETLELIDHMQCG VTPTAPARPGKVHELWGEWGQRPPESSRRRSQAELETLVLSRSLTQELQGTVEALESSV WGLPAGAQEKVAEVRRSVDALQTAFADARCFRDVVD		
	SEQ ID NO: 79	1401 bp	
NOV18d, 188822733 DNA Sequence	AGATCTATGTCTGAAGAAGAGGCGGCTCAGATCCCCAGATCCAGTGTGTGGGAGCAGG ACCAGCAGAACGTGGTGCAGCGTGTGGTGGCTCTGCCCCTGGTCAGGGCCACGTGCAC CGCGGTCTGCGATGTTTACAGTGCAGCCAAGGACAGGCACCCGCTGCTGGGCTCCGCC TGCCGCTGGCTGAGAACTGCGTGTGCGGCCTGACCACCCGTGCCCTGGACCACGCCC AGCCGCTGCTCGAGCACCTGCAGCCCCAGCTGGCCACTATGAACAGCCTCGCCTGCAG GGGCTGGACAAGCTGGAAGAGAAGCTTCCCTTTCTCCAGCAACCTTCGGAGACGGTG GTGACCTCAGCCAAGGACGTGGTGGCCAGCAGTGTACGGGTGTGGTGGACCTGGCCC GGAGGGGCCGCGCTGGAGCGTGGAGCTGAAGCGCTCCGTGAGCCATGCTGTGGATGT TGTA CTGGA AAAATCAGAGGAGCTGGTGGATCACTTCTGCCCATGACGGAGGAAGAG CTCGCGGCACTGGCGGCTGAGGCTGAAGGCCCTGAAGTGGGTTCGGTGGAGGATCAGA GGAGACAGCAGGGCTACTTTGTGCGCCTCGGCTCCCTGTCAGCACGGATCCGCCACCT GGCCTACGAGCACTCTGTGGGGAACTGAGGCAGAGCAAACACCGTGCCAGGACACC CTGGCCCAGCTGCAGGAGACGCTGGAGCTGATAGACCACATGCAGTGTGGGGTGACCC CCACCGCCCCGGCCCGCCCTGGGAAGGTGCACGAGCTGTGGGGGAATGGGGCCAGCG CCCTCCGGAGAGCCGCCCGCGGAGCCAGGCAGAGCTGGAGACGCTGGTGTGCTCCCGC AGCCTGACCCAGGAGCTGCAGGGCACGGTAGAGGCTCTGGAGTCCAGCGTGTGGGGCC TGCCCGCCGGCGCCAGGAGAAGGTGGCTGAGGTGCGGCGCAGTGTGGATGCCCTGCA GACCGCCTTCGCTGATGCCCGCTGCTTCAGGGACGTGCCAGCGGCCGCGCTGGCCGAG GGCCGGGGTTCGCTGGCCCACGCGCACGCTGCGTGGACGAGCTGCTGGAGCTGGTGG TGCAGGCCGTGCCGCTGCCCTGGCTGGTGGGACCCCTTCGCGCCCATCCTTGTGGAGCG ACCCGAGCCCCTGCCCGACCTGGCGGACCTGGTGGACGAGGTCATCGGGGGCCCTGAC CCCCGCTGGGCGCACCTGGACTGGCCGGCCAGCAGAGAGCCTGGGAGGCAGAGCACA GGGACGGGAGTGGGAATGGGGATGGGGACAGGATGGGTGTTGCCGGGGACATCTGCGA GCAGGAACCCGAGACCCCAGCTGCCCGGTCAAGCACACCCTGATGCCCGAGCTGGAC TTCGTTCGAC		
	ORF Start: AGA at 1	ORF Stop:	
	SEQ ID NO: 80	467 aa	MW at 51331.4kD
NOV18d, 188822733 Protein Sequence	RSMSEEEAAQIPRSSVWEQDQQNVVQRVVALPLVRATCTAVCDVYSAKDRHPLLGSA CRLAENCVCGLTTRALDHAQPLLEHLQPQLATMNSLACRGLDKLEEKLPFLQQPSETV VTSKDVVASSVTGVVDLARRGRRWSVELKRSVSHAVDVVLEKSEELVDHFLPMTEEE LAALAAEAEGPEVGSVEDQRRQQGYFVRLGSLSARIRHLAYEHSVGLRQSKHRAQDT LAQLQETLELIDHMQCGVTPTAPARPGKVHELWGEWGQRPPESSRRRSQAELETLVLSR SLTQELQGTVEALESSVWGLPAGAQEKVAEVRRSVDALQTAFADARCFRDVPAAALAE GRGRVAHAHACVDELLELVQAVPLPWLVGPFAPILVERPEPLPDLADLVDEVIGGPD PRWAHLDWPAQQRAWAEAEHRDGSNGDGRMGVAGDICEQEPETPSCPVKHTLMPELD FVD		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 18B.

Table 18B. Comparison of NOV18a against NOV18b through NOV18d.

Protein Sequence	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV18b	18..339 3..324	307/322 (95%) 308/322 (95%)
NOV18c	18..339 3..324	306/322 (95%) 307/322 (95%)
NOV18d	1..463 3..465	447/463 (96%) 448/463 (96%)

Further analysis of the NOV18a protein yielded the following properties shown in Table 18C.

Table 18C. Protein Sequence Properties NOV18a	
PSort analysis:	0.3000 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV18a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
5 homologous proteins shown in Table 18D.

Table 18D. Geneseq Results for NOV18a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY67240	Human adipophilin-like protein (HALP) amino acid sequence - Homo sapiens, 434 aa. [US5989820-A, 23-NOV-1999]	19..385 22..428	163/407 (40%) 234/407 (57%)	1e-77
AAW59883	Amino acid sequence of the cDNA clone ADF (HFKFY79) - Homo sapiens, 452 aa. [WO9831800-A2, 23-JUL-1998]	19..388 22..431	149/411 (36%) 223/411 (54%)	4e-64
AAM25962	Human protein sequence SEQ ID NO:1477 - Homo sapiens, 139 aa. [WO200153455-A2, 26-JUL-2001]	1..117 23..139	116/117 (99%) 117/117 (99%)	4e-62
AAY99534	Human adipocyte-specific differentiation-related protein ADRP - Homo sapiens, 437 aa. [WO200031532-A1, 02-JUN-2000]	12..384 2..411	140/416 (33%) 222/416 (52%)	1e-59

AAW53264	Human adipocyte-specific differentiation-related protein - Homo sapiens, 437 aa. [US5739009-A, 14-APR-1998]	12..384 2..411	140/416 (33%) 222/416 (52%)	1e-59
----------	---	-------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV18a protein was found to have homology to the proteins shown in the BLASTP data in Table 18E.

Table 18E. Public BLASTP Results for NOV18a				
Protein Accession Number	Protein/Organism/Length	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9D6M0	2310076L09RIK PROTEIN - Mus musculus (Mouse), 448 aa.	1..463 1..448	329/463 (71%) 368/463 (79%)	0.0
Q9BS03	CARGO SELECTION PROTEIN (MANNOSE 6 PHOSPHATE RECEPTOR BINDING PROTEIN) - Homo sapiens (Human), 434 aa.	19..385 22..428	163/407 (40%) 234/407 (57%)	4e-77
O60664	Cargo selection protein TIP47 (47 kDa mannose 6-phosphate receptor- binding protein) (47 kDa MPR-binding protein) (Placental protein 17) - Homo sapiens (Human), 434 aa.	19..385 22..428	163/407 (40%) 234/407 (57%)	6e-77
Q9DBG5	1300012C15RIK PROTEIN (RIKEN CDNA 1300012C15 GENE) - Mus musculus (Mouse), 437 aa.	19..385 22..432	160/411 (38%) 232/411 (55%)	4e-73
Q9CZK1	1300012C15RIK PROTEIN - Mus musculus (Mouse), 437 aa.	19..385 22..432	160/411 (38%) 232/411 (55%)	6e-73

PFam analysis predicts that the NOV18a protein contains the domains shown in the Table 18F.

Table 18F. Domain Analysis of NOV18a			
Pfam Domain	NOV18a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Man-6-P_recep: domain 1 of 1	156..168	9/13 (69%) 9/13 (69%)	0.7
perilipin: domain 1 of 1	10..369	139/411 (34%) 240/411 (58%)	1.4e-76

EXAMPLE 19.

The NOV19 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 19A.

Table 19A. NOV19 Sequence Analysis			
	SEQ ID NO: 81	774 bp	
NOV19a, CG59342-01 DNA Sequence	GTAGAGTTTTTCAGGTTGCTCCTGGAAACCATGCCGAAAGTAGTGTCTCGGTCAGTAG TCTGCTCTGACACTCGGGACCGGGAGGAATATGACGACGGCGAGAAGCCCCCTCCATGT GTACTACTGTTTGTGCGGCCAGGTGGTCCTAGTGCTGGACTGTCAGTTAGAGAAATTG CCCATGAGGCCCCGGGACCGGTCCCGTGTGATTGATGCTGCCAAACATGCCCATAAGT TTTGTAACACAGAAGACGAAGAGACTATGTATCTGCGGAGACCTGAAGGCATTGAACT ACAGTACAGAAAGAAATGTGCAAAGTGTGGACTGCTGCTCTTCTACCAATCCCAGCCG AAGAATGCTCCCGTTACCTTCATTGTGGATGGAGCAGTCGTCAAGTTTGGCCAGGGCT TTGGGAAAACGAACATATATACTCAGAAACAAGAGCCTCCTAAGAAGGTGATGATGAC CAAACGGACCAAAGACATGGGCAAGTTCAGTTCTGTCACTGTGTCTACCATTGATGAA GAGGAAGAGGAGATTGAGGCTAGGGAAGTTGCTGACTCGTATGCACAGAATGCCAAAG TGATTGAAAAACAGCTGGAGCGCAAAGGCATGAGCAAGAGGCCACTGCAAGAGCTGGC TGAATGGGAACCCCAGGAAAAGAGGACATATGACACAGGTTCTCCCTCTGCAAAAAG TGGCAGATGCGTGGCTCAGGGGCCTTCCACTGTCCAGGTCCTCCTCAGATGGCCCTGG GAATGAGCGGCCACCATTAA		
	ORF Start: ATG at 31	ORF Stop: TAA at 772	
	SEQ ID NO: 82	247 aa	MW at 28211.1kD
NOV19a, CG59342-01 Protein Sequence	MPKVVSRSVVCSDTRDREYDDGEKPLHVYYCLGQVVLVLDQCLEKLPMRPRDRSRV IDAAKHAHKFCNTEDEETMYLRRPEGIELQYRKKCAKCGLLLFYQSQPKNAPVTFIVD GAVVKFGQGFGKTNITYTQKQEPKKVMMTKRTKDMGKFSSVTVSTIDEEEEIEAREV ADSYAQNAKVIEKQLERKGMSKRPLQELAWEWPEKRTYDTGSPSAKKWQMRGSGAFH CPGPPQMALGMSGHH		

Further analysis of the NOV19a protein yielded the following properties shown in

5 Table 19B.

Table 19B. Protein Sequence Properties NOV19a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3600 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV19a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 19C.

Table 19C. Geneseq Results for NOV19a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

AAG01784	Human secreted protein, SEQ ID NO: 5865 - Homo sapiens, 87 aa. [EP1033401-A2, 06-SEP-2000]	1..86 1..86	85/86 (98%) 86/86 (99%)	3e-46
AAM41425	Human polypeptide SEQ ID NO 6356 - Homo sapiens, 92 aa. [WO200153312-A1, 26-JUL-2001]	139..215 9..85	65/77 (84%) 69/77 (89%)	4e-28
AAM39639	Human polypeptide SEQ ID NO 2784 - Homo sapiens, 80 aa. [WO200153312-A1, 26-JUL-2001]	143..215 1..73	63/73 (86%) 66/73 (90%)	6e-27
AAG60283	Arabidopsis thaliana protein fragment SEQ ID NO: 78065 - Arabidopsis thaliana, 236 aa. [EP1033405-A2, 06-SEP-2000]	1..119 1..122	44/127 (34%) 64/127 (49%)	2e-09
AAG59843	Arabidopsis thaliana protein fragment SEQ ID NO: 77448 - Arabidopsis thaliana, 230 aa. [EP1033405-A2, 06-SEP-2000]	1..119 1..116	42/123 (34%) 63/123 (51%)	8e-09

In a BLAST search of public sequence databases, the NOV19a protein was found to have homology to the proteins shown in the BLASTP data in Table 19D.

Table 19D. Public BLASTP Results for NOV19a				
Protein Accession Number	Protein/Organism/Length	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H5V9	CDNA: FLJ22965 FIS, CLONE KAT10418 - Homo sapiens (Human), 222 aa.	1..215 1..215	202/215 (93%) 206/215 (94%)	e-114
AAH21479	HYPOTHETICAL 25.6 KDA PROTEIN - Mus musculus (Mouse), 222 aa.	1..215 1..215	201/215 (93%) 205/215 (94%)	e-113
Q9CWC1	C330007P06RIK PROTEIN - Mus musculus (Mouse), 250 aa.	1..202 1..202	197/202 (97%) 198/202 (97%)	e-111
Q9V412	BG:DS00941.3 PROTEIN - Drosophila melanogaster (Fruit fly), 247 aa.	1..193 1..218	106/220 (48%) 145/220 (65%)	2e-50
Q95Q06	Y66D12A.8 PROTEIN - Caenorhabditis elegans, 244 aa.	13..194 29..207	79/187 (42%) 114/187 (60%)	1e-30

PFam analysis predicts that the NOV19a protein contains the domains shown in the Table 19E.

Table 19E. Domain Analysis of NOV19a			
Pfam Domain	NOV19a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

EXAMPLE 20.

The NOV20 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 20A.

Table 20A. NOV20 Sequence Analysis			
	SEQ ID NO: 83	324 bp	
NOV20a, CG59486-01 DNA Sequence	ATTTTTTTGTTGTTATTGTTGTAGATATGTGGTTTCCCCATGTTGCCAGCTGGCCTCG AACTCCTGGCCTCAAGATCCACCCGCTCGACCTCCCAAAGGCCAGCCCTCTCTTT CCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTGTTTTTAAAAAAAAAAAA AAAGGCCAGGCGCAGTGGCTCATGTCTGTAATCCCAGCACTCTGGGAGGCCAAGGCAG GCAGATCACAAGGTCAGGAGATCAAGACCATCCTGGCTAACACAGTGAAACCCCATCT CTACTAAAAAATACAAAAAAAATTAGCCAGGCG		
	ORF Start: ATG at 40	ORF Stop: TAA at 295	
	SEQ ID NO: 84	85 aa	MW at 9476.2kD
NOV20a, CG59486-01 Protein Sequence	MLPAGLELLASRSTRDLPLKAQPLSFLPSFLPSFLVFKKKKGQAQWLMSVIPAL LWEAKAGRSQGQEIKTILANTVKPHLY		

Further analysis of the NOV20a protein yielded the following properties shown in
5 Table 20B.

Table 20B. Protein Sequence Properties NOV20a	
PSort analysis:	0.6238 probability located in microbody (peroxisome); 0.6000 probability located in nucleus; 0.3600 probability located in mitochondrial matrix space; 0.1830 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV20a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 20C.

Table 20C. Geneseq Results for NOV20a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

AAB95050	Human protein sequence SEQ ID NO:16847 - Homo sapiens, 112 aa. [EP1074617-A2, 07-FEB-2001]	35..85 62..112	38/51 (74%) 42/51 (81%)	4e-15
ABB11422	Human Zn finger protein homologue, SEQ ID NO:1792 - Homo sapiens, 670 aa. [WO200157188-A2, 09-AUG-2001]	47..85 632..670	32/39 (82%) 35/39 (89%)	1e-12
AAM85296	Human immune/haematopoietic antigen SEQ ID NO:12889 - Homo sapiens, 81 aa. [WO200157182-A2, 09-AUG-2001]	37..84 19..67	35/49 (71%) 41/49 (83%)	1e-12
AAM94124	Human reproductive system related antigen SEQ ID NO: 2782 - Homo sapiens, 107 aa. [WO200155320-A2, 02-AUG-2001]	41..85 63..107	33/45 (73%) 38/45 (84%)	3e-12
AAM91494	Human immune/haematopoietic antigen SEQ ID NO:19087 - Homo sapiens, 58 aa. [WO200157182-A2, 09-AUG-2001]	49..85 22..58	32/37 (86%) 34/37 (91%)	3e-12

In a BLAST search of public sequence databases, the NOV20a protein was found to have homology to the proteins shown in the BLASTP data in Table 20D.

Table 20D. Public BLASTP Results for NOV20a				
Protein Accession Number	Protein/Organism/Length	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UI59	PRO0478 - Homo sapiens (Human), 87 aa.	8..85 7..87	50/82 (60%) 59/82 (70%)	4e-18
P39189	Alu subfamily SB sequence contamination warning entry - Homo sapiens (Human), 587 aa.	47..85 1..39	30/39 (76%) 35/39 (88%)	1e-10
P39192	Alu subfamily SC sequence contamination warning entry - Homo sapiens (Human), 585 aa.	47..85 1..39	29/39 (74%) 34/39 (86%)	5e-10
P39191	Alu subfamily SB2 sequence contamination warning entry - Homo sapiens (Human), 603 aa.	47..85 1..39	29/39 (74%) 33/39 (84%)	9e-10
P39190	Alu subfamily SB1 sequence contamination warning entry - Homo sapiens (Human), 587 aa.	47..85 1..39	28/39 (71%) 33/39 (83%)	3e-09

PFam analysis predicts that the NOV20a protein contains the domains shown in the Table 20E.

Table 20E. Domain Analysis of NOV20a			
Pfam Domain	NOV20a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

EXAMPLE 21.

The NOV21 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 21A.

Table 21A. NOV21 Sequence Analysis			
	SEQ ID NO: 85		1572 bp
NOV21a, CG59446-01 DNA Sequence	GGTGTGCAGGATATAAGGTTGGACTTCCAGACCCACTGCCCGGGAGAGGAGAGGAGCC GGCCGAGGACTCCAGCGTGCCCGAGGTCTGGCATCCTGCACTTGCTGCCCTCTGACACCC TGGGAAGATGGCCGGCCCGTGGACCTTCACCCTTCTCTGTGGTTTGCTGGCAGCCACC TTGATCCAAGCCACCCTCAGTCCCAGTGCAGTTCTCATCCTCGGCCCAAAGTCATCA AAGAAAAGCTGACACAGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCT GCCGCTGCTCAGTGCCATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGC CTGGTGAACACCGTCTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCC TCCAGCTGCAGGTGAAGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCT GGACATGGTGGCTGGATTCAACACGCCCTGGTCAAGACCATCGTGGAGTTCCACATC ACGACTGAGGCCCAAGCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCC TGGTCCTCAGTGACTGTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCATAA GCTCTCCTTCTGGTGAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCC CTGCCCAATCTAGTGAAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCA TGTATGCAGACCTCCTGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCT GGAGTTTGACCTTCTGTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGG GCCAAGTTGTTGGACTCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTT CCCTGACAATGCCACCCCTGGACAACATCCCGTTCAGCCTCATCGTGAGTCACCCGTT CAGCCTCATCGTGAGTCAGGACGTGGTGAAGGCTGCAGTGGCTGCTGTGCTCTCTCCA GAAGAATTCATGGTCCTGTTGGACTCTGTGCTTCCTGAGAGTGCCCATCGGCTGAAGT CAAGCATCGGGCTGATCAATGAAAAGGAAGCCAGCTCGGAAGCTCAGTTTTACACCAA AGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCTGATC AACTCTGGGATTGGCTGGTTCACCTGATGTTCTGAAAAACATCATCACTGAGATCA TCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAGTGT ATTGGTGAAGGCCTTGGGATTTCGAGGCAGCTGAGTCCTCACTGACCAAGGATGCCCTT GTGCTTACTCCAGCCTCCTTGTGGAAACCAGCTCTCCTGTCTCCAGTGAAGACTTG GATGGCAGCCATCAGGGAAGGCTGGGTCCCAGTTGGGAGTATGGGTGTGAGCTCTATA GACCATCCCTCTCTGCAATCAATAAACACTTGCCTGTGAAAAAAAAAAAAAAAAATAAAA AAAAAA		
	ORF Start: ATG at 124		ORF Stop: TGA at 1441
	SEQ ID NO: 86		439 aa MW at 47572.2kD
NOV21a, CG59446-01 Protein Sequence	MAGPWTFTLLCGLLAATLIQATLSPTAVLILGPKVIKEKLTQELKDHNATSILQQLPL LSAMREKPAGGIPVLGSLVNTVLKHIIWLKVITANILQLQVKPSANDQELLVKIPLDM VAGFNTPLVKTI VEFHMTTEAQATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLS FLVNALAKQVMNLLVPSLPNLVKNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEF DLLYPAIKGDTIQLYLGAKLLDSQGKVTKWFNNSAASLTMPITLDNIPFSLIVSHPFSL IVSQDVVKA AVAVLSPEEFMVLLDSVLPESAHLKSSIGLINEKEASSEAQFYTKGD QLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITEIIHSILLPNQNGKLRSQVPSLV		

	KALGFEEAESSLT KDALV LTPASLWKPSSPVSQ		
	SEQ ID NO: 87	1392 bp	
NOV21b, 174308261 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAACACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAC T GCTGCATAAGCTCTCCTTCCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCGTCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATT CAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTT CAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCCTGTTGGACTCTGTGCTTCCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGACCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTT CACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCAG TGTCAATTGGTGAAGGCCTTGGGATTGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop: LV at 1393	
	SEQ ID NO: 88	464 aa	MW at 50459.5kD
NOV21b, 174308261 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSI LQQLPLLSAMREKPAGGI PVLGSLVNTV LKHIIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPI SLSIDRLEFDLLYP AIKGD TIQLYLGA KLLD SQGKVTKWFNNSAASLTMP TLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGP TQIVKILTQDTPEFFIDQGHAKVAQLIVLEVFP SSE ALRPLFTLGI EASSEAQFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNI ITE IIHSILLPNQNGKL RSGVPVSLVKALGFEEAESSLT KDALV LTPASLWKPSSPVSQLE		
	SEQ ID NO: 89	1392 bp	
NOV21c, 174308266 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAGTCATCAAAGAAAAGCCGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAC T GCTGCATAAGCTCTCCTTCCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATT CAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTT CAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCCTGTTGGACTCTGTGCTTCCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTT CACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT		

	GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTTCGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 90	464 aa	MW at 50433.4kD
NOV21c, 174308266 Protein Sequence	KLPTAVLILGPKVIEKEKPTQELKDNATSILQQLPLLSAMREKPAGGI PVLGSLVNTV LKHI IWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPI SLSIDRLEFDLLYP AIKGDTIQLYLGAKLLD SQGKVTKWFNNSAASLTMP TLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKL GSTQIVKILTQDTPEFFIDQGHAKVAQLIVLEVFP SSE ALRPLFTLGIEASSEAQFYTKGDQLILNLN NISSDRIQLMNSGIGWFQPDVLKNI ITE IIHSILLPNQNGKL RSGVPVSLVKALGFEEAESSLT KDALVLT PASLWKPSSPVSQLE		
	SEQ ID NO: 91	1392 bp	
NOV21d, 174308278 DNA Sequence	AAGCTTCCCAGTGCAGTTCTCATCCTCGGCCCAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCCTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCACATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCTCAGTGA GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCTGTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTACCCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTTCGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 92	464 aa	MW at 50430.4kD
NOV21d, 174308278 Protein Sequence	KLPTAVLILGPKVIEKEKLTQELKDNATSILQQLPLLSAMREKPAGGI PVLGSLVNTV LKHI IWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIHMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPI SLSIDRLEFDLLYP AIKGDTIQLYLGAKLLD SQGKVTKWFNNSAASLTMP TLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKL GSTQIVKILTQDTPEFFIDQGHAKVAQLIVLEVFP SSE ALRPLFTLGIEASSEAQFYTKGDQLILNLN NISSDRIQLMNSGIGWFQPDVLKNI ITE IIHSILLPNQNGKL RSGVPVSLVKALGFEEAESSLT KDALVLT PASLWKPSSPVSQLE		
	SEQ ID NO: 93	1392 bp	
NOV21e, 174308283 DNA Sequence	AAGCTTCCCAGTGCAGTTCTCATCCTCGGCCCAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTTAAGATCCCCCTGGACATGGTGGCTGG		

	ATTCAACACGCCCCCTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAACTGCTGCATAAGCTCTCCTTCCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTACAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCCTGTTGGACTCTGTGCTTCCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTTACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCCCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCAG CGTCATTGGTGAAGGCCTTGGGATTGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 94	464 aa	MW at 50431.4kD
NOV21e, 174308283 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILOQLPLLSAMREKPAGGI PVLGSLVNTV LKHIIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPI SLSIDRLEFDLLYP AIKGDITQLYLGAKLLD SQGKVTKWFNNSAASLTMP TLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPEFFIDQGHAKVAQLIVLEVFP SSE ALRPLFTLGLIEASSEAQFYTKGDLILNLNNISPDR IQLMNSGIGWFPDVLKNI ITE I IHSILLPNQNGKLRS GVPASLVKALGFEEAESSLT KDALVLT PASLWKPSSPVSQLE		
	SEQ ID NO: 95	1392 bp	
NOV21f, 174308287 DNA Sequence	AAGCTTCCCCTGACAGTTCTCATCCTCGGCCCAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCCTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAACTGCTGCATAAGCTCTCCTTCCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCGTTACAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCCTGTTGGACTCTGTGCTTCCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAGGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTTACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCAG TGTCAATTGGTGAAGGCCTTGGGATTGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 96	464 aa	MW at 50435.4kD

NOV21f, 174308287 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILQQLPLLSAMREKPAGGIPVLGSLVNTV LKHI IWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLYP AIKGDTVQLYLGAKLLD SQGKVTKWFNNSAASLTMP TLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPEFFIDQGHAKVAQLIVLEVFP SSE ALRPLFTLGIEASSEAQFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNI ITE IIHSILLPNQNGKLRSGVPVSLVKALGFEEAESSLT KDALVLT PASLWKPSSPVSQLE		
	SEQ ID NO: 97	1392 bp	
NOV21g, 174308293 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCCTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCATAAGCTCTCCTTCCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCTGTGGACTCTGTGCTTCCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTACCCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAAGTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAGATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTTGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 98	464 aa	MW at 50477.5kD
NOV21g, 174308293 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILQQLPLLSAMREKPAGGIPVLGSLVNTV LKHI IWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLYP AIKGDTIQLYLGAKLLD SQGKVTKWFNNSAASLTMP TLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPEFFIDQGHAKVAQLIVLEVFP SSE ALRPLFTLGIEASSEAQFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNI ITE IIHSILLPNQNGRLRSGVPVSLVKALGFEEAESSLT KDALVLT PASLWKPSSPVSQLE		
	SEQ ID NO: 99	1392 bp	
NOV21h, 174308301 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACGTATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCCTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCATAAGCTCTCCTTCCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA		

	CCCTGGACAACATCCCGTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCCTGTTGGACTCTGTGCTTCCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCAAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTACCCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACGCTGGGATTGGCTGGTTC AACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTTCGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 100	464 aa	MW at 50418.5kD
NOV21h, 174308301 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILQQLPLLSAMREKPAAGGIPVLGSLVNTV LKHVIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTIVEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLYP AIKGDTIQLYLGAKLLD SQGKVTKWFNNSAASLTMP TLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPKFFIDQGHAKVAQLIVLEVFPSSSE ALRPLFTLGIEASSEAQFYTKGDQLILNLNNISSDRIQLMNAGIGWFQPDVLKNIITE IIHSILLPNQNGKLRS GVPVSLVKALGFEEAESSLT KDALVLTPASLWKPSSPVSQLE		
	SEQ ID NO: 101	1392 bp	
NOV21i, 174308311 DNA Sequence	AAGCTTCCCAGTGCAGTTCTCATCCTCGGCCAAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCCGGTGAACACCGTC CTGAAGCACGTCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCCTGGTCAAGACCATCGTGGAGTTCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGGCCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCCTGTTGGACTCTGTGCTTCCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCAAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTACCCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACCTCTGGGATTGGCTGGTTC AACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTTCGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 102	464 aa	MW at 50360.4kD
NOV21i, 174308311 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILQQLPLLSAMREKPAAGGIPVLGSPVNTV LKHVIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTIVEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIGRLEFDLLYP AIKGDTIQLYLGAKLLD SQGKVTKWFNNSAASLTMP TLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPKFFIDQGHAKVAQLIVLEVFPSSSE ALRPLFTLGIEASSEAQFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQNGKLRS GVPVSLVKALGFEEAESSLT KDALVLTPASLWKPSSPVSQLE		

	SEQ ID NO: 103	1392 bp
NOV21j, 174308315 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCTGGTCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCATAAGCTCTCCTTCCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCCTGTTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAGGGTGGCCCAACTGATCGTGCTGGAAGTGTCTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTTACA CCAAAGGTGACCAACTTATACTCAACTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTCGAGGCAGATGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCAGCTCGAG	
	ORF Start: AAG at 1	ORF Stop:
	SEQ ID NO: 104	464 aa MW at 50461.4kD
NOV21j, 174308315 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILQQLPLLSAMREKPAGGIPVLGSLVNTV LKHIIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTIVEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLYPAIKGDITQLYLGAALLD SQGKVTKWFNNSAASLTMPITLDNIPFSLIVSQDVVKAAVAVALSPEEFMVL LDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPEFFIDQGHARVAQLIVLEVSPSSE ALRPLFTLGIEASSEAQFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQNGKLRSVGPVSLVKALGFEADESSLTKDALVLTPASLWKPPSPVSQL	
	SEQ ID NO: 105	1392 bp
NOV21k, 174308321 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCTGGTCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCATAAGCTCTCCTTCCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCCTGTTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTCCTCCAGTGTA GCCCTCCGCCCTTTGTTACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTTACA CCAAAGGTGACCAACTTATACTCAACTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG	

	ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCAATTGGTGAAGGCCTTGGGATTCGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 106	464 aa	MW at 50419.5kD
NOV21k, 174308321 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILQQLPLLSAMREKPAAGGIPVLGSLVNTV LKHI IWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLYP AIKGDTIQLYLGAKLLD SQGKVTKWFNNSAASLTMP TLDNIPFSLIVSQDVVKA AAVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTP EFFIDQGHAKVAQLIVLEVFPSSV ALRPLFTLGLIEASSEAQFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQNGKLRS GVPVSLVKALGFEEAESSLT KDALVLTPASLWKPPSPVSQL E		
	SEQ ID NO: 107	1392 bp	
NOV21l, 174308327 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCTGGTCAAGACCATCGTGGAGTTCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTATGAACCTCCTAGTGCCATCCCTGCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCCCC TGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATT CAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTT CAGCCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATT CATGGTCCTGTTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACGCTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTCAACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCAATTGGTGAAGGCCTTGGGATTCGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 108	464 aa	MW at 50403.4kD
NOV21l, 174308327 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILQQLPLLSAMREKPAAGGIPVLGSLVNTV LKHI IWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADPLQLVKVPISLSIDRLEFDLLYP AIKGDTIQLYLGAKLLD SQGKVTKWFNNSAASLTMP TLDNIPFSLIVSQDVVKA AAVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDAPEFFIDQGHAKVAQLIVLEVFPSS E ALRPLFTLGLIEASSEAQFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQNGKLRS GVPVSLVKALGFEEAESSLT KDALVLTPASLWKPPSPVSQL E		
	SEQ ID NO: 109	1392 bp	
NOV21m, 174308337 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACGTATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCTGGCCAAGACCATCGTGGAGTTCACATGACGACTGAGGCCCAA		

	GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTGTGCATAAGCTCTCCTTCCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GCATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCTTGACAATGCCCA CCCTGGACAACATCCCGTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCTGTTGGACTCTGTGCTTCCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCAAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTACCCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTACTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTCGAGGCAGCTGAGTCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGGGGAAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 110	464 aa	MW at 50251.2kD
NOV21m, 174308337 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILQQLPLLSAMREKPAGGIPVLGSLVNTV LKHVIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLAKTIVEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLHPAIGDITIQLYLGAALLD SQGKVTKWFNNSAASLTMPDLNIPFSLIVSQDVVKAAVAVALSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPKFFIDQGHAKVAQLIVLEVFPSSSE ALRPLFTLGIEASSEAFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQNGKLRSVPSLVKALGFEEAESSLTKDALVLT PASLGKPPSPVSQL E		
	SEQ ID NO: 111	1023 bp	
NOV21n, CG59446-02 DNA Sequence	CCTCTGACACCTGGGAAGATGGCCGGCCCGTGGACCTTCACCCTTCTCTGTGGTTTGC TGGCAGCCACCTTGATCCAAGCCACCCTCAGTCCCAGTTCATCCTCGGCCC AAAAGTCATCAAAGAAAAGCTGACACAGGAGCTGAAGGACCACAACGCCACCAGCATC CTGCAGCAGCTGCCGCTGCTCAGTGCCATGCGGGAAAAGCCAGCCGGAGGCATCCCTG TGCTGGGCAGCCTGGTGAACACCGTCTGAAGCACGTCTCTGGCTGAAGGTCATCAC AGCTAACATCCTCCAGCTGCAGGTGAAGCCCTCGGCCAATGACCAGGAGCTGCTAGTC AAGATCCCCCTGGACATGGTGGCTGGATTCAACACGCCCTGGTCAAGACCATCGTGG AGTTCCACATGACGACTGAGGCCCAAGCCACCATCCGCATGGACACCAGTGCAAGTGG CCCCACCCGCCTGGTCTCAGTGACTGTGCCACCAGCCATGGGAGCCTGCGCATCCAA CTGCTGCATAAGCTCTCCTTCCTGGTGAACGCCTTAGCTAAGCAGGTATGAACCTCC TAGTGCCATCCCTGCCCAATCTAGTGAAAAACCAGCTGTGTCCCGTGATCGAGGCTTC CTTCAATGGCATGTATGCAGACCTCCTGCAGCTGGTGAAGGTGCCATTTCCCTCAGC ATTGACCGTCTGGAGTTTGACCTTCTGTATCCTGCCATCAAGGGTGACACCATTGAGC TCTACCTGGGGGCCAAGTTGTTGGACTCACAGGGAAAGGTGACCAAGTGGTTCAATAA CTCTGCAGCTTCCTTGACAATGCCACCCTGGACAACATCCCGTTCAGCCTCATCGTG AGTCAGGACGTGGTGAAGCTGCAGTGGCTGCTGTGCTCTCTCCAGAAGAATTCATGG TCCTGTTGGACTCTGTGGTAAACCTCAGCACAAGGCAGAGAATAGGGCCGCCAGGCC ACATCATAGGAATTTCTGAACACAGGGTGCCCCCTAA		
	ORF Start: ATG at 19	ORF Stop: TAA at 1021	
	SEQ ID NO: 112	334 aa	MW at 36309.5kD
NOV21n, CG59446-02 Protein Sequence	MAGPWTFTLLCGLLAATLIQATLSPTAVLILGPKVIKEKLTQELKDHNATSILQQLPL LSAMREKPAGGIPVLGSLVNTVLKHVIWLKVITANILQLQVKPSANDQELLVKIPLDM VAGFNTPLVKTIVEFHMTTEAQATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLS FLVNALAKQVMNLLVPSLPNLVKNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEF DLLYPAIKGDTIQLYLGAALLDSQGKVTKWFNNSAASLTMPDLNIPFSLIVSQDVVK AAVAVALSPEEFMVLLDSVNLSTRQRIGPPRPHHRNFLNTGCP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 21B.

Table 21B. Comparison of NOV21a against NOV21b through NOV21n.		
Protein Sequence	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV21b	25..439 3..462	388/472 (82%) 393/472 (83%)
NOV21c	25..439 3..462	405/468 (86%) 405/468 (86%)
NOV21d	25..439 3..462	405/468 (86%) 405/468 (86%)
NOV21e	25..439 3..462	404/468 (86%) 404/468 (86%)
NOV21f	25..439 3..462	405/468 (86%) 406/468 (86%)
NOV21g	25..439 3..462	405/468 (86%) 406/468 (86%)
NOV21h	25..439 3..462	404/468 (86%) 406/468 (86%)
NOV21i	25..439 3..462	403/468 (86%) 404/468 (86%)
NOV21j	25..439 3..462	405/468 (86%) 405/468 (86%)
NOV21k	25..439 3..462	406/468 (86%) 406/468 (86%)
NOV21l	25..439 3..462	405/468 (86%) 405/468 (86%)
NOV21m	25..439 3..462	402/468 (85%) 404/468 (85%)
NOV21n	1..318 1..310	308/318 (96%) 310/318 (96%)

Further analysis of the NOV21a protein yielded the following properties shown in Table 21C.

Table 21C. Protein Sequence Properties NOV21a	
PSort analysis:	0.6138 probability located in outside; 0.4772 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)

SignalP analysis:	Likely cleavage site between residues 25 and 26
-------------------	---

A search of the NOV21a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 21D.

Table 21D. Geneseq Results for NOV21a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY77126	Human neurotransmission-associated protein (NTAP) 2799056 - Homo sapiens, 484 aa. [WO200001821-A2, 13-JAN-2000]	1..439 1..484	431/492 (87%) 431/492 (87%)	0.0
AAG63976	Amino acid sequence of a human Lng103 polypeptide - Homo sapiens, 484 aa. [WO200161055-A2, 23-AUG-2001]	1..439 1..484	430/492 (87%) 431/492 (87%)	0.0
AAU29163	Human PRO polypeptide sequence #140 - Homo sapiens, 484 aa. [WO200168848-A2, 20-SEP-2001]	1..439 1..484	430/492 (87%) 431/492 (87%)	0.0
AAB87564	Human PRO1357 - Homo sapiens, 484 aa. [WO200116318-A2, 08-MAR-2001]	1..439 1..484	430/492 (87%) 431/492 (87%)	0.0
AAB66124	Protein of the invention #36 - Unidentified, 484 aa. [WO200078961-A1, 28-DEC-2000]	1..439 1..484	430/492 (87%) 431/492 (87%)	0.0

In a BLAST search of public sequence databases, the NOV21a protein was found to have homology to the proteins shown in the BLASTP data in Table 21E.

Table 21E. Public BLASTP Results for NOV21a				
Protein Accession Number	Protein/Organism/Length	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96HK6	SIMILAR TO DNA SEGMENT, CHR 2, MASSACHUSETTS INSTITUTE OF TECHNOLOGY 19 - Homo sapiens (Human), 484 aa.	1..439 1..484	428/492 (86%) 431/492 (86%)	0.0
Q61114				e-127

	GLAND PROTEIN - Mus musculus (Mouse), 474 aa.	1..473	324/482 (66%)	
Q9BWZ6	DJ1187J4.1.1 (NOVEL PROTEIN SIMILAR TO MOUSE VON EBNER SALIVARY GLAND PROTEIN, ISOFORM 1.) - Homo sapiens (Human), 285 aa (fragment).	200..439 1..285	232/293 (79%) 232/293 (79%)	e-116
Q9BQP8	BA49G10.6 (SIMILAR TO MURINE VON EBNER MINOR SALIVARY GLAND PROTEIN, ISOFORM 1) - Homo sapiens (Human), 199 aa (fragment).	1..199 1..199	199/199 (100%) 199/199 (100%)	e-107
Q9H4V6	DJ1187J4.1.2 (NOVEL PROTEIN SIMILAR TO MOUSE VON EBNER SALIVARY GLAND PROTEIN, ISOFORM 2.) - Homo sapiens (Human), 213 aa.	272..439 1..213	160/221 (72%) 160/221 (72%)	1e-73

PFam analysis predicts that the NOV21a protein contains the domains shown in the Table 21F.

Table 21F. Domain Analysis of NOV21a			
Pfam Domain	NOV21a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

EXAMPLE 22.

5 The NOV22 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 22A.

Table 22A. NOV22 Sequence Analysis		
	SEQ ID NO: 113	2020 bp
NOV22a, CG59375-01 DNA Sequence	CATGAGTGAATGAAGGCACATGACAAACCTCCAGACCTGTGGAGACTGAAGGCTGAGA GCCTTTATAGATGCTGTGGGGCCGAGGAGTTTGCCAACTACAGCAGGTCATGCCCAGC GCTGCAAGAGGCCTACGTGCGGGTGGTCACCGAGAAGTCCCCGACCGACTGGGCTCTC TTTACCTATGAAGGCAACAGCAATGACATCCGCGTGGCTGGCACAGGGGAGGGTGGCC TGGAGGAGATGGTGGAGGAGCTCAACAGCGGGAAGGTGATGTACGCCTTCTGCAGAGT GAAGGACCCCAACTCTGGACTGCCCAAATTTGTCCTCATCAACTGGACAGGCGAGGGC GTGAACGATGTGCGGAAGGGAGCCTGTGCCAGCCACGTCAGCACCATGGCCAGCTTCC TGAAGGGGGCCCATGTGACCATCAACGCACGGGGCCGAGGAGGATGTGGAGCCTGAGTG CATCATGGAGAAGGTGGCCAAGGCTTCAGGTGCCAACTACAGCTTTCACAAGGAGAGT GGCCGCTTCCAGGACGTGGGACCCCAGGCCCCAGTGGTGAGTGGCTCTGTGTACCAGA AGACCAATGCCGTGTCTGAGATTAAAGGGTTGGTAAAGACAGCTTCTGGGCCAAAGC AGAGGACCCTGAGACCTTGTGAGAAAGAAATAAAAGAGAAAGAGAGGAGGAGGCACAG CGGCAGCTGGAGCAGGAGCGCCGGGAGCGTGAGCTGCGTGAGGCTGCACGCCGAGAGC AGCGCTATCAGGAGCAGAGGTGGCGAGGCCAGAGCAGGACGTGGGAGCAGCAGCAAGA AGTGGTTTCAAGGAACCGAAATGAGCAGGGGTCAACATGTGCTTCCCTCCAGGAGTCT	

	GCCGTGCACCCGAGGGAGATTTTCAAGCAGAAGGAGAGGGCCATGTCCACCACCTCCA TCTCCAGTCCTCAGCCTGGCAAGCTGAGGAGCCCCTTCCTGCAGAAGCAGCTCACCCA ACCAGAGACCCACTTTGGCAGAGAGCCAGCTGCTGCCATCTCAAGGCCCAGGGCAGAT CTCCCTGCTGAGGAGCCGGCGCCAGCACTCCTCCATGTCTGGTGCAGGCAGAAGAGG AGGCTGTGTATGAGGAACCTCCAGAGCAGGAGACCTTCTACGAGCAGCCCCACTGGT GCAGCAGCAAGGTGCTGGCTCTGAGCACATTGACCACCACATTCAGGGCCAGGGGCTC AGTGGGCAAGGGCTCTGTGCCCCGTGCCCTGTACGACTACCAGGCAGCCGACGACACAG AGATCTCCTTTGACCCCGAGAACCTCATCACGGGCATCGAGGTGATCGACGAAGGCTG GTGGCGTGGCTATGGGCCGGATGGCCATTTTGCATGTTCCCTGCCAACTACGTGGAGC TCATTGAGTGAGGCTGAGGGCACATCTTGCCCTTCCCCTCTCAGACATGGCTTCCTTA TTGCTGGAAGAGGAGGCCTGGGAGTTGACATTCAGCACTCTTCCAGGAATAGGACCCC CAGTGAGGATGAGGCCTCAGGGCTCCCTCCGGCTTGGCAGACTCAGCCTGTCAACCCA AATGCAGCAATGGCCTGGT GATT CCCCACACATCCTTCTGCATCCCCGACCCTCCCA GACAGCTTGGCTCTTGCCCCTGACAGGATACTGAGCCAAGCCCTGCCTGTGGCCAAGC CCTGAGTGGCCACTGCCAAGCTGCGGGGAAGGGTCCTGAGCAGGGGCATCTGGGAGGC TCTGGCTGCCTTCTGCATTTATTTGCCTTTTTTCTTTTTCTTTGCTTCTAAGGGGTG GTGGCCACCACTGTTTAGAATGACCCTTGGGAACAGTGAACGTAGAGAATNGTTTTTA GCAGAGTTGTGACCAAAGTCAGAGTGGATCATGGTGGTTTGGCAGCAGGGAATCTGTC TTGTTGGAGCCTGCTCTGTGCTCCCCACTCCATTTCTGTCCCTCTGCCTGGGCTAT GGGAAGTGGGGATGCAGATGGCAAGCTCCCACCCTGGGTATTCAAAA		
	ORF Start: ATG at 10	ORF Stop: TGA at 1585	
	SEQ ID NO: 114	525 aa	MW at 58507.2kD
NOV22a, CG59375-01 Protein Sequence	MKAHDKPPDLWRLKAESLYRCCGAEEFANYSRSCPALQEAYVRVTEKSPTDWALFTY EGNSNDIRVAGTGEGGLEEMVEELNSGKVMYAFCRVKDPNSGLPKFVLINWTGEGVND VRKGACASHVSTMASFLKGAHVTINARAEEDVEPECIMEKVAKASGANYSFHKESGRF QDVGPPQAPVVSQSVYQKTNVSEIKRVGKDSFWAKAEDPETLSERNKREREEEAQRQL EQERRERELREARREQRYQEQRWRGQSRTWEQQQEVVSRNRNEQGSTCASLQESAVH PREIFKQKERAMSTTSISSPQPGKLRSFPLQKQLTQPETHFGREPAAAI SRPRADLPA EEPAPSTPPCLVQAEAAVEEPEPEQETFYEQPPLVQQQAGSEHIDHHIQGQGLSGQ GLCARALYDYQAADDTEISFDPENLITGIEVIDEGWWRGYGPDGHFACSLPTTWSSLS EAEGTSCPSPLRHGFLIAGRGLGVDIQHSSRNRTPSEDEASGLPPAWQTQPVTPNAA MAW		

Further analysis of the NOV22a protein yielded the following properties shown in Table 22B.

Table 22B. Protein Sequence Properties NOV22a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV22a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
5 homologous proteins shown in Table 22C.

Table 22C. Geneseq Results for NOV22a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV22a Residues/	Identities/ Similarities for	Expect Value

		Residues	Region	
AAB93895	Human protein sequence SEQ ID NO:13840 - Homo sapiens, 439 aa. [EP1074617-A2, 07-FEB-2001]	28..465 3..439	407/440 (92%) 411/440 (92%)	0.0
AAY85662	Human tyrosine kinase substrate tks118/Dresh protein sequence - Homo sapiens, 431 aa. [WO200061750-A2, 19-OCT-2000]	28..465 3..431	399/440 (90%) 403/440 (90%)	0.0
AAB20896	Human dreblin-like protein and SH3 domain sequence SEQ ID NO:1 - Homo sapiens, 431 aa. [JP2000197489-A, 18-JUL-2000]	28..465 3..431	398/440 (90%) 403/440 (91%)	0.0
AAM79569	Human protein SEQ ID NO 3215 - Homo sapiens, 458 aa. [WO200157190-A2, 09-AUG-2001]	28..465 31..458	397/439 (90%) 401/439 (90%)	0.0
AAM78585	Human protein SEQ ID NO 1247 - Homo sapiens, 430 aa. [WO200157190-A2, 09-AUG-2001]	28..465 3..430	397/439 (90%) 401/439 (90%)	0.0

In a BLAST search of public sequence databases, the NOV22a protein was found to have homology to the proteins shown in the BLASTP data in Table 22D.

Table 22D. Public BLASTP Results for NOV22a				
Protein Accession Number	Protein/Organism/Length	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96K74	CDNA FLJ14461 FIS, CLONE MAMMA1000173, HIGHLY SIMILAR TO HOMO SAPIENS SRC HOMOLOGY 3 DOMAIN-CONTAINING PROTEIN HIP-55 MRNA - Homo sapiens (Human), 439 aa.	28..465 3..439	407/440 (92%) 411/440 (92%)	0.0
Q96F30	SIMILAR TO SRC HOMOLOGY 3 DOMAIN-CONTAINING PROTEIN HIP-55 - Homo sapiens (Human), 431 aa.	28..465 3..431	399/440 (90%) 403/440 (90%)	0.0
Q9UJU6	SRC HOMOLOGY 3 DOMAIN-CONTAINING PROTEIN HIP-55 (DREBRIN F) - Homo sapiens (Human), 430 aa.	28..465 3..430	397/439 (90%) 401/439 (90%)	0.0
Q9NR72	CERVICAL SH3P7 (MUCIN-	28..465 3..430	395/439 (89%) 401/439 (90%)	0.0

	sapiens (Human), 430 aa.			
Q62418	DREBRIN-LIKE SH3 DOMAIN-CONTAINING PROTEIN SH3P7 - Mus musculus (Mouse), 433 aa.	29..465 4..433	345/439 (78%) 371/439 (83%)	0.0

PFam analysis predicts that the NOV22a protein contains the domains shown in the Table 22E.

Table 22E. Domain Analysis of NOV22a			
Pfam Domain	NOV22a Match Region	Identities/ Similarities for the Matched Region	Expect Value
cofilin_ADF: domain 1 of 1	35..158	27/151 (18%) 101/151 (67%)	7.8e-21
SH3: domain 1 of 1	408..455	16/58 (28%) 31/58 (53%)	0.0038
Peptidase_M36: domain 1 of 1	486..509	11/24 (46%) 13/24 (54%)	2.9

EXAMPLE 23.

The NOV23 clone was analyzed, and the nucleotide and predicted polypeptide
5 sequences are shown in Table 23A.

Table 23A. NOV23 Sequence Analysis		
	SEQ ID NO: 115	2898 bp
NOV23a, CG59321-01 DNA Sequence	CTCTGGTCACAACCTGCATCAATGACATGCAGAAAGCAGGTCAGGTCACAAGATGCAGC CCTTTCCAGACTCTCTTCGGAGAGTTGGATCTGATGGCAGGAAGTACAAATGGCGAA GCCCTTCCGAATCCATCCCATCAGCTCCTGGGACACTGCCTCATTTCATAGAGGAGC CAGATGATGCTTATATTATCAAGAGCAACCCTATTGCACTCAGGTGCAAAGCGAGGCC AGCCATGCAGATATTCTTCAAATGCAACGGCGAGTGGGTCCATCAGAACGAGCACGTC TCTGAAGAGACTCTGGACGAGAGCTCAGGTTTGAAGGTCCGCGAAGTGTTTCATCAATG TTACTAGGCAACAGGTGGAGGACTTCCATGGGCCCCGAGGACTATTGGTGCCAGTGTGT GGCGTGGAGCCACCTGGGTACCTCCAAGAGCAGGAAGGCCCTCTGTGCGCATAGCCTAT TTACGGAAAACTTTGAACAAGACCCACAAGGAAGGGAAGTTCCCATTTGAAGGCATGA TTGTACTGCACTGCCGCCACCAGAGGGAGTCCCTGCTGCCGAGGTGGAATGGCTGAA AAATGAAGAGCCCATTTGACTCTGAACAAGACGAGAACATTGACACCAGGGCTGACCAT AACCTGATCATCAGGCAGGCACGGCTCTCGGACTCAGGAAATTACACCTGCATGGCAG CCAACATCGTGGCTAAGAGGAGAAGCCTGTGCGCCACTGTTGTGGTCTACGTGAATGG AGGCTGGTCTTCTGACAGAGTGGTCAGCCTGCAATGTTGCTGTGGTAGAGGATGG CAGAAACGTTCCCGACCTGCACCAACCCAGCTCCTCTCAATGGTGGGGCCTTTTGTG AGGGAATGTCAGTGCAGAAAATAACCTGCACTTCTCTTTGTCTGTGGATGGGAGCTG GGAAGTGTGGAGCGAATGGTCCGTCTGCAGTCCAGAGTGTGAACATTTGCGGATCCGG GAGTGCACAGCACCACCCCGAGAAATGGGGGCAAATTCTGTGAAGGTCTAAGCCAGG AATCTGAAAACCTGCACAGATGGTCTTTGCATCCTAACTCCACCACCATGCAGGAACC CAAGGTGACTGCCCTTCAGACGCTATGCCAAATTGAGAATGCCAGCGACATTGCTTTG TACTCGGGCTTGGGTGCTGCCGTGCTGGCGTTGCAGTCTGGTCATTGGTGTACCC TTTACAGACGGAGCCAGAGTGAATATGGCGTGGACGTGATTGACTCTTCTGCATTGAC AGGTGGCTTCCAGACCTTCAACTTCAAACAGTCCGTCAAGGGAAGTCCCTGCTCCTG	

	AATTCTGCCATGCAGCCAGATCTGACAGTGAGCCGGACATACAGCGGACCCATCTGTC TGCAGGACCCTCTGGACAAGGAGCTCATGACAGAGTCCTCACTCTTTAACCCTTTGTC GGACATCAAAGTGAAAGTCCAGAGCTCGTTCATGGTTTCCCTGGGAGTGTCTGAGAGA GCTGAGTACCACGGCAAGAATCATTCCAGGACTTTTCCCCATGGAAACAACCACAGCT TTAGTACAATGCATCCCAGAAATAAAATGCCCTACATCCAAAATCTGTCATCACTCCC CACAAGGACAGAACTGAGGACAACTGGTGTCTTTGGCCATTTAGGGGGGCGCTTAGTA ATGCCAAATACAGGGGTGAGCTTACTCATAACACACGGTGCCATCCCAGAGGAGAATT CTTGGGAGATTTATATGTCCATCAACCAAGGTGAACCCAGGTGAGATGGCTCTGAGGT GCTCCTGAGTCCTGAAGTCACCTGTGGTCCTCCAGACATGATCGTCACCACTCCCTTT GCATTGACCATCCCGCACTGTGCAGATGTGAGTTCTGAGCATTGGAATATCCATTTAA AGAAGAGGACACAGCAGGGCAAATGGGAGGAAGTGATGTGAGTGAAGATGAATCTAC ATCCTGTTACTGCCTTTTGGACCCCTTTGCGTGTGATGTGCTCCTGGACAGCTTTGGG ACCTATGCGCTCACTGGAGAGCCAATCACAGACTGTGCCGTGAAGCAACTGAAGGTGG CGGTTTTTGGCTGCATGTCTGTAACTCCCTGGATTACAACCTTGAGAGTTTACTGTGT GGACAATACCCCTTGTGCATTTTCAAGGAAGTGGTTTTCAGATGAAAGGCATCAAGGTGGA CAGCTCCTGGAAGAACC AAAATTGCTGCATTTCAAAGGGAATACCTTTAGTCTTCAGA TTTCTGTCCTTGATATTCCCCCATTCCTCTGGAGAATTAAACCATTCACTGCCTGCCA GGAAGTCCCGTTCTCCCGCGTGTGGTGCAGTAACCGGCAGCCCCTGCACTGTGCCTTC TCCCTGGAGCGTTATACGCCCACTACCACCCAGCTGTCCTGCAAAAATCTGCATTCCGC AGCTCAAAGGCCATGAACAGATCCTCCAAGTGCAGACATCAATCCTAGAGAGTGAACG AGAAACCATCACTTTCTTCGCACAAGAGGACAGCACTTTCCCTGCACAGACTGGCCCC AAAGCCTTCAAATTCCTACTCCATCAGACAGCGGATTTGTGCTACATTTGATACCC CCAATGCCAAAGGCAAGGACTGGCAGATGTTAGCACAGAAAAACAGCATCAACAGGAG GAATTTATCTTATTTTCGCTACACAAAGTAGCCCATCTGCTGTGATTTTGAACCTGTGG GAAGCTCGTCATCAGCATGATGGTGATCTTGACTCCCTGGCCTGTGCCCTTGAAGAGA TTGGGAGGACACACAGAACTCTCAAACATTTGAGAATCCAGCTTGATGAAGCCGA CTTCAACTACAGCAGGCAAATGGACTCTAGTCCACTTCCTCCCATGAGACAGAGT		
	ORF Start: ATG at 21	ORF Stop: TAG at 2871	
	SEQ ID NO: 116	950 aa	MW at 105960.6kD
NOV23a, CG59321-01 Protein Sequence	MTCRKQVRSQDAALSQTLFGELDLMAGTDNGEALPESIP SAPGTLPHFIEEPDDAYII KSNPIALRCKARPAMQIFFKCNGEVWHQNEHVSEETLDESSGLKVREVFINVTRQQVE DFHGPEDYWCQCVAWSHLGTSKSRKASVRIAYLRKNFEQDPQGREVP IEGMIVLHCRP PEGVPAAEVEWLKNEEPIDSEQDENIDTRADHNLII RQARLSDSGNYTCMAANIVAKR RSLSATVVVYVNGGWSSWTEWSACNVRRCGRGWQKRSRTCTNPAPLNGGAFCEGMSVQK ITCTSLCPVDGSWEVWSEWSVCSPECEHLRI RECTAPPPRNGGKFCEGLSQESENCTD GLCILNSTTMQEPKV TALQTL CQIENASDIALYSGLGAAVVAVAVLVIGVTLYRRSQS DYGVDVIDSSALTGGFQTFNFKTVRQNSLLLSAMQPD LTVSRTYSGPICLQDPLDK ELMTESSLFNPLSDIKVKVQSSFMVSLGV SERAEYHGKNHSRTFPHGNNHSFSTMHPR NKMPYIQNLSSLPTRTELRTTG VFGHLGGRLVMPNTGVSLLI PHGAIPEENSWEIYMS INQGEPRSDGSEVLLSPEVTCGPPDMIVTTPFALT I PHCADVSSEHWN IHLKRTQQG KWEVMSVEDESTSCYCLLDPFACHVLLDSFGTYALTGEPI TDCAVKQLKVAVFGCMS CNSLDYNLRVYCV DNTPCA FQEVVSDERHQGGQLLEEPKLLHFKGNTFSLQISVLDIP PFLWRIKPFTACQEV PFSRVWCSNRQPLHCAFSLERYTPTTTQLSCKICIRQLKGHEQ ILQVQTSILESERETITFFAQEDSTFPAQTGPKAFKIPYSIRQRICATFDTPNAKGKD WQMLAQKNSINRRNLSYFATQSSPSAVILNLWEARHQHDGDLDSLACALEEIGRTHTK LSNISESQLDEADFNYSRQNL		
	SEQ ID NO: 117	2181 bp	
NOV23b, CG59321-02 DNA Sequence	CGGCCAGTCAGAACAACTCCTCCTGTTTTTAATGAATTGGGTTTACCATTGACAATGCT TCCTGATTTTCGGTTGTTGACTTAAGCATGAATAGTAAGAGGCTCTGGTCACAACTGCA TCAATGACATGCAGAAAGCAGGTGCGCGCCGCTGGCTCCCGTGGCTGGGGCTGTGTTTC TGGGCGGGAGGGAACGGGGGTGGCCCAAGGAACTGACAATGGCGAAGCCCTTCCCGAA TCCATCCCATCAGCTCCTGGGACACTGCCTCATTTTCATAGAGGAGCCAGATGATGCTT ATATTATCAAGAGCAACCCTATTGCACTCAGGTGCAAAGCGAGGCCAGCCATGCAGAT ATTCTTCAAATGCAACGGCGAGTGGGTCCATCAGAACGAGCACGTCTCTGAAGAGACT CTGGACGAGAGCTCAGGTTTGAAGGTCCGCGAAGTGTTTCATCAATGTTACTAGGCAAC AGGTGGAGGACTTCATGGGCCCAGGACTATTGGTGCCAGTGTGTGGCGTGGAGCCA CCTGGGTACCTCCAAGAGCAGGAAGGCCTCTGTGCGCATAGCCTATTTACGGAAAAAC TTTGAACAAGACCCACAAGGAAGGGAAGTTCCCATGGAAGGCATGATTGTACTGCACT		

	GCCGCCCACCAGAGGGAGTCCCTGCTGCCGAGGTGGAATGGCTGAAAAATGAAGAGCC CATTGACTCTGAACAAGACGAGAACATTGACACCAGGGCTGACCATAACCTGATCATC AGGCAGGCACGGCTCTCGGACTCAGGAAATTACACCTGCATGGCAGCCAACATCGTGG CTAAGAGGAGAAGCCTGTCGGCCACTGTTGTGGTCTACGTGAATGGAGGCTGGTCTTC CTGGACAGAGTGGTCAGCCTGCAATGTTGCTGTGGTAGAGGATGGCAGAAACGTTCC CGGACCTGCACCAACCCAGCTCCTCTCAATGGTGGGGCCTTTTGTGAGGGAATGTCAG TGCAGAAAATAACCTGCACCTTCTTTGTCTGTGGATGGGAGCTGGGAAGTGTGGAG CGAATGGTCCGTCTGCAGTCCAGAGTGTGAACATTTGCGGATCCGGGAGTGCACAGCA CCACCCCGAGAAATGGGGGCAAATTCTGTGAAGGTCTAAGCCAGGAATCTGAAAACCT GCACAGATGGTCTTTGCATCCTAGATAAAAAACCTCTTCATGAAATAAAACCCCAAAG CATTGAGAATGCCAGCGACATTGCTTTGTAAGTCTGGGCTGGGTGCTGCCGTCGTGGCC GTTGCAGTCCTGGTCATTGGTGTCAACCTTTACAGACGGAGCCAGAGTGACTATGGCG TGGACGTCAATTGACTCTTCTGCATTGACAGGTAACCTCCCTGCTCCTGAATGCGAGCAC ACTCCAGCCTCTGGAGAGACGACAACGCGTGAAGCAACTGAAGGTGGCGGGTTTGGC TGCATGTCCTGTAACCTCCCTGGATTACAACTGGAGAGTTTACTGTGTGGACAAAACCC CTTGGGCTTTTTCAGGAAGTGGTTTTCAGATGAAAGGCATCAAGGGGGACAGCTCCTGGA AGAACCAAAATTGCTGCATTTCAAAGGGAATACCTTTAGTCTTCAGATTTCTGTCCTT GATATTCCCCCATTCCTCTGGAGAATTAAACCATTCACTGCCTGCCAGGAAGTCCCGG TCTCCCGCGTGTGGTGCAGTAACCGGCAGCCCCTGCACTGTGCCTTCTCCCTGGAGCG TTATACGCCCCTACCACCCAGCTGTCCTGCAAAATCTGCATTGCGCAGCTCAAAGGC CATGAACAGATCCTCCAAGTGCAGACATCAATCCTAGAGACTGGCCCCAAAGCCTTCA AAATTCCCTACTCCATCAGACAGCGGATTTGTGCTACATTTGATACCCCAATGCCAA AGGCAAGGACTGGCAGATGTTAGCACAGAAAAACAGCATCAACAGGAATTTATCTTAT TTCGCTACACAAAGTAGCCCATCTGCTGTCATTTTGAACCTGTGGGAAGCTCGTCATC AGCATGATGGTGTCTTACTCCCTGGCCTGTGCCCTTGAAGAGATTGGGAGGACACA CACGAACTCTCAAACATTTCAAGATCCAGCTTGATGAAGCCGACTTCAACTACAGC AGGCAAAATGGACTCTAGTCCACTTCCTCCCATGA		
	ORF Start: ATG at 125	ORF Stop: TAG at 2162	
	SEQ ID NO: 118	679 aa	MW at 75724.8kD
NOV23b, CG59321-02 Protein Sequence	MQKAGRAAGSRGWGCVSGREGTGVAQGTDNGEALPESIP SAPGTLPHFIEEPDDAYII KSNPIALRCKARPAMQIFFKCNGEVWHQNEHVSEETLDESSGLKVREVFINVTRQQVE DFHGPEDYWCQCVAWSHLGTSKSRKASVRIAYLRKNFEQDPQGREVP IEGMIVLHCRP PEGVPAAEVEWLKNEPIDSEQDENIDTRADHNLII RQARLSDSGNYTCMAANIVAKR RSLSATVVVYVNGGWSSWTEWSACNVRRCGRGWQKRSRTCTNPAPLNGGAFCEGMSVQK ITCTSLCPVDGSWEVWSEWSVCSPECEHLRI RECTAPPPRNGGKFCEGLSQESENCTD GLCILDKPLHEIKPQSIENASDIALYSGLGA VAVAVLVIGVTLYRRSQSDYGV DV IDSSALTGNSLLLNASTLQPLERRQRVKQLKVAGFGCMSCNSLDYNWRVYCVDKTPWA FQEVVSDERHQGGQLLEPKLLHFKGNTFSLQISVLDIPPFLWRIKPFTACQEVVSR VWC SNRQPLHCAFS LERYTPTTTQLSKICIRQLKGHEQILQVQTSILETGPKAFKIP YSIRQRICATFDTPNAKGKDWM LAQKNSINRNLSYFATQSSPSAVILNLWEARHQHD GDLD SLACALEEIGRTHTKLSNISESQLDEADFNYSRQNGL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 23B.

Table 23B. Comparison of NOV23a against NOV23b and NOV23c.		
Protein Sequence	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV23b	27..444 27..426	357/419 (85%) 361/419 (85%)

Further analysis of the NOV23a protein yielded the following properties shown in Table 23C.

Table 23C. Protein Sequence Properties NOV23a	
PSort analysis:	0.8411 probability located in mitochondrial inner membrane; 0.7000 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.2057 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV23a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 23D.

Table 23D. Geneseq Results for NOV23a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU12244	Human PRO4326 polypeptide sequence - Homo sapiens, 945 aa. [WO200140466-A2, 07-JUN-2001]	26..925 26..933	499/915 (54%) 651/915 (70%)	0.0
AAW78900	Rat UNC-5 homologue UNC5H-2 - Rattus sp, 943 aa. [WO9837085-A1, 27-AUG-1998]	25..925 23..931	487/916 (53%) 648/916 (70%)	0.0
AAB50691	Human UNC5C protein SEQ ID NO:90 - Homo sapiens, 931 aa. [WO200073328-A2, 07-DEC-2000]	34..936 49..930	465/921 (50%) 621/921 (66%)	0.0
AAW78898	Rat UNC-5 homologue UNC5H-1 - Rattus sp, 898 aa. [WO9837085-A1, 27-AUG-1998]	26..936 23..897	417/921 (45%) 582/921 (62%)	0.0
AAM79128	Human protein SEQ ID NO 1790 - Homo sapiens, 943 aa. [WO200157190-A2, 09-AUG-2001]	24..936 31..942	422/955 (44%) 588/955 (61%)	0.0

In a BLAST search of public sequence databases, the NOV23a protein was found to have homology to the proteins shown in the BLASTP data in Table 23E.

Table 23E. Public BLASTP Results for NOV23a				
Protein Accession Number	Protein/Organism/Length	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
BAB83663	KIAA1777 PROTEIN (UNC5H4) - Homo sapiens (Human), 948 aa.	27..950 30..948	906/926 (97%) 910/926 (97%)	0.0

O08722	TRANSMEMBRANE RECEPTOR UNC5H2 - Rattus norvegicus (Rat), 945 aa.	25..925 25..933	488/916 (53%) 649/916 (70%)	0.0
Q9D398	6330415E02RIK PROTEIN - Mus musculus (Mouse), 945 aa.	1..925 1..933	491/940 (52%) 656/940 (69%)	0.0
O08747	ROSTRAL CEREBELLAR MALFORMATION PROTEIN - Mus musculus (Mouse), 931 aa.	34..936 49..930	468/921 (50%) 622/921 (66%)	0.0
O95185	TRANSMEMBRANE RECEPTOR UNC5C - Homo sapiens (Human), 931 aa.	34..936 49..930	465/921 (50%) 621/921 (66%)	0.0

PFam analysis predicts that the NOV23a protein contains the domains shown in the Table 23F.

Table 23F. Domain Analysis of NOV23a			
Pfam Domain	NOV23a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ig: domain 1 of 1	165..225	16/63 (25%) 43/63 (68%)	5.2e-07
tsp_1: domain 1 of 2	248..297	23/54 (43%) 33/54 (61%)	1.6e-07
tsp_1: domain 2 of 2	304..351	23/54 (43%) 32/54 (59%)	0.0014
ZU5: domain 1 of 1	538..638	33/115 (29%) 68/115 (59%)	7.8e-21
death: domain 1 of 1	855..933	21/87 (24%) 61/87 (70%)	4.4e-13

EXAMPLE 24.

The NOV24 clone was analyzed, and the nucleotide and predicted polypeptide
5 sequences are shown in Table 24A.

Table 24A. NOV24 Sequence Analysis		
	SEQ ID NO: 119	898 bp
NOV24a, CG59591-01 DNA Sequence	CCTGTGACTCCTCCATCCAGCTATGCCCCTGCTGCCCAGCACCGTGGGCCTGGCAGGC CTGCTCTTCTGGGCTGGCCAGGCAGTGAACGCCTTGATAATGCCTAATGCTACCCAG CCCCGGCCCAGCCCGAGAGCACGGCTATGCGGCTCCTGAGTGGCCTGGAGGTGCCAG GTACCGCCGGAAGCGCCACATCTCTGTGAGAGACATGAATGCCTTACTGGATTATCAC AACCACATCCGGGCCAGTGTGTACCCACCTGCCGCCAACATGGAATACATGGTGTGGG ACAAGCGGCTGGCCAGGGCTGCCGAAGCCTGGGCCACCCAGTGCATCTGGGCACATGG GCCTTCACAGCTGATGAGATACGTGGGCCAGAACCTCTCCATCCATTCTGGCCAGTAC CGGTCCGTAGTGGATCTCATGAAGTCCTGGTCTGAGGAGAAGTGGCATTACTTGTTC CGGCCCCAAGGGACTGTAACCCACACTGCCCTGGCGCTGCGATGGCCCCACCTGCTC	

	CCATTATACCCAGATGGTGTGGGCATCCTCCAATCGGCTGGGCTGTGCCATCCACACC TGTAGTAGCATCAGTGTCTGGGGCAACACCTGGCATCGGGCGGCATACCTGGTCTGCA ACTATGCCATTAAGGGCAACTGGATTGGCGAGTCCCCGTACAAGATGGGAAAGCCGTG CTCCTCCTGTCCCCCAGTTATCAAGGCAGCTGCAATAGCAACATGTGCTTCAAGGGG CTGAAATCCAACAAGTTCACGTGGTTCTGAATTTTCTCTGGGCTTTGGTGCGCCTCCA GCTGGGCCTGACCCTCCATGTCTGCCCCTCAAAAACTGGGTGGAGAAATAATTGTTT CTTTAAAGGATATGAGTTAGAATCACCC		
	ORF Start: ATG at 23	ORF Stop: TGA at 782	
	SEQ ID NO: 120	253 aa	MW at 28604.6kD
NOV24a, CG59591-01 Protein Sequence	MPLLPSTVGLAGLLFWAGQAVNALIMP NATPAPAQPESTAMRLLSGLEVPYRRKRHI SVRDMNALLDYHNHIRASVYPPAANMEYMWWDKRLARAAEAWATQCIWAHGPSQLMRY VGQNLSIHSGQYRSVVDLMKSWSEEKWHYLFAPRDCNPHCPWRCDGPTCSHYTQMVW ASSNRLGCAIHTCSSISVWGNTWHRAAYLVCNYAIKGNWIGESPYKMGKPCSSCPSY QGSCNSNMCFKGLKSNKFTWF		

Further analysis of the NOV24a protein yielded the following properties shown in Table 24B.

Table 24B. Protein Sequence Properties NOV24a	
PSort analysis:	0.4400 probability located in lysosome (lumen); 0.3798 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 24 and 25

A search of the NOV24a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
5 homologous proteins shown in Table 24C.

Table 24C. Geneseq Results for NOV24a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR79914	Trypsin inhibitory protein, isolated from human T98G cells - Homo sapiens, 198 aa. [JP07242700-A, 19-SEP-1995]	57..253 1..197	130/197 (65%) 159/197 (79%)	3e-86
AAR79915	Human trypsin inhibitory protein, residues 11-198 - Homo sapiens, 188 aa. [JP07242700-A, 19-SEP-1995]	67..253 1..187	125/187 (66%) 152/187 (80%)	1e-83
AAU29058	Human PRO polypeptide sequence #35 - Homo sapiens, 500 aa. [WO200168848-A2, 20-SEP-2001]	19..243 10..242	112/235 (47%) 155/235 (65%)	3e-70
AAM41693	Human polypeptide SEQ ID NO 6624	19..243 93..325	112/235 (47%) 155/235 (65%)	3e-70

	[WO200153312-A1, 26-JUL-2001]			
AAM39907	Human polypeptide SEQ ID NO 3052 - Homo sapiens, 300 aa. [WO200153312-A1, 26-JUL-2001]	19..243 10..242	112/235 (47%) 155/235 (65%)	3e-70

In a BLAST search of public sequence databases, the NOV24a protein was found to have homology to the proteins shown in the BLASTP data in Table 24D.

Table 24D. Public BLASTP Results for NOV24a				
Protein Accession Number	Protein/Organism/Length	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H3Y0	DJ881L22.3 (NOVEL PROTEIN SIMILAR TO A TRYPSIN INHIBITOR) - Homo sapiens (Human), 253 aa.	1..253 1..253	253/253 (100%) 253/253 (100%)	e-159
O43692	25 KDA TRYPSIN INHIBITOR - Homo sapiens (Human), 258 aa.	37..253 41..257	137/217 (63%) 170/217 (78%)	6e-90
Q98ST6	SUGARCRISP - Gallus gallus (Chicken), 258 aa.	22..253 20..257	140/238 (58%) 179/238 (74%)	8e-90
Q99MM7	SUGARCRISP - Mus musculus (Mouse), 258 aa.	3..253 2..257	144/256 (56%) 186/256 (72%)	1e-89
Q98ST5	COCOACRISP - Gallus gallus (Chicken), 523 aa.	14..243 14..242	111/230 (48%) 158/230 (68%)	6e-71

PFam analysis predicts that the NOV24a protein contains the domains shown in the Table 24E.

Table 24E. Domain Analysis of NOV24a			
Pfam Domain	NOV24a Match Region	Identities/ Similarities for the Matched Region	Expect Value
SCP: domain 1 of 1	67..215	54/173 (31%) 96/173 (55%)	2.9e-21

5

EXAMPLE 25.

The NOV25 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 25A.

Table 25A. NOV25 Sequence Analysis		
	SEQ ID NO: 121	2345 bp
NOV25a,	AGGAGCCGCGATGTTCCCCCTTCGGGCCCTGTGGTTGGTCTGGGCGCTTCTAGGAGTG	

CG59588-01 DNA Sequence	GCCGGATCATGCCCCGAGCCGTGCGCCTGCGTGGACAAGTACGCTCACCAGTTCGCGG ACTGCGCTTACAAAGAGTTGCGTGAGGTGCCGGAAGGACTGCCTGCCAACGTGACGAC GCTTAGTCTGTCCGCGAACAAGATCACTGTGCTGCGGCGCGGGGCTTCGCCGACGTC ACACAGGTCACGTCGCTGTGGCTGGCGCACAAATGAGGTGCGCACCGTGGAGCCAGGCG CACTGGCCGTGCTGAGTCAGCTCAAGAACCTCGATCTGAGCCACAACCTTCATATCCAG CTTCCGTGGAGCGACCTGCGCAACCTGAGCGCGCTGCAGCTGCTCAAAATGAACCAC AACCGCCTGGGCTCTCTGCCCCGGGACGCACTCGGTGCGCTACCCGACCTGCGTTCCC TGCGCATCAACAACAACCGGCTGCGTACGTGGCGCCTGGCACCTTCGACGCGCTTAG CGCGCTGTCACACTTGCAACTCTATCACAATCCCTTCCACTGCGGCTGCGGCCTTGTG TGGCTGCAGGCCTGGGCCGCGAGCACCCGGGTGTCCTTACCCGAGCCCGACTCCATTG CTTGTGCCTCGCCTCCCGCGCTGCAGGGGGTGCCGGTGTACCGCCTGCCCGCCCTGCC CTGTGCACCGCCAGCGTGATCTGAGTGCCGAGCCACCGCTTGAAGCACCCGGCACC CCACTGCGCGCAGGACTGGCGTTTCGTGTTACTGTCATCGCCGACGGCCACCCTACGC CTCGCCTGCAATGGCAACTTCAGATCCCCGGTGGCACCGTAGTCTTAGAGCCACCGGT TCTGAGCGGGGAGGACGACGGGGTTGGGGCGGAGGAAGGAGAGGGAGAAGGAGATGGG GATTTGCTGACGCAGACCCAAGCCCAAACGCCGACTCCAGCACCCGCTTGGCCGGCGC CCCCAGCCACACCGCGCTTCCTGGCCCTCGCAAATGGCTCCCTGTTGGTGCCCTCCT GAGTGCCAAGGAGGCGGGCGTCTACACTTGCCGTGCACACAATGAGCTGGGCGCCAAC TCTACGTCAATACGCGTGCGGGTGGCAGCAACCGGGCCCCCAAACACGCGCCTGGCG CCGGGGGAGAACCCGACGGACAGGCCCGACCTCTGAGCGCAAGTCCACAGCCAAGGG CCGGGGCAACAGCGTCCTGCCTTCCAAACCCGAGGGCAAAATCAAAGGCCAAGGCCTG GCCAAGGTCAGCATTCTCGGGGAGACCGAGACGGAGCCGGAGGAGGACACAAGTGAGG GAGAGGAGGCCGAAGACCAGATCCTCGCGGACCCGGCGGAGGAGCAGCGCTGTGGCAA CGGGGACCCCTCTCGGTACGTTTCTAACCACGCGTTCAACCAGAGCGCAGAGCTCAAG CCGCACGTCTTCGAGCTGGGCGTCATCGCGCTGGATGTGGCGGAGCGCGAGGCGCGGG TGCAGCTGACTCCGCTGGCTGCGCGCTGGGGCCCTGGGCCCGGCGGGGCTGGCGGAGC CCCGCGACCCGGGCGGCGACCCCTGCGCCTACTCTATCTGTGTCCAGCGGGGGGCGGC GCGGCAGTGCAGTGGTCCCGCGTAGAGGAAGGCGTCAACGCCTACTGGTTCGCGGGCC TGCGGCCGGGTACCAACTACTCCGTGTGCCTGGCGCTGGCGGGCGAAGCCTGCCACGT GCAAGTGGTGTTCCTACCAAGAAGGAGCTCCCATCGCTGCTGGTCATAGTGGCAGTG AGCGTATTCTCCTGGTGCTGGCCACAGTGCCCCCTTCTGGGCGCCGCTGCTGCCATC TGCTGGCTAAACACCCGGGCAAGCCCTACCGTCTGATCCTGCGGCCTCAGGCCCTGA CCCTATGGAGAAGCGCATCGCCGCACTTCGACCCGCGTGCTTCGTACCTCGAGTCC GAGAAAAGCTACCCGGCAGGCGGCGAGGCGGGCGGCGAGGAGCCAGAGGACGTGCAGG GGGAGGGCCTTGATGAAGACGCGGAGCAGGGAGACCCAAGTGGGGACCTGCAGAGAGA GGAGAGCCTGGCGGCCTGCTCACTGGTGGAGTCCCAGTCCAAGGCCAACCAAGAGGAG TTCGAGGCGGGCTCTGAGTACAGCGATCGGCTGCCCCCTGGGCGCCGAGGCGGTCAACA TCGCCCAGGAGATTAATGGCAACTACAGGCAGACGGCAGGCTGAACCTCCGCCCGTCC GGCCCGCCCATTCCTGACCTCCACCTAGGGTGCCTGGGAGCAGCAGTCTAGGGCTGGC AGGACTTATGTCCCCCGTCCCCAAC
	ORF Start: ATG at 11 ORF Stop: TGA at 2246
	SEQ ID NO: 122 2345 aa MW at 78989.2 kD
NOV25a, CG59588-01 Protein Sequence	MFPLRALWLWVALLGVAGSCPEPCACVDKYAHQFADCAKELREVPEGLPANVTTLSL SANKITVLRGAFADVTQVTSWLAHNEVRTVEPGALAVLSQLKNLDSLHNFISFPW SDLRNLALQLLKMNHNLGSLPRDALGALPDLRLRINNNRLRTLAPGTFDALSALS HLQLYHNPFFHCGGLVWLQAWAASRVSLPEPDSIACASPPALQGVVPVYRLPALPCAP PSVHLSAEPPEAPGTPLRAGLAFVLHLCIADGHPTPRLQWQLQIPGGTVVLEPPVLSG EDDGVGAEEGEGEGDGLLTQTQAQTPTPAPAWPAPPATPRFLALANGSLLVPLLSAK EAGVYTCRAHNELGANSTSIRVAVAATGPPKHAPGAGGEPDGQAPTSEKSTAKGRGN SVLPSKPEGKIKQGLAKVSILGETETETEPEEDTSEGEAEQDQILADPAEEQRCNGDP SRYVSNHAFNQSAELKPHVFELGVIALDVAEREARVQLTPLAARWGPGPGGAGGAPRP GRRPLRLLYLCPAGGGAAVQWSRVEEGVNAYWFRGLRPGTNYSVCLALAGEACHVQVV FSTKKELPSLLVIVAVSVFLLVLATVPLLGAACCHLLAKHPGKPYRLILRPQAPDME KRIAADFDPRASYLESEKSYAGGEAGGEEPEDVQGEGLDEDAEQGDPGDLQREESL AACSLVESQSKANQEEFEAGSEYSDRLPLGAEAVNIAQEINGNYRQTAG

Further analysis of the NOV25a protein yielded the following properties shown in Table 25B.

Table 25B. Protein Sequence Properties NOV25a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Likely cleavage site between residues 19 and 20

A search of the NOV25a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 25C.

Table 25C. Geneseq Results for NOV25a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE09450	Human sbg34976IGBa protein #1 - Homo sapiens, 745 aa. [WO200160850-A1, 23-AUG-2001]	1..745 1..745	745/745 (100%) 745/745 (100%)	0.0
AAU12205	Human PRO4329 polypeptide sequence - Homo sapiens, 745 aa. [WO200140466-A2, 07-JUN-2001]	1..745 1..745	745/745 (100%) 745/745 (100%)	0.0
AAB40448	Human ORFX ORF212 polypeptide sequence SEQ ID NO:424 - Homo sapiens, 209 aa. [WO200058473-A2, 05-OCT-2000]	265..472 2..209	208/208 (100%) 208/208 (100%)	e-120
AAM93734	Human polypeptide, SEQ ID NO: 3699 - Homo sapiens, 428 aa. [EP1130094-A2, 05-SEP-2001]	1..417 1..385	217/426 (50%) 260/426 (60%)	e-107
AAU12317	Human PRO215 polypeptide sequence - Homo sapiens, 428 aa. [WO200140466-A2, 07-JUN-2001]	1..417 1..385	217/426 (50%) 260/426 (60%)	e-107

In a BLAST search of public sequence databases, the NOV25a protein was found to have homology to the proteins shown in the BLASTP data in Table 25D.

Table 25D. Public BLASTP Results for NOV25a				
Protein Accession Number	Protein/Organism/Length	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9P263	KIAA1465 PROTEIN - Homo	104..745 1..642	642/642 (100%) 642/642 (100%)	0.0

	(fragment).			
O14498	ISLR PRECURSOR - Homo sapiens (Human), 428 aa.	1..417 1..385	217/426 (50%) 260/426 (60%)	e-106
BAA85972	ISLR PRECURSOR - Mus musculus (Mouse), 428 aa.	4..417 1..385	209/421 (49%) 258/421 (60%)	e-102
O88279	MEGF4 - Rattus norvegicus (Rat), 1531 aa.	20..246 734..933	77/232 (33%) 113/232 (48%)	1e-25
Q9WVB5	SLIT1 - Mus musculus (Mouse), 1531 aa.	20..246 734..933	77/232 (33%) 113/232 (48%)	1e-25

PFam analysis predicts that the NOV25a protein contains the domains shown in the Table 25E.

Table 25E. Domain Analysis of NOV25a			
Pfam Domain	NOV25a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LRRNT: domain 1 of 1	19..50	12/33 (36%) 20/33 (61%)	0.27
LRR: domain 1 of 5	52..75	7/25 (28%) 18/25 (72%)	1.4
LRR: domain 2 of 5	76..99	5/25 (20%) 18/25 (72%)	31
LRR: domain 3 of 5	100..123	11/25 (44%) 20/25 (80%)	0.0026
LRR: domain 4 of 5	124..147	10/25 (40%) 18/25 (72%)	0.099
LRR: domain 5 of 5	148..171	9/25 (36%) 19/25 (76%)	0.14
LRRCT: domain 1 of 1	181..231	19/54 (35%) 41/54 (76%)	1.5e-14
ig: domain 1 of 1	253..357	13/108 (12%) 70/108 (65%)	0.01

EXAMPLE 26.

The NOV26 clone was analyzed, and the nucleotide and predicted polypeptide
5 sequences are shown in Table 26A.

Table 26A. NOV26 Sequence Analysis		
	SEQ ID NO: 123	4426 bp
NOV26a,	ATGTACTGTTTTTTCGATCTGCCGTTTCGTTCTTCTGCCTCAGCCTCCCCAGGTGCT	

CG59584-01 DNA Sequence	GGGGTTATAGGTGTGAGCCACTGTGCCTGGCTATTCTTTTATTACAGTATGTTCTGCT GATTCCTTCTGTTCTACAAGAAGGCTCTTTGGATAAAGCTTGTGCCAGCTTTTAAAT CTCACTGAATCTGTTGTTTTGACGGTCTCCCTCAACTATGGTGAGGTCCAGACCAAAA TATTTGAAGAAAATGTTACTGGAGAAAATTTCTTCAAATGCATCAGCTTTGAGGTTCC TCAGGCCAGATCTGACCCACTGGCATTATTACATTTTCTGCTAAAGGAGCCACTCTC AACCTGGAAGAGAGGAGATCTGTGGCAATCAGATCCAGAGAGAATGTGGTCTTTGTAC AGACTGATAAACCACCTACAAGCCTGGACAGTATAATAAAAAGCCGATCAGTCACAT AATGCCAGTGATAGCAGTCACTGAACAGGATCCAGAAGGCAATCGAATACAACAGTGG GTGAATGAGGAGTCTGTGGGAGGGATTCTACAACCTCTCCTTCCAGTTAATCTCAGAGC CCATCCTCGGATGGTATGAAATCACCGTGGAGATGCTCAATGAGAAGAAAACATATCA CTCCTTCTCTGTGGAAGAATATGTGTTACCCAAATTTCAAATGACTGTGGATGCACCA GAAAATATCTTAGTTGTGGACTCTGAATTCAAAGTGAATGTCTGTGCCTTGTATACCT ATGGTGAACCTGTGGACGGGAAGGTCCAACCTTAGTGTGTGCAGAGAATCTACGGCTTA TCATTCATGTGCTCATCTTATCAGTTCACCTCTGTAAAAATTTTACCTTGGGGAAAGAT GGCTGTGTCTCCAAGTTTATTAACACAGATGCTTTTGAGTTAAATCGGGAAAGGATACT GGAGTTTCCTCAAAGTGCATGCTCTTGTTACAGAGCTTACAGGCTCCAAGTACGTATA CATAGACTCATCAGTGGTGAAGATTAGTTTTGAGAATATGGATATGTCTACAAACAG GGACTCCCTTATTTTGGCCAGATTAAATTGCTTAATCCAGACAACCTCTCCAATCCCAA ATGAAGTTGTTCAAGTGCATCTGAAGGACAAAATCGTGGGAAACTACACCACAGATGT AAATGGCATCGCTCAGTTTTTCTTGGACACATATACGTTTACATACCCAAATATCACT TTGAAAGCAGCCTACAAGGCCAATGAAAATTGCCAGGCTCATGGCTGGGTGTTGCCTC AATACCCTCAGCCCGAGTACTTTGCATATCGATTTTACTCCAAGATGAATAGCTTCCT AAAGATTGTCCAAGAGATGGAAGAACTGAGATGCAACCAGCAGAAGAGGGTGCTAGTG CACTGCATTCTCAATATGGAAGACTTTGAAGACAAAACCTACACAGCAGACTTCAATT ATTTGGTGATTTCAAAGGTGTAATCATTCTTCATGGGCAACAGAAAATTGAGATCAA CGAAAATGGGAGGAAGGGCATATTTTCCATTTCTATAGACATTAACCCTGAATTAGCG CCCTCAGTACATATGCTTGTCTATAGCTTGCATCCTGGAGGAGAAATGGTCACTGATA GCACCCAATTCCAATTGAGAAATGTTAACATAAAGTTCTCTAACGAGCAGGGCTTACC TGGTTCCAATGCTAGTCTCTGTCTTCAAGCGGCGCCTGTCTTATTCTGTGCCCTCAGG GCTGTGGATAGGAATGTCCTTCTACTGAAATCTGAACAACAGCTGTCAGCTGAAAGTG TGTATAACATGGTTCCAAGTATAGAGCCGTATGGTTATTTCTACCATGGCCTCAATCT TGATGATGGCAAGGAAGACCCTTGCAATTCCTCAGAGGGATATGTTCTACAATGGTTTA TATTACACACCTGTAAGCAACTATGGGGATGGAGATATCTATAATATTGTCAGGAACA TGGGTCTAAAAGTCTTTACCAATCTCCATTACCGAAAACCAGAAGTATGTGTGATGGA GAGAAGGCTGCCACTCCCTAAGCCGCTTTATCTGGAAACAGAAAATTATGGTCCAATG CGTAGTGTTCCGTCTAGAATTGCATCTAGTGGAAATCAGAGGGGAGAATGCTGACTATG TAGAACAGGCTATAATTCAAACAGTAAGAACAACTTCCCAGAGACATGGATGTGGGA CCTCGTCAGTGTGATTCTCAGGCTCTGCCAATCTTTCGTTCCCTCATTCTGATACG ATAACCCAATGGGAGGCAAGTGGCTTTTGTGTGAATGGTGACGTTGGATTTGGCATT CCTCTACAACCACTCTAGAAGTCTCCCAACCTTTCTTTATTGAGATTGCCTCACCTT TTCGGTTGTTCAAATGAACAATTTGATTTGATTGTCAATGTCTTCAGCTACCGGAAT ACATGTGTAGAGATTTCTGTTCAAGTGGAGGAGTCTCAGAATTATGAAGCAAATATTC ATACCTTGAAAATCAATGGCAGTGAGGTTATTCAAGCTGGAGGGAGGAAACAAACGT CTGGACTATTATACCTAAGAAATTGGGCAAAGTGAATATCACTGTAGTTGCTGAGTCC AAACAAAGCAGTGCTTGCCCAAATGAAGGAATGGAGCAGCAAAGCTAAACTGGAAAG ACACTGTGGTCCAAAGCTTCTTAGTAGAGCCTGAAGGTATTGAAAAGGAAAGGACCCA GAGTTTCCTTATCTGTACAGAAGGTGCCAAAGCCTCCAAGCAGGGAGTTTTGGACTTG CCAAACGATGTAGTAGAAGGGTCAGCCAGAGGCTTTTTCACTGTTGTGGGGATATT TAGGACTTGCCTTGCAGAATCTGGTTGTTCTCCAAATGCCCTATGGAAGTGGAGAGCA GAATGCTGCCCTACTAGCATCTGATACTTATGTTCTGGACTATCTGAAATCTACTGAG CAACTGACAGAGGAAGTTCAATCTAAGGCTTTCTTTCTCTTATCTAATGGTTATCAA GGCAATTATCTTCAAACAACTCTGATGGTTCCTATAGTGTGTTTTGGCAGCAGAGTCA GAAAGGAAGCATATGGCTCAGTGCTCTTACTTTTAAGACATTGGAGAGAATGAAAAAA TATGTATTCAATTGATGAAAATGTTCAAACACAGACCTTAATCTGGCTTTCAAGCCAAC AGAAAACAAGCGGCTGCTTTAAGAATGATGGCCAGCTTTTCAACCACGCCTGGCAGGG TGGAGATGAAGAGGACATTTCACTCACTGCGTATGTTGTTGGGATGTTCTTTGAAGCT GGGGCGGCATTGGACAGTGGTGTCACTAATGGCTATAATCATGCAATTCTAGCTTATG CTTTTGCCTTAGCTGGAAAAGAGAAGCAAGTGGAAATCTTTACTCCAACCTGGATCA ATCTGCCCCAAACTAAATAATGTCATCTACTGGGAAAGAGAAAGGAAACCAAGACA GAAGAATTTCCATCCTTTATTCCCTGGGCACCTTCTGCTCAGACTGAGAAGAGTTGCT ACGTGCTGTTGGCTGTCATTTCCCGGAAATTCCTGACCTCACCTATGCTAGTAAGAT
----------------------------	--

	TGTGCAGTGGCTTGCCCAACGGATGAATTCCCATGGAGGCTTTTCTTCCAACCAGGAT CAAAACACTGTCACCTTTAGCAGTGAAGGATCCAGTGAGATTTTCCAGGTTAACGGTC ATAACCGCCTACTGGTCCAACGTTTCAAGTAACACAGGCACCTGGAGAATACACAGT AGATGTGGAAGGACACGGTTGTACATTTATCCAGGCCACCCTTAAGTACAATGTTCTC CTACCTAAGAAGGCATCTGGATTTTCTCTTTCCTTGGAATAGTAAAGAACTACTCTT CGACTGCTTTTGACCTCACAGTGACCCTCAAATACACTGGAATTCGCAATAAATCCAG TATGGTGGTTATAGATGTAAAAATGCTATCAGGATTTACTCCAACCATGTCATCCATT GAAGAGCTTGAAAACAAGGGCCAAGTGATGAAGACTGAAGTCAAGAATGACCATGTTT TTTTCTACTTGGAATGTAGGTTTTGGTTCGAGCAGACAGTTTCCCTTTTTCTGTTGA GCAGAGCAACCTTGTGTTCAACATTCAGCCAGCCCCAGCCATGGTCTACGATTATTAT GAAAAAGAAGAATATGCCCTAGCTTTTTTACAACATCGACAGTAGTTCAGTTTCCGAGT GAGACAAAGCAATTACTAGAAGAGTTGGAGAAGCATTCTTGTAACAACTGATTCTT CTGTATCAAACCTGGAAAAAATCATGAACCATCTGACATCGTGAACAGTCTGCAGTG GGCTATGGTTTCTTGTCAAGTCTTATTTCTTATCATCCCATTAATGTTGTCATTTT GCAAAAAAAAAAAAAA		
	ORF Start: ATG at 1	ORF Stop: TGA at 4234	
	SEQ ID NO: 124	1411 aa	MW at 158867.0kD
NOV26a, CG59584-01 Protein Sequence	MYCFLRSAVSFFCLSLPRCWGYRCEPLCLAILLLQYVLLIPSVLQEGSLDKACAQLFN LTESVVLTVSLNYGEVQTKIFEENVVTGENFFKCISFEVPQARSDPLAFITFSAGKATL NLEERRSVAIRSRENVVVFQTDKPTYKPGQYNKKPI SHIMPVIAVTEQDPEGNRIQQW VNEESVGGILQLSFQLISEPILGWYEITVEMLNEKKTYHSFSVEEYVLPKFQMTVDAP ENILVVDSEFKVNVCALYTYGEPVDGKVQLSVCRESTAYHSCAHLISSLCKNFTLGKD GCVSKFINTDAFELNREGYWSFLKVHALVTELTGSKYVYIDSSVVKISFENMDMSYKQ GLPYFGQIKLLNPDNSPIPNEVVQLHLKDKIVGNYTTDVNGIAQFFLDITYTFTYPNIT LKAAYKANENCQAHGWVLPQYPQPEYFAYRFYSKMNSFLKIVQEMEELRCNQQRVLV HCILNMEDFEDKTYTADFNVLVISKGVIIILHGQQKIEINENGRKGIFISISIDINPELA PSVHMLVYSLHPGGEMVTDSTQFQLRNVNIKFSNEQGLPGSNASLCLQAAPVLFALR AVDRNVLLKSEQQLSAESVYNMVPSIEPYGYFYHGLNLDGKEDPCIPQRDMFYNGL YYTPVSNYGDGDIYNIVRNMGLKVFTNLHYRKPEVCMERRLPKPLYLETENYGP RSVPSRIASSGIRGENADYVEQAI IQTVRTNFPETWMWDLVSVDSGSANLSFLIPDT ITQWEASGFCVNGDVGFGISSTTTLEVSQPFIEIASPFSVVQNEQFDLIVNVFSYRN TCVEISVQVEESQNYEANIHTLKINGSEVIQAGGRKTNVWTIIPKKLGKVNITVVAES KQSSACPNEGMEQQKLNWKDTVVSFLVEPEGIEKERTQSFLICTEGAKASKQGVLDL PNDVVEGSARGFFTIVGDI LGLALQNLVVLQMPYGSGEQNAALLASDTYVLDYLKSTE QLTEEVQSKAFFLLSNGYQRQLSFKNSDGSYSVFWQQSQKGSIWLSALTFKTLERMKK YVFIDENVQKQTLIWLSSQKQTSKCFKNDGQLFNHAWQGGDEEDISLTAYVVGMMFFEA GAALDSGVTNGYNHAILAYAFALAGKEKQVESLLQTLQDQSAKLNVIYWERERKPKT EEFPSFIPWAPSAQTEKSCYVLLAVISRKIPDLTYASKIVQWLAQRMNSHGGFSSNQD QNTVTFSSEGSSEIFQVNGHNRLLVQRSEVTQAPGEYTVDVEGHGCTFIQATLKYNVL LPKKASGFSLSLEIVKNYSSTAFDLTVTLKYTGIRNKSSMVVIDVKMLSGFTPTMSSI EELENKGQVMKTEVKNDHVLFYLENVGFGRADSFPFSVEQSNLVFNIQPAPAMVYDYY EKEEYALAFYNIDSSSVSE		
	SEQ ID NO: 125	4501 bp	
NOV26b, CG59584-02 DNA Sequence	TCCATTTCTATAGACATTAACCCTGAATTAGCGCCCTCAGTAGATATGCTTGTCTATA GCTTGCATCCTGGAGGAGAAATGGTCACTGATAGCACCCAATTCCGAATTGAGAAATG CTTCGAAAATCAGGTCAACTTAAATTTTTCTAAAGAAAAAAGTTTACCAGGATCCAAT ATTGATCTTCAAGTCTCGGCTGCTTCAAACCTCTCTTTGTGCTCTTTGGGCTGTAGACC AGAGTGTATTGCTACTAAGGAATTATGGTCAGCTGTCAGCACAACTGTGTATAGTCA GCTATATTCCAGGGAACATACATGGCTATTACTTCAGAGGACTTAACTTAGAAGATGGC CTTAAAGTGCCGTGTCTTGAAGATGAACATATCCTTTACAATGGAATTTATTACACAC CTGCATGGGCTGACTTTGGAAAAGATGGCTATGACCTTGTGAAGGATCCTCAAACAA TCGGATTTTTCAAAGGCAAAATGTGACTTCTTTCCGAAATATTACCCAACCTCTCGTTC CAACTGATTTTCAGAACCAATGTTTGGAGATTACTGGATTGTTGTGAAAAGAACTCAA GGGAGACAGTGACACACCAATTTGCTGTAAAAGATATGTGCTGCCCAAGTTTGAAGT TACAGTCAATGCACCACAAACAGTAACTATTTTCAAGATGATGAATTCCAAGTGGATGTA TGTGCTAAGTACAACCTTTGGCCAACCTGTGCAAGGGGAAACCCAAATCCGGGTGTGCA GAGAGTATTTTTCTTCAAGCAATTGTGAGAAAAATGAAAATGAAATATGTGAGCAATT TATTGCACAGTTGGAAAATGGTTGTGTTTCTCAAATTGTAAATACAAAAGTCTTCCAA		

CTCTACCGTTCGGGATTGTTTCATGACATTTTCATGTCGCTGTAATTGTTACAGAATCTG
 GGACAGTTATGCAGATCAGCGAGAAGACCTCAGTTTTTATCACTCAATTGCTTGGAAC
 TGTAACCTTTGAGAACATGGATACATTCTATAGAAGAGGGATTCTTATTTTGGAAC
 CTTAAATTTTCGGATCCCAATAATGTACCTATGGTGAACAAGTTGTTGCAACTGGAGC
 TCAATGATGAATTTATAGGAAATTACACTACGGATGAGAATGGCGAAGCTCAATTTTC
 CATTGACACTTCAGACATATTTGATCCAGAGTTCAACCTAAAAGCCACATATGTTCTGA
 CCTGAGAGCTGCTATCTTCCCAGCTGGTTGACGCCTCAGTACTTGGATGCTCACTTCT
 TAGTCTCACGCTTTTACTCCCGAACCAACAGCTTCCTGAAGATTGTTCCAGAACCAAA
 GCAGCTTGAATGTAATCAACAGAAGGTTGTTACTGTGCATTACTCCCTAAACAGTGAA
 GCATATGAGGATGATTCCAATGTAAAGTTCTTCTATTTGATGATGGTAAAAGGAGCTA
 TCTTACTCAGTGGACAAAAGGAAATCAGAAACAAAGCCTGGAATGGAACTTCTCGTT
 CCCAATCAGCATCAGTGCTGATCTGGCTCCTGCAGCCGTCTGTTTGTCTATACCCTT
 CACCCCAGTGGGGAAATTGTGGCTGACAGTGTGAGATTCCAGGTTGACAAGTGCTTTA
 AACACAAGGTTAACATAAAGTTCTCTAACGAGCAGGGCTTACCTGGTTCCAATGCTAG
 TCTCTGTCTTCAAGCGGCGCCTGTCTTATTCTGTGCCCTCAGGGCTGTGGATAGGAAT
 GTCCTTCTACTGAAATCTGAACAACAGCTGTGAGCTGAAAGTGTGTATAACATGGTTC
 CAAGTATAGAGCCGTATGGTTATTTCTACCATGGCCTCAATCTTGATGATGGCAAGGA
 AGACCCCTGCATTCTCAGAGGGATATGTTCTACAATGGTTTATATTACACACCTGTA
 AGCAACTATGGGGATGGAGATATCTATAATATTGTCAGGAACATGGGTCTAAAAGTCT
 TTACCAATCTCCATTACCGAAAACCAGAAAAAATATGGTCCAATGCGTAGTGTTCCG
 TCTAGAATTGCATGTAGCTAGTGGAAATCAGAGGGGAGAATGCTGACTATGTAGAACAG
 GCTATAATTCAAACAGTAAGAACAACTTCCCAGAGACATGGATGTGGGACCTCGTCA
 GTGTCGATTCTCAGGCTCTGCCAATCTTTCGTTCTCCTCATTCTGATACGATAACCCA
 ATGGGAGGCAAGTGGCTTTTGTGTGAATGGTGACGTTGGATTGTCATTCTCTACA
 ACCACTCTAGAAGTCTCCCAACCTTTCTTTATTGAGATTGCCTCACCTTTTCGGTTG
 TTCAAAATGAACAATTTGATTGATTGTCAATGTCTTCAGCTACCGGAATACATGTGT
 AGAGATTTCTGTTCAAGTGGAGGAGTCTCAGAATTATGAAGCAAATATTCATACCTTG
 AAAATCAATGGCAGTGAGGTTATTCAAGCTGGAGGGAGGAAAACAAACGTCTGGACTA
 TTATACCTAAGAAATTGGGTAAAGTGAATATCACTGTAGTTGCTGAGTCCAAACAAAG
 CAGTGCTTGCCCAAATGAAGGAATGGAGCAGCAAAAGCTAAACTGGAAAGACACTGTG
 GTCCAAAGCTTCTTAGTAGAGCCTGAAGGTATTGAAAAGGAAAGGACCCAGAGTTTCC
 TTATCTGTACAGAAGGTGCCAAAGCCTCCAAGCAGGGAGTTTGGACTTGCCAAACGA
 TGTAGTAGAAGGGTCAGCCAGAGGCTTTTCTACTGTTGTGGGGGATATTCTAGGACTT
 GCCTTGACAGAACTGCTGTTGTTCTCCAAATGCCCTATGGAAGTGGAGAGCAGAAATGCTG
 CCCTACTAGCATCTGATACTTATGTTCTGGACTATCTGAAATCTACTGAGCAACTGAC
 AGAGGAAGTTCAATCTAAGGCTTTCTTTCTCTTATCTAATGGTTATCAAAGGCAATTA
 TCTTTCAAAAACCTCTGATGGTTCCTATAGTGTGTTTTGGCAGCAGAGTCAGAAAGGAA
 GCATATGGCTCAGTGCTCTTACTTTTAAGACATTGGAGAGAATGAAAAAATATGTATT
 CATTGATGAAAATGTTCAAAAACAGACCTTAATCTGGCTTTCAAGCCAACAGAAAACA
 AGCGGCTGCTTTAAGAATGATGGCCAGCTTTTCAACCACGCCTGGGAGGGTGGAGATG
 AAGAGGACATTTCACTCACTGCGTATGTTGTTGGGATGTTCTTTGAAGCTGGGCTCAA
 TTTCACTTTTCTGCTCTACGAAACGCACTCTTTTGCTTGAAGCGGCATTGGACAGT
 GGTGTCACCTAATGGCTATAATCATGCAATTCTAGCTTATGCTTTTGCCTTAGCTGGAA
 AAGAGAAGCAAGTGGAAATCTTACTCCAAACCCTGGATCAATCTGCCCCAAACTAAA
 TAATGTCATCTACTGGGAAAGAGAAAGGAAACCCAAGACAGAAGAATTTCCATCCTTT
 ATTCCCTGGGCACCTTCTGCTCAGACTGAGAAGAGTTGCTACGTGCTGTTGGCTGTCA
 TTTCCCGGAAAATTCCTGACCTCACCTATGCTAGTAAGATTGTGCAGTGGCTTGCCCA
 ACGGATGAATTCCCATGGAGGCTTTTCTTCCAACCAGACACCTGATGATACTCTGTTT
 AAATTATATACGGGCCAAAAGAAAGCTTTTCGCTCTAGTTCTGTGGGCTATACACTGG
 GAAAAGCAAATGAAAAGAAGGAAAACAGGAGAAATGGGGGTGAAGGATCCAGTGAGAT
 TTTCCAGGTTAACGGTCATAACCGCCTACTGGTCCAACGTTTCAAGTAACACAGGCA
 CCTGGAGAATACACAGTAGATGTGGAAGGACACGGTTGTACATTTATCCAGGCCACCC
 TTAAGTACAATGTTCTCCTACCTAAGAAGGCATCTGGATTTTCTCTTTCTTGGAAAT
 AGTAAAGAACTACTCTTCGACTGCTTTTGACCTCACAGTGACCCTCAAATACACTGGA
 ATTCGCAATAAATCCAGTATGGTGGTTATAGATGTAAAAATGCTATCAGGATTTACTC
 CAACCATGTCATCCATTGAAGAGCTTGAAAACAAGGGCCAAGTGATGAAGACTGAAGT
 CAAGAATGACCATGTTCTTTTCTACTTGAAAATGTAGGTTTGGTTCGAGCAGACAGT
 TTCCCTTTTCTGTTGAGCAGAGCAACCTTGTGTTCAACATTCAGCCAGCCCCAGCCA
 TGGTCTACGATTATTATGAAAAGAAGAATATGCCCTAGCTTTTACAACATCGACAG
 TAGTTTCAAGTTTCCGAGTGAGACAAAGCAATTACTAGAAGAGTTGGAGAAGCATTTCTT
 GTAACAACTGATTCTTCTGTATCAAACCTGGAAAAAATCATGAACCATCTGACATC

	GTGAACAGTCTGCAGTGGGCTATGGTTTCTTGTCAAGTCTTATTTCTTATCATCCCA TTAAATGTTGTGTCATTTTGCAAAAAAAAAAAAAAAAAA		
	ORF Start: TCC at 1	ORF Stop: TGA at 4309	
	SEQ ID NO: 126	1436 aa	MW at 161836.4kD
NOV26b, CG59584-02 Protein Sequence	SISIDINPELAPSVDMVLVYSLHPGGEMVTDSTQFRIEKCFCENQVNLNFSKEKSLPGSN IDLQVSAASNSLCALWAVDQSVLLLRNYGQLSAQTVYSQLYSRELHGYIFRGLNLEDG LKVPCLDEDEHILYNGIYYTPAWADFGKDGDLVKDPQNNRIFQRQNVTSFRNITQLSF QLISEPMFGDYWIVVKRNSRETIVTHQFAVKRYVLPKFEVTVNAPQTVTISDDEFQVDV CAKYNFGQPVQGETQIRVCREYFSSSNCEKNENEICEQFIAQLENGCVSQIVNTKVFO LYRSGLFMTFHVAVIVTESGTVMQISEKTSVFITQLLGTNVFENMDTFYRRGISYFGT LKFSDPNNVPMVNKLLQLELNDEFIGNYTTDENGAEQFSIDTSDIFDPEFNLKATYVR PESCYLPSWLTPOYLDAHFLVSRFYSTRNSFLKIVPEPKQLECNQKVVTVHYSLNSE AYEDDSNVKFFYLMMVKGAILLSGQKEIRNKAWNGNFSFPI SISADLAPAAVLFFVYTL HPSGEIVADSVRFQVDKCFKHVNIKFSNEQGLPGSNASLCLQAAPVLFCALRAVDRN VLLLKSEQQLSAESVYNMVPSTIEPYGYFYHGLNLDGKEDPCIPQRDMFYNGLYYTPV SNYGDGDIYNIVRNMGLKVFTNLHYRKPEKIMVQCVVFRLELHVASGIRGENADYVEQ AIIQTVRTNFPETWMWDLVSDSSGSANLSFLIPDTITQWEASGFCVNGDVGFGISST TTLEVSQPFIEIASPFSVVQNEQFDLIVNVFSYRNTCVEISVQVEESQNYEANIHTL KINGSEVIQAGGRKTNVWTIIPKKLGKVNITVVAESKQSSACPNEGMEQQKLNWKDTV VQSFLVEPEGIEKERTQSFLICTEGAKASKQGVLDLPNDVVEGSARGFFTVDGILGL ALQNLVVLQMPYGSGEQNAALLASDTYVLDYLKSTEQLTEEVQSKAFFLLSNGYQRQL SFKNSDGSYSVFWQQSQKGSIWLSALTFKTLERMKKYVFIDENVQKQTLIWLSSQQKT SGCFKNDGQLFNHAWEGGDEEDISLTAYVVGMMFFEAGLNFTFPALRNALFCLEAALDS GVTNGYNHAILAYAFALAGKEKQVESLLQTLQDQAPKLNNVIYWERERKPKTEEFPSF IPWAPSAQTEKSCYVLLAVISRKIPDLTYASKIVQWLAQRMNSHGGFSSNQTPDDTLF KLYTGQKESFRSSSVGYTLGKANEEKENRRNGGEGSSEIFQVNGHNRLLVQRSEVTQA PGEYTVDVEGHGCTFIQATLKYNVLLPKKASGFSLSLEIVKNYSSTAFDLTVTLKYTG IRNKSSMVVIDVKMLSGFTPTMSSIELENKGQVMKTEVKNDHVLFFYLENVGFGRADS FPFSVEQSNLVFNIQPAPAMVYDYEEKEEYALAFYNIDSSSVSE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 26B.

Table 26B. Comparison of NOV26a against NOV26b and NOV26c.		
Protein Sequence	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV26b	164..1411 150..1436	997/1311 (76%) 1072/1311 (81%)

Further analysis of the NOV26a protein yielded the following properties shown in Table 26C.

Table 26C. Protein Sequence Properties NOV26a	
PSort analysis:	0.8200 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1380 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Likely cleavage site between residues 46 and 47

A search of the NOV26a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 26D.

Table 26D. Geneseq Results for NOV26a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB50673	Human alpha-2 macroglobulin protein SEQ ID NO:59 - Homo sapiens, 1474 aa. [WO200073328-A2, 07-DEC-2000]	35..1407 30..1468	552/1458 (37%) 846/1458 (57%)	0.0
AAY97157	Human alpha-2-macroglobulin - Homo sapiens, 1474 aa. [WO200046246-A1, 10-AUG-2000]	35..1407 30..1468	551/1458 (37%) 846/1458 (57%)	0.0
AAR11334	Recombinant human alpha-2 macroglobulin - Homo sapiens, 1474 aa. [WO9103557-A, 21-MAR-1991]	35..1407 30..1468	549/1458 (37%) 844/1458 (57%)	0.0
AAR11749	Human alpha-2 macroglobulin bait region mutant - Homo sapiens, 1484 aa. [WO9103557-A, 21-MAR-1991]	35..1407 30..1478	546/1460 (37%) 842/1460 (57%)	0.0
AAB43949	Human cancer associated protein sequence SEQ ID NO:1394 - Homo sapiens, 1285 aa. [WO200055350-A1, 21-SEP-2000]	187..1407 2..1279	497/1295 (38%) 753/1295 (57%)	0.0

In a BLAST search of public sequence databases, the NOV26a protein was found to have homology to the proteins shown in the BLASTP data in Table 26E.

Table 26E. Public BLASTP Results for NOV26a				
Protein Accession Number	Protein/Organism/Length	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P20740	Ovostatin precursor (Ovomacroglobulin) - Gallus gallus (Chicken), 1473 aa.	1..1402 1..1461	640/1482 (43%) 931/1482 (62%)	0.0
P01023	Alpha-2-macroglobulin precursor (Alpha-2-M) - Homo sapiens (Human), 1474 aa.	35..1407 30..1468	552/1458 (37%) 846/1458 (57%)	0.0
CAA01532	ALPHA 2-MACROGLOBULIN	35..1407 30..1468	550/1458 (37%) 845/1458 (57%)	0.0

	1474 aa.			
P06238	Alpha-2-macroglobulin precursor (Alpha-2-M) - Rattus norvegicus (Rat), 1472 aa.	26..1408 13..1467	552/1477 (37%) 852/1477 (57%)	0.0
CAA01533	ALPHA 2-MACROGLOBULIN 690-740 - Homo sapiens (Human), 1484 aa.	35..1407 30..1478	547/1460 (37%) 844/1460 (57%)	0.0

PFam analysis predicts that the NOV26a protein contains the domains shown in the Table 26F.

Table 26F. Domain Analysis of NOV26a			
Pfam Domain	NOV26a Match Region	Identities/ Similarities for the Matched Region	Expect Value
A2M_N: domain 1 of 1	35..611	178/655 (27%) 381/655 (58%)	3.3e-96
A2M: domain 1 of 3	717..1096	137/414 (33%) 268/414 (65%)	2.2e-95
prenyltrans: domain 1 of 1	1194..1214	7/21 (33%) 15/21 (71%)	4.4
A2M: domain 2 of 3	1114..1218	45/110 (41%) 72/110 (65%)	1e-19
A2M: domain 3 of 3	1226..1402	61/242 (25%) 125/242 (52%)	1.1e-35

EXAMPLE 27.

The NOV27 clone was analyzed, and the nucleotide and predicted polypeptide
5 sequences are shown in Table 27A.

Table 27A. NOV27 Sequence Analysis		
	SEQ ID NO: 127	880 bp
NOV27a, CG59417-01 DNA Sequence	ACTCACTATAGGGCTCGAGCGGCACCATGGCTTTCCTCTGGCTCCTCTCCTGCTGGGC CCTCCTGGGTACCACCTTCGGCTGCGGGGTCCCCGCCATCCACCCTGTGTTCAGCGGC CTGTCCAGGATCGTGAATGGGGAGGACGCCGTCCCCGGCTCCTGGCCCTGGCAGGTGT CCCTGCAGGACAAAACCGGCTTCCACTTCTGCGGGGGCTCCCTCATCAGCGAGGACTG GGTGGTCACCGCTGCCCACTGCGGGGTGAGGACCTCCGACGTGGTCTGGCTGGGGAG TTTGACCAGGGCTCTGACGAGGAGAACATCCAGGTCCTGAAGATCGCCAAGGTCTTCA AGAACCCCAAGTTCAGCATTCTGACCGTGAACAATGACATCACCTGTGAAGCTGGC CACACCTGCCCCGCTTCTCCAGACAGTGTCCGCCGTGTGCCTGCCAGCGCCGACGAC GACTTCCCCGCGGGGACACTGTGTGCCACCACAGGCTGGGGCAAGACCAAGTACAACG CCAACAAGACCCCTGACAAGCTGCAGCAGGCAGCCCTGCCCTCCTGTCCAATGCCGA ATGCAAGAAGTCCTGGGGCAGGAGGATCACCGACGTGATGATCTGTGCCGGGGCCAGT GGCGTCTCCTCCTGCATGGGTGACTCTGGAGGCCCCCTGGTCTGCCAGAAGGACGGAG CCTGGACCCTGGTGGGCATTGTGTCTGGGGCAGCCGCACCTACTCTACCACCACGCC	

	CGCTGTGTACGCCCCGTGTCACCAAGCTCATACCCTGGGTGCAGAAGATCCTGGCCGCC AACTGAGCCCGCAGCTCCTGCCACCCCTGCCTTAAGATTCCCATTAAATGCATCTGT TTAGAAAAAA		
	ORF Start: ATG at 27	ORF Stop: TGA at 816	
	SEQ ID NO: 128	263 aa	MW at 28046.9kD
NOV27a, CG59417-01 Protein Sequence	MAFLWLLSCWALLGTTFGCGVPAIHPVFSGLSRIVNGEDAVPGSWPWQVSLQDKTGFH FCGGSLLISEDWVVTAHCGVRTSDVVVAGEFDQGSDEENIQVLKIAKVFKNPKFSILT VNNDITLLKLATPARFSQTVSAVCLPSADDDFPAGTLCATTGWGKTKYNANKTPDKLQ QAALPLLSNAECKKSWGRRITDVMICAGASGVSSCMGDSGGPLVCQKDGAWTLVGIVS WGSRTYSTTTPAVYARVTKLIPWVQKILAN		

Further analysis of the NOV27a protein yielded the following properties shown in Table 27B.

Table 27B. Protein Sequence Properties NOV27a	
PSort analysis:	0.3700 probability located in outside; 0.1040 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 19 and 20

A search of the NOV27a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
5 homologous proteins shown in Table 27C.

Table 27C. Geneseq Results for NOV27a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB98504	Human chymotrypsin serine protease catalytic domain - Homo sapiens, 231 aa. [WO200129056-A1, 26-APR-2001]	33..263 1..231	226/231 (97%) 228/231 (97%)	e-132
AA99596	Bovine chymotrypsinogen A - Bos taurus, 245 aa. [WO200032759-A1, 08-JUN-2000]	19..263 1..245	197/245 (80%) 213/245 (86%)	e-116
AAB11711	Mouse serine protease BSSP5 (mBSSP5) SEQ ID NO:4 - Mus sp, 264 aa. [WO200031243-A1, 02-JUN-2000]	1..263 1..264	150/264 (56%) 188/264 (70%)	5e-87
AAB11710	Human serine protease BSSP5 (hBSSP5) SEQ ID NO:2 - Homo sapiens, 264 aa. [WO200031243-A1, 02-JUN-2000]	1..263 1..264	141/264 (53%) 184/264 (69%)	2e-82

AAB54190	Human pancreatic cancer antigen protein sequence SEQ ID NO:642 - Homo sapiens, 133 aa. [WO200055320-A1, 21-SEP-2000]	132..263 2..133	127/132 (96%) 129/132 (97%)	1e-71
----------	---	--------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV27a protein was found to have homology to the proteins shown in the BLASTP data in Table 27D.

Table 27D. Public BLASTP Results for NOV27a				
Protein Accession Number	Protein/Organism/Length	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P17538	Chymotrypsinogen B precursor (EC 3.4.21.1) - Homo sapiens (Human), 263 aa.	1..263 1..263	257/263 (97%) 259/263 (97%)	e-152
P04813	Chymotrypsinogen 2 precursor (EC 3.4.21.1) - Canis familiaris (Dog), 263 aa.	1..263 1..263	228/263 (86%) 241/263 (90%)	e-135
Q9CR35	2200008D09RIK PROTEIN - Mus musculus (Mouse), 263 aa.	1..263 1..263	223/263 (84%) 246/263 (92%)	e-135
P07338	Chymotrypsinogen B precursor (EC 3.4.21.1) - Rattus norvegicus (Rat), 263 aa.	1..263 1..263	222/263 (84%) 244/263 (92%)	e-135
Q9DC86	2200008D09RIK PROTEIN - Mus musculus (Mouse), 263 aa.	1..263 1..263	222/263 (84%) 246/263 (93%)	e-134

PFam analysis predicts that the NOV27a protein contains the domains shown in the Table 27E.

Table 27E. Domain Analysis of NOV27a			
Pfam Domain	NOV27a Match Region	Identities/ Similarities for the Matched Region	Expect Value
trypsin: domain 1 of 1	34..256	109/261 (42%) 194/261 (74%)	5.6e-102

5

EXAMPLE 28.

The NOV28 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 28A.

Table 28A. NOV28 Sequence Analysis		
	SEQ ID NO: 129	1749 bp

NOV28a, CG59415-01 DNA Sequence	GCGGTCCCCAGCCTGGGTAAAGATGGCCCCATGGCCCCGAAGGGCCTAGTCCCAGCT GTGCTCTGGGGCCTCAGCCTCTTCTCAACCTCCCAGGACCTATCTGGCTCCAGCCCT CTCCACCTCCCCAGTCTTCTCCCCCGCCTCAGCCCCATCCGTGTCATACCTGCCGGGG ACTGGTTGACAGCTTTAACAAGGGCCTGGAGAGAACCATCCGGGACAACCTTTGGAGGT GGAAACACTGCCTGGGAGGAAGAGAATTTGTCAAATACAAAGACAGTGAGACCCGCC TGGTAGAGGTGCTGGAGGGTGTGTGCAGCAAGTCAGACTTCGAGTGCCACCGCCTGCT GGAGCTGAGTGAGGAGCTGGTGGAGAGCTGGTGGTTTTCACAAGCAGCAGGAGGCCCG GACCTCTTCCAGTGGCTGTGCTCAGATTCCCTGAAGCTCTGCTGCCCCGCAGGCACCT TCGGGCCCTCCTGCCTTCCCTGTCCTGGGGGAACAGAGAGGCCCTGCGGTGGCTACGG GCAGTGTGAAGGAGAAGGGACACGAGGGGGCAGCGGGCAGTGTGACTGCCAAGCCGGC TACGGGGGTGAGGCCTGTGGCCAGTGTGGCCTTGGCTACTTTGAGGCAGAACGCAACG CCAGCCATCTGGTATGTTTCGGCTTGTTTTGGCCCCCTGTGCCCCGATGCTCAGGACCTGA GGAATCAAACCTGTTTGAATGCAAGAAGGGCTGGGCCCTGCATCACCTCAAGTGTGTA GACATTGATGAGTGTGGCACAGAGGGAGCCAACTGTGGAGCTGACCAATTCTGCGTGA ACACTGAGGGCTCCTATGAGTGCCGAGACTGTGCCAAGGCCTGCCTAGGCTGCATGGG GGCAGGGCCAGGTCGCTGTAAGAAGTGTAGCCCTGGCTATCAGCAGGTGGGCTCCAAG TGTCTCGATGTGGATGAGTGTGAGACAGAGGTGTGTCCCGGGAGAGAACAAAGCCAGT GTGAAAACACCGAGGGCGGTTATCGCTGCATCTGTGCCGAGGGCTACAAGCAGATGGA AGGCATCTGTGTGAAGGAGCAGATCCCAGAGTCAGCAGGCTTCTTCTCAGAGATGACA GAAGACGAGTTGGTGGTGTGCTGCAGCAGATGTTCTTTGGCATCATCATCTGTGCACTGG CCACGCTGGCTGCTAAGGGGGACTTGGTGTTCACCGCCATCTTCATTGGGGCTGTGGC GGCCATGACTGGGTACTGGTTGTGAGAGCGCAGTGACCGTGTGCTGGAGGGCTTCATC AAGGGCAGATAATCGCGGCCACCACCTGTAGGACCTCCTCCCACCCACGCTGCCCCCA GAGCTTGGGCTGCCCTCCTGCTGGACACTCAGGACAGCTTGGTTTATTTTGGAGAGTG GGGTAAGCACCCCTACCTGCCTTACAGAGCAGCCAGGTACCCAGGCCCGGGCAGACA AGGCCCTGGGGTAAAAAGTAGCCCTGAAGGTGGATACCATGAGCTCTTCACCTGGCG GGGACTGGCAGGCTTACAATGTGTGAATTTCAAAGTTTTTCTTAATGGTGGCTGC TAGAGCTTTGGCCCCCTGCTTAGGATTAGGTGGTCTCAGAGGGGTGGGGCCATCACAG CTCCCTCCTGCCAGCTGCATGCTGCCAGTTCCTGTTCTGTGTTACACATCCCCACA CCCCATTGCCACTTATTTATTCATCTCAGGAAATAAAGAAAGGTCTTGAAAGTTAA AAAAAAA
	ORF Start: ATG at 23 ORF Stop: TAA at 1286
	SEQ ID NO: 130 421 aa MW at 45520.1kD
NOV28a, CG59415-01 Protein Sequence	MAPWPPKGLVPAVLWGLSLFLNLP GPIWLQSPPPQSSPPQPHPCHTCRGLVDSFNK GLERTIRDNFGGNTAWEEENLSKYKDSETRLVEVLEGVCSKSDFECHRLLELSEELV ESWWFHKQOEAPDLFQWLCSDSLKLCCPAGTFGPSCLPCPGGTERPCGGYGQCEGEGT RGGSGHCDQCAGYGGGEACGQCGLGYFEAERNASHLVCSACFGPCARCSGPEESNCLQC KKGWALHHLKCVDIDECGTEGANCGADQFCVNTEGSYECRDCAKACLGCMGAGPGRCK KCSPGYQQVGSKCLDVDECETEVC PGREQAQCENTEGGYRCICAEGYQMEGICVKEQ IPESAGFFSEMTEDELVVLQOMFFGIIICALATLAAKGLVFTAI FIGAVAAMTGYWL SERSDRVLEGFIKGR
	SEQ ID NO: 131 1011 bp
NOV28b, 191815704 DNA Sequence	GGATCCCAGCCCTCTCCACCTCCCCAGTCTTCTCCCCCGCCTCAGCCCCATCCGTGTC ATACCTGCCGGGGACTGGTTGACAGCTTTAACAAGGGCCTGGAGAGAACCATCCGGGA CAACTTTGGAGGTGGAAACACTGCCTGGGAGGAAGAGAATTTGTCAAATACAAAGAC AGTGAGACCCGCCTGGTAGAGGTGCTGGAGGGTGTGTGCAGCAAGTCAGACTTCGAGT GCCACCGCCTGCTGGAGCTGAGTGAGGAGCTGGTGGAGAGCTGGTGGTTTTCACAAGCA GCAGGAGGCCCCGGACCTCTTCCAGTGGCTGTGCTCAGATTCCCTGAAGCTCTGCTGC CCCGCAGGCACCTTCGGGCCCTCCTGCCTTCCCTGTCCTGGGGGAACAGAGAGGCCCT GCGGTGGCTACGGGCAGTGTGAAGGAGAAGGGACACGAGGGGGCAGCGGGCAGTGTGA CTGCCAAGCCGGCTACGGGGGTGAGGCCTGTGGCCAGTGTGGCCTTGGCTACTTTGAG GCAGAACGCAACGCCAGCCATCTGGTATGTTTCGGCTTGTTTTGGCCCCCTGTGCCCGAT GCTCAGGACCTGAGGAATCAAACCTGTTTGAATGCAAGAAGGGCTGGGCCCTGCATCA CCTCAAGTGTGTAGACATTGATGAGTGTGGCACAGAGGGAGCCAACTGTGGAGCTGAC CAATTCTGCGTGAACACTGAGGGCTCCTATGAGTGCCGAGACTGTGCCAAGGCCTGCC TAGGCTGCATGGGGGCAGGGCCAGGTCGCTGTAAGAAGTGTAGCCCTGGCTATCAGCA GGTGGGCTCCAAGTGTCTCGATGTGGATGAGTGTGAGACAGAGGTGTGTCCGGGAGAG AACAAAGCAGTGTGAAAACACCGAGGGCGGTTATCGCTGCATCTGTGCCGAGGGCTACA

	AGCAGATGGAAGGCATCTGTGTGAAGGAGCAGATCCCAGAGTCAGCAGGCTTCTTCTC AGAGATGACAGAAGACGAGCTCGAG		
	ORF Start: GGA at 1	ORF Stop:	
	SEQ ID NO: 132	337 aa	MW at 36352.1kD
NOV28b, 191815704 Protein Sequence	GSQPSPPPQSSPPPQPHPCHTCRGLVDSFNKGLERTIRDNFGGGNTAWEEENLSKYKD SETRLVEVLEGVCSKSDFECHRLLELSEELVESWWFHKQQEAPDLFQWLCSDSLKLCC PAGTFGPSCLPCPGGTERPCGGYGQCEGEGTRGSGHCDQAGYGGEACGQCGLGYFE AERNASHLVCSACFGPCARCSGPESNCLQCKKGWALHHLKCVDIDECGTEGANCGAD QFCVNTEGSYECRDCAKACLGCMGAGPGRCKKCSPGYQQVGSKCLDVDECETEVCPGE NKQCENTEGGYRCICAEGYKQMEGICVKEQIPESAGFFSEMTEDELE		
	SEQ ID NO: 133	1011 bp	
NOV28c, 191815724 DNA Sequence	GGATCCCAGCCCTCTCCACCTCCCAAGTCTTCTCCCCGCCTCAGCCCCATCCGTGTC ATACCTGCCGGGACTGGTTGACAGCTTTAACAAGGGCCTGGAGAGAACCATCCGGGA CAACTTTGAGAGTGGAACACTGCCTGGGAGGAAGAGAATTTGTCCAAATACAAAGAC AGTGAGACCCGCCTGGTAGAGGTGCTGGAGGGTGTGTGCAGCAAGTCAGACTTCGAGT GCCACCGCTGCTGGAGCTGAGTGAGGAGCTGGTGGAGAGCTGGTGGTTTCAACAAGCA GCAGGAGGCCCCGACCTCTTCCAGTGGCTGTGCTCAGATTCCCTGAAGCTCTGCTGC CCCGCAGGCACCTTCGGGCCCTCCTGCCTTCCCTGTCTTGGGGGAACAGAGAGGCCCT GCGGTGGCTGCGGGCAGTGTGAAGGAGAAGGGACACGAGGGGGCAGCGGGCACTGTGA CTGCCAAGCCGGCTACGGGGGTGAGGCCTGTGGCCAGTGTGGCCTTGGCTACTTTGAG GCAGAACGCAACGCCAGCCATCTGGTATGTTTCGGCTTGTTTTGGCCCCCTGTGCCCGAT GCTCAGGACCTGAGGAATCAAAGTGTGCAATGCAAGAAGGGCTGGGCCCTGCATCA CCTCAAGTGTGTAGACATTGATGAGTGTGGCACAGAGGGAGCCAACTGTGGAGCTGAC CAATTCTGCGTGAACACTGAGGGCTCCTATGAGTGCCGAGACTGTGCCAAGGCCTGCC TAGGCTGCATGGGGGCAGGGCCAGGTCGCTGTAAGAAGTGTAGCCCTGGCTATCAGCA GGTGGGCTCCAAGTGTCTCGATGTGGATGAGTGTGAGACAGAGGTGTGTCCGGGAGAG AACAAGCAGTGTGAAAACACCGAGGGCGGTTATCGCTGCATCTGTGCCGAGGGCTACA AGCAGATGGAAGGCATCTGTGTGAAGGAGCAGATCCCAGAGTCAGCAGGCTTCTTCTC AGAGATGACAGAAGACGAGCTCGAG		
	ORF Start: GGA at 1	ORF Stop:	
	SEQ ID NO: 134	337 aa	MW at 36292.1kD
NOV28c, 191815724 Protein Sequence	GSQPSPPPKSSPPPQPHPCHTCRGLVDSFNKGLERTIRDNFGGGNTAWEEENLSKYKD SETRLVEVLEGVCSKSDFECHRLLELSEELVESWWFHKQQEAPDLFQWLCSDSLKLCC PAGTFGPSCLPCPGGTERPCGGCGQCEGEGTRGSGHCDQAGYGGEACGQCGLGYFE AERNASHLVCSACFGPCARCSGPESNCLQCKKGWALHHLKCVDIDECGTEGANCGAD QFCVNTEGSYECRDCAKACLGCMGAGPGRCKKCSPGYQQVGSKCLDVDECETEVCPGE NKQCENTEGGYRCICAEGYKQMEGICVKEQIPESAGFFSEMTEDELE		
	SEQ ID NO: 135	1646 bp	
NOV28d, CG59415-02 DNA Sequence	GGCGACGCGGTCCCCAGCCTGGGTAAAGATGGCCCCATGGCCCCGAAGGGCCTAGTC CCAGCTGTGCTCTGGGGCCTCAGCCTCTTCTCAACCTCCCAGGACCTATCTGGCTCC AGCCCTCTCCACCTCCCCAGTCTTCTCCCCGCCTCAGCCCCATCCGTGTCATACCTG CCGGGGACTGGTTGACAGCTTTAACAAGGGCCTGGAGAGAACCATCCGGGACAACCTT GGAGGTGGAACACTGCCTGGGAGGAAGAGAATTTGTCCAAATACAAAGACAGTGAGA CCCGCCTGGTAGAGGTGCTGGAGGGTGTGTGCAGCAAGTCAGACTTCGAGTGCCACCG CCTGCTGGAGCTGAGTGAGGAGCTGGTGGAGAGCTGGTGGTTTCAACAAGCAGCAGGAG GCCCCGGACCTCTTCCAGTGGCTGTGCTCAGATTCCCTGAAGCTCTGCTGCCCCGCAG GCACCTTCGGGCCCTCCTGCCTTCCCTGTCTTGGGGGAACAGAGAGGCCCTGCGGTGG CTACGGGCAGTGTGAAGGAGAAGGGACACGAGGGGGCAGCGGGCACTGTGACTGCCAA GCCGGCTACGGGGGTGAGGCCTGTGGCCAGTGTGGCCTTGGCTACTTTGAGGCAGAAC GCAACGCCAGCCATCTGGTATGTTTCGGCTTGTTTTGGCCCCCTGTGCCCGATGCTCAGG ACCTGAGGAATCAAAGTGTGCAATGCAAGAAGGGCTGGGCCCTGCATCACCTCAAG TGTGTAGACTGTGCCAAGGCCTGCCTAGGCTGCATGGGGGCAGGGCCAGGTCGCTGTA AGAAGTGTAGCCCTGGCTATCAGCAGGTGGGCTCCAAGTGTCTCGTGAGTCTCCTGCT GATGGACACAGGCACCGGCTCACCCAGCATGAATGGTGAAGAGGCTGGAATATGGGCA GGTGGGGGAAGGAAGGGTGGAATGTTGCCTGGGCAGAGGGGAGGAGATGGACAAGATG		

	GAGTCAGGTGCTGGGTGGGGGGCCCTAGCAGGACTCTGACCCCTCCCTCCCCTCAAGA TGTGGATGAGTGTGAGACAGAGGTGTGTCCGGGAGAGAACAAGCAGTGTGAAAACACC GAGGGCGGTTATCGCTGCATCTGTGCCGAGGGCTACAAGCAGATGGAAGGCATCTGTG TGAACAGAAGACGAGTTGGTGGTGTGTCAGCAGATGTTCTTTGGCATCATCATCTGTG CACTGGCCACGCTGGCTGCTAAGGGCGACTTGGTGTTCACCGCCATCTTCATTGGGGC TGTGGCGGCCATGACTGGCTACTGGTTGTCAGAGCGCAGTGACCGTGTGCTGGAGGGC TTCATCAAGGGCAGATAATCGCGGCCACCACCTGTAGGACCTCCTCCCACCCACGCTG CCCCAGAGCTTGGGCTGCCCTCCTGCTGGACACTCAGGACAGCTTGGTTTATTTTGTG AGAGTGGGGTAAGCACCCCTACCTGCCTTACAGAGCAGCCAGGTACCCAGGCCCGGG CAGACAAGGCCCTGGGGTAAAAAGTAGCCCTGAAGGTGGATACCATGAGCTCTTCAC CTGGCGGGGACTGGCAGGCTTCACAATGTGTGAATTCAAAGTTTTTCCTTAATGGTG GCTGCTAGAGCTTTGGCCCCCTG		
	ORF Start: ATG at 29	ORF Stop: TAA at 1238	
	SEQ ID NO: 136	403 aa	MW at 42961.2kD
NOV28d, CG59415-02 Protein Sequence	MAPWPPKGLVPAVLWGLSLFLNLPGPPIWLQPSPPPQSSPPPQPHPCHTCRGLVDSFNK GLERTIRDNFGGGNTAWEEENLSKYKDSETRLVEVLEGVCSKSDFECHRLLLELSEELV ESWWFHKQQEAPDLFQWLCSDSLKLCCPAGTFGPSCLPCPGGTERPCGGYGQCEGEGT RGGSGHCDQCAGYGGEACGQCGLGYFEARNASHLVCSACFGPCARCSGPESNCLQC KKGWALHHLKCVDCAKACLGCMGAGPGRCKKCSPGYQQVGSKCLVSLLLMDTGTGSPS MNGEEAGIWAGGGRKGGMLPGQRGGDGDGVRWCWVGGPSRTLTPPSPQDVDECETEVC PGENKQCENTEGGYRCICAEGYKQMEGICVNRRRVGGAAADVLWHHHLCTGHAGC		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 28B.

Table 28B. Comparison of NOV28a against NOV28b through NOV28d.		
Protein Sequence	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV28b	44..364	307/321 (95%)
	17..336	307/321 (95%)
NOV28c	46..364	286/319 (89%)
	19..336	286/319 (89%)
NOV28d	1..348	270/384 (70%)
	1..381	275/384 (71%)

Further analysis of the NOV28a protein yielded the following properties shown in Table 28C.

Table 28C. Protein Sequence Properties NOV28a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 30 and 31

5 A search of the NOV28a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 28D.

Table 28D. Geneseq Results for NOV28a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU12316	Human PRO214 polypeptide sequence - Homo sapiens, 420 aa. [WO200140466-A2, 07-JUN-2001]	1..421 1..420	418/421 (99%) 418/421 (99%)	0.0
AAM41685	Human polypeptide SEQ ID NO 6616 - Homo sapiens, 513 aa. [WO200153312-A1, 26-JUL-2001]	1..421 94..513	418/421 (99%) 418/421 (99%)	0.0
AAM39899	Human polypeptide SEQ ID NO 3044 - Homo sapiens, 420 aa. [WO200153312-A1, 26-JUL-2001]	1..421 1..420	418/421 (99%) 418/421 (99%)	0.0
AAB68594	PRO214 - Homo sapiens, 420 aa. [WO200105836-A1, 25-JAN-2001]	1..421 1..420	418/421 (99%) 418/421 (99%)	0.0
AAB27228	Human EXMAD-6 SEQ ID NO: 6 - Homo sapiens, 420 aa. [WO200068380-A2, 16-NOV-2000]	1..421 1..420	418/421 (99%) 418/421 (99%)	0.0

In a BLAST search of public sequence databases, the NOV28a protein was found to have homology to the proteins shown in the BLASTP data in Table 28E.

Table 28E. Public BLASTP Results for NOV28a				
Protein Accession Number	Protein/Organism/Length	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9Y409	HYPOTHETICAL 44.9 KDA PROTEIN - Homo sapiens (Human), 417 aa.	1..418 1..417	413/418 (98%) 413/418 (98%)	0.0
Q91XD7	UNKNOWN (PROTEIN FOR MGC:18896) - Mus musculus (Mouse), 420 aa.	1..421 1..420	383/421 (90%) 405/421 (95%)	0.0
Q96HD1	UNKNOWN (PROTEIN FOR MGC:8447) - Homo sapiens (Human), 422 aa.	1..362 1..361	348/362 (96%) 353/362 (97%)	0.0
Q9CYA0	5730592L21RIK PROTEIN - Mus musculus (Mouse), 350 aa.	33..346 16..330	154/316 (48%) 200/316 (62%)	e-100
Q60438	HT PROTEIN - Cricetulus griseus (Chinese hamster), 348 aa.	9..339 3..324	156/333 (46%) 202/333 (59%)	4e-97

PFam analysis predicts that the NOV28a protein contains the domains shown in the Table 28F.

Table 28F. Domain Analysis of NOV28a			
Pfam Domain	NOV28a Match Region	Identities/ Similarities for the Matched Region	Expect Value
laminin_EGF: domain 1 of 1	168..211	11/60 (18%) 32/60 (53%)	0.11
zf-MYND: domain 1 of 1	218..243	10/43 (23%) 15/43 (35%)	4.7
PHD: domain 1 of 1	217..277	12/64 (19%) 38/64 (59%)	2.8
TIL: domain 1 of 1	249..309	17/79 (22%) 37/79 (47%)	8.1
Furin-like: domain 1 of 1	189..310	36/188 (19%) 76/188 (40%)	6.5
EB: domain 1 of 1	292..344	15/62 (24%) 35/62 (56%)	0.3

EXAMPLE 29.

The NOV29 clone was analyzed, and the nucleotide and predicted polypeptide
5 sequences are shown in Table 29A.

Table 29A. NOV29 Sequence Analysis		
	SEQ ID NO: 137	6997 bp
NOV29a, CG59297-01 DNA Sequence	ATGGGGTGGAGGGGGCTGATCGCAGCCTTGCCCCTGCTCTCCTTGGTGCAGCCTGCTC TGGGAACCAGCTCAAAGGATGAGGATGTAGGAAGAAGCTGGTCTGCTGACTGTCATAC TTGTGACCAGCTTGCACAGGACATGGCCGAGGAGGCAGCCAGAACATTTCTGATGAC CAGGAAAGGTGTCTCCAGGCTGCCTGCTGCCTTTCTTTGGTGGTGAGCTGTCTGTGA GCACTGACAAGAGCTGGGGTCTTCATCTGTGCAGCTGTAGCCCTCCTGGAGGTGGATT GTGGGTGAGGTCTATGCTAATCATGTGCTTCTTATGAGTGATGGGAAGTGTGGCTGT CCTTGGTGTGCTCTGAATGGAAAGGCAGAAGACCGGGAATCACAGAGCCCATCCTCAT CAGCTTCCAGGCAGAAGAACATTTGGAAAACAAGTAGTGAAGCAGCGTTAAGTGTGT TAATGAAAAAACACAGGCTGTTGTTAATGAAAAAACACAGGCGCCTCTGGATTGTGAT AACAGTGCTGATAGTCTCCGAGTCTTTGCTGACAGCAGTATTGGGGAGAATTGGACCC TTCAGATGGTTTGTGACCCAGACACTTGGATGCGTGGGCCAGCTCCCACGGCCTTCC GCCTGGCATTCTCGCACCCCCAGCTTCACGGCATCGCAGTCTGGTTCTGAGATCCTC TATCCCCCTACTCAGCATCCTCCTGTGGCCATCCTAGCTCGAAATTCTGATAACTTCA TGAACCCTGTTCTTAATTGCTCCCTGGAAGTGGAAGCTCGGGCACCTCCAAATCTGGG ATTCCGTGTTTCATATGGCTTCTGGAGAGGCTCTCTGTCTGATGATGGATTTGGGGAC AGTTCTGGGGTTGAAATGAGGCTACACAACATGTCTGAGGCAATGGCGGTGACTGCCT ACCACCAGTACTCAAAGAAGGAGTCTATATGCTCAAGGCTGTTATTTATAACGAGTT TCATGGAACCGAAGTGGAGCTTGGGCCTTATTATGTGGAGATTGGCCATGAGGCCGTG TCTGCGTTCATGAACTCCAGCAGTGTCCATGAAGATGAAGTGCTTGTCTTTGCTGACT CCCAAGTGAATCAGAAAACCGTCTCTGTCTACACAAATGGAAGTGTGTTGCCACAGA CACAGACATTACATTTACAGCTGTTACCAAGGAAACAATACCCCTGGAATTTGAGTGG	

TATTTTGGAGAGGATCCACCAGTGAGGACAACCTTCAAGAAGCATTAAAAAAGACTCA
 GCATCCCCCAATGGTATCGTGTGATGGTTAAGGCTTCCAACAGGATGAGCAGTGTGGT
 CTCTGAGCCCCATGTCATCAGGGTGCAGAAGAAAATTGTGGCCAATCGGCTCACGTCC
 CCCTCCTCAGCTCTGGTAAATGCCAGTGTGGCCTTTGAGTGCTGGATCAACTTCGGCA
 CAGATGTTGCCTACCTGTGGGACTTTGGGGATGGCACCGTCAGCCTGGGGAGCAGCTC
 CAGCAGCCATGTCTACAGTAGGGAAGGAGAATTTACAGTGGAGGTCCTTGCCTTCAAT
 AATGTCAGTGCCTCCACTCTAAGACAGCAACTTTTCATCGTGTGCGAGCCCTGCCAGC
 CACCCCTGGTGAAGAACATGGGGCCTGGGAAAGTCCAGATATGGAGGTCTCAGCCTGT
 GAGGCTGGGAGTGACGTTTGAAGCTGCAGTCTTCTGTGATATTTCCCAAGGTCTTTCT
 TACACCTGGAACCTTGATGGACTCTGAAGGGCTCCCTGTCTCCCTCCCTGCTGCTGTGG
 ACACTCACAGACAGACCCTCATCCTCCCGAGCCACACCTTGGAGTATGGGAACTACAC
 TGCCCTTGCCAAGGTTTCAATTGAAGGCAGTGTGGTGTACAGCAACTACTGTGTGGGC
 CTGGAGGTGCGAGCCCAGGCCCTGTGAGTGTGATCTCCGAGGGGCACACACCTATTCT
 TCTCCAGGACCACCTCATCCCCATTGTCTCAGAGGGACCCAGTCCTTCGACCCTGA
 CGACCCTGGGGCGACTCTCAGCCACTCCGATGACTTCTCCAACAGGTATCACTGGGAA
 TGCGCCACCGCTGGCTCCCCAGCACATCCCTGCTTCGACTCCTCCACTGCACACCAAC
 TGGATGCCGCGGCTCCCACTGTTTCTTTGAGGCACAATGGCTCAGTGACAGCTATGA
 TCAGTTCCTTGTGATGCTGAGGGTCTCCAGTGGTGGCCGGAACCTTCTGAGACCCGG
 GTGTTCTGTCCCCCTACCCTGACTCGGCGTTCAGATTCGTCCACATCTCCTGGGTCA
 GCTTTAAAGACACCTTCGTCAACTGGAATGACGAACTCTCTCTTCAAGCTATGTGTGA
 GGACTGCAGTGAAATACCGAATCTGTCTTATTCTGCTGGGATCTCTTTTTAGTCAATGCA
 ACAGAAAAGAATAGGATAGAAGTCCGTGTGAACACTTGTGAGGCACCAGCAGAAGAGG
 TGACACACTCAAGGGCTGCTTCTGACCTCAGTGTGATATGGAAGGCTGCGCCCAACAC
 CTGTGTAGGCAGTTTTATTGAGCTAAAGCCACAGTTCAGAAGGACCTGCGATGTGACA
 CACTGCCCAGAGGGCTGTGTCCGTACCACTCGCCTGTCTTTGGGCCAGCTGCACA
 AGTCATCACAGTTAAACCTGCTGCCCACTGAGCCTGGCACTGCAGATCCTGATGCAAC
 GACCACACCATTTCTACGGGAACCTTCACCCGTGACCCTTGGCCAACCTGCCACTTCA
 GCTCCAAGGGGAACCCCCACAGAGCCCATGACTGGAGTCTACTGGATTCTCCTGCGG
 GGGACTCTGCAGTCCTGGGGGAGGCTCCAGAGGAAGGTTCACTAGACCTAGAGCCAGG
 GCCACAGAGCAAGGGATCCCTGATGACTGGCCGCTCTGAGAGAAGTCAGCCCACCCAC
 AGCCCTGACCCTCACCTCTCTGCTAAGGACACCAGCTTTCCAGGATCAGGACCTAGCT
 TGAGTGCCGAGGAGAGCCCTGGAGATGGGGATAACCTGGTGGACCCCTCCCTGTCTGC
 AGGCAGAGCCGAGCCTGTCTCATGATTGACTGGCCCAAGGCCCTGCTGGGTGAGCA
 GTTTTCCAAGGCTATTATCCTCAGGTATTACAGAACAGACAGTGACAATCAAGCCAT
 ACTCTCTGAGCAGTGGAGAGACGTACGTCTGCAAGTGTCTGTGGCTTCAAGCATGG
 CTTACTGGGTAAAGCTCAGCTGTACTTGACAGTCAACCCGGCTCCTCGGGACATGGCC
 TGTGAGGTGCAGCCCCACCATGGTCTGGAAGCACACACCGTCTTCAGTGTCTTCTGCA
 TGTCTGGAAAACCGGACTTCCATTATGAATTTAGTTACCAGATAGGAAACACCTCCAA
 ACACACTTTGTACCATGGGAGAGACACCCAGTATTATTTTGTGTTGCCAGCTGGTGAG
 CACTTGGACAATTACAAAGTCATGGTTTCCACTGAAATCACAGATGGCAAAGGCTCCA
 AGGTCCAGCCGTGCACTGTGGTGGTGAAGTGTGCTGCCCCGCTACCATGGAAATGACTG
 TCTGGGCGAGGACCTGTATAATTCCAGCCTGAAAAACCTTTCTACCCTCCAGCTGATG
 GGGAGTTACACAGAAATCAGGAACTACATCACTGTGATCACCAGAATCCTGAGTCGTT
 TGTCTAAGGAGGACAAAACCTGCCTCCTGCAACCAATGGTCACGAATACAGGATGCATT
 AATTTCTTCAGTATGCAGATTGGCTTTTGTAGATCAGCTAGGCTTTATGAGTGCAGTT
 CTCATCCTCAAGTACACCCGGGCACTCCTTGCTCAAGGCCAGTTCCTCGGGGCCATTG
 TGATTGACAAAGGAGTGAGGCTTGAAGTCTCATCGGTCTCATATCCAGAGTCTGGGAAGT
 CTCTGAGCAAGAAAACCTCGAAGGAGGAAGTCTATCGACATGAAGAAGGAATTACAGTC
 ATCTCAGATTTATTGTTGATTGGTGGAGTTGTGGGCCTCAACCTCTATACCTGCTCCA
 GCAGAAGACCCATCAACAGGCAATGGCTAAGGAAACCCGTGATGGTTCGAGTTTGGGGA
 GGAGGATGGCCTGGATAATAGGAGAAATAAAACGACATTTGTATTACTTCGGGATAAA
 GTGAATCTCCATCAGTTCACTGAGCTTTCCGAAAACCCCAAGGAATCTCTACAGATAG
 AAATTGAATTTTCCAAACCTGTTACAAGGGCATTTCCTCGTCATGTTGCTAGTAAGATT
 CTCTGAGAAACCTACTCCCTCTGATTTTCTTGTGAAGCAGATCTACTTCTGGGATGAG
 TCAATTGTGCAGATTTATATACCTGCTGCTTCTCAGAAAGATGCCAGTGTAGGCTATT
 TATCCTTATTGGATGCTGACTATGACAGAAAACCTCCAAACAGATATTTAGCTAAGGC
 AGTGAACCTATACAGTACATTTCCAGTGGATCCGATGCCTGTTTTGGGACAAGAGAGAG
 TGGAATCTGAACGTTTCTCTCCACAACCAGGGACTTCTCCTGAAAAAGTGAAGTGA
 GCTACCATCGCCTCGCGGCATTCTGCTCTCCTAAGGAGAAAGCTGAAGGCCAGTTTTGA
 AGTGAGTGACATTTCCAAGCTACAGAGCCACCCAGAAAACCTTGCTTCCAGTATTTTT
 ATTATGGGTTCTGTGATTCTTTATGGATTTTTTGGTTCGCTAAAAGTAGACAAGTAGATC

	ATCATGAAAAAAGAAAGCTGGTTACATCTTTCTGCAAGAAGCTTCCCTGCCGGGCCA TCAGCTATATGCGGTCGTCATTGACACTGGCTTCCGAGCTCCGGCCAGCGCTCCTGCC CAACTGGGCCTGCTGAGGAAGATCCGCCTCTGGCACGACAGCCGTGGGCCTTCCCCAG GCTGGTTCATCAGCCACGTGATGGTGAAGGAGCTGCACACGGGACAGGGCTGGTTCTT CCCTGCCCAGTGCTGGCTGTCTGCCGGCAGGCATGATGGTCGCGTGGAGCGGGAGCTC ACCTGTCTGCAAGGGGGACTCGGCTTCCGGAAGCTTTTCTATTGCAAGTTCACAGAGT ACCTGGAGGATTTCCATGTCTGGCTGTGCGGTGTACAGCAGGCCCTCCTCCAGCCGCTA CCTGCACACGCCGCGCCTCACCGTGTCTTCTCCCTGCTGTGCGTCTACGCGTGTCTC ACTGCCCTGGTTGCTGCTGGAGGGCAAGAGCAGGTGAGAGCCATCGCTTTTCTTATA GCAGCTTCCAGATCCGACTACACTGTGGCCCCCTTTTGCCTAAGAAATCAACAAAGCT CACAGTTCTCCGAGAAAAGTTTAAACCAGGGGAAGCAAGCCTGGCTGCCTGGGGACCA GAGAAGGAACAGGAGGGCTCTGCCCCGGCTCAGCAAGGTACCTGCGACTTGTCTCATG GCCCTGTTCTCCTGAGCAGCCCCCTTATTGCTGGGGAACACGCTTGGCGAACCACCTC TTTCTTCTGTCAGGAAGCCCCGGGGTCTGCCCGAGTGGAGCCACACAGCCCCTTAGA GGAGGAGCACAGACCGAGGCACCCCATGGTGGGTGAGAAAGAGGGGTCTCAGCAGAG GCCTGAAACAGGAAGGAAGTGAAGCCCAGAAGAATTGAGAAAGCCCTGTGTGTCTACT CAGTAAATACCGGCAGGACCGTGGGAGAGACACTGTGGAGCAGCAAGGCTCGGGCACC CAGCAGTGGTTTGGAGGGACTAATGCCCCAGTGGTCAAGGGCCCTTCAGCCTTGGTGG AGCTCTGCAGTGTGGGCCATTTGTGGGACCGCTTCTTGGCCTGCAGTTTGGGGACAG GATTTCTAGCCTACAGGTATGCCTCATGGCCTTGGGTTTTGCTTGGAAAAGAAGAGCT GACAACCACTTTTTTACTGAGTCTTTATGTGAGGCTACCAGGGATCTGGACTCTGAAT TGGCAGAACGTTCTGGACTCGCCTCCCCTTCTCTTCAAGCTGCAGTATTCCTGACTG TGCAGGCGAGGTTGAAAAAGTCTTGGCTGCCCGACAACAAGCTCGCCACCTGCGCTGG GCGCATCCACCATCAAGGCCAGCTGAGGGGCACCAGACAGAGGATGAGGAGAGAGA GTCGCACACGGGCTGCCCTGAGAGACATTTCCATGGACATCCTCATGCTGCTTCTGCT TTTGTGTGTAATATATGGGAGATTTTCCCAAGATGAATACTCCCTCAATCAAGCTATC CGGAAAGAATTTACAAGAAATGCCAGAACTGCTTGGGTGGCCTGAGAAACATCGCTG ACTGGTGGGACTGGAGTCTGACCACACTTCTGGATGGCCTGTACCCGGGAGGCACCCC GTCAGCCCGTGTGCCGGGGGCTCAGCCTGGAGCTCTTGGAGGAAAATGCTACCTAATA GGCAGTTCCGTAATTAGGCAGCTAAAAGTTTTTCTAGGCATTTATGCAAGCCTCCCA GGCCATTTTCAGCACTCATCGAAGACTCTATTCTACATGTAGTCCCGAAGTTGGAGG CCCTGAGAACCCTACCTGATAGACCCAGAGAACCAGAACGTCAGCCCTGAATGGTCCT GGGGGCTGTGGGACAAGGGAGGACTGTGTGCTCAGCCTGGGCAGAACAGGACTGAAG CCCACACAGCCCTGTCCCGACTCAGGGCCAGCATGTGGATTGACCGCAGCACCAGGGC TGTGTCTGTGCACTTCACTCTCTATAACCCTCCAACCCAACTCTTACCAGCGTGTCC CTGAGAGTGGAGATCCTCCCTACGGGGAGTCTCGTCCCCTCATCCCTGGTGGAGTCAT TCAGCATCTTCCGCAGCGACTCAGCCCTGCAGTACCACCTCATGCTTCCCCAGGTGAG CTGACCTGCCTCTTGGGCCTCCTGGAGGTGCACAGGAAGATGGGGCTTACCTGGGCT GGGCTTCTCCACCAGACAGGACTAGTTCCCTACCCAT		
	ORF Start: ATG at 1	ORF Stop: TGA at 6904	
	SEQ ID NO: 138	2301 aa	MW at 254558.5kD
NOV29a, CG59297-01 Protein Sequence	MGWRGLIAALPLLSLVQPALGTSSKDEDVGRSWSADCHTCDQLAQDMAEEAAQNI SDD QERCLQAACCLSFGGELSVSTDKSWGLHLCS CSPGGGLWVEVYANHVLLMSDGKCGC PWCALNGKAEDRESQSPSSASRQKNIWKT TSEAALSVVNEKTQAVVNEKTQAPLDCD NSADSLRVFADSSIGENWTLQMVCDPDTWMRGPSSHGLPPGI PRTPSFTASQSGSEIL YPPTQHPPVAILARNSDNFMNPVLNCSLEVEARAPPNLGFRVHMASGEALCLMDFGD SSGVEMRLHNMSEAMAVTAYHQYSKEGVYMLKAVIYNEFHGTEVELGPYYVEIGHEAV SAFMNSSSVHEDEVLFVADSQVNQKTVSVYTNGTVFATDITFTAVTKETI PLEFEW YFGEDPPVRTTSRSIKKRLSIPQWYRVMVKASNRMSVSEPHVIRVQKKIVANRLTS PSSALVNASVAFECWINFGTDVAYLWDFGDGTVSLGSSSSSHVYSREGEFTVEVLAFN NVSASTLRQQLFIVCEPCQPPLVKNMGPGKVQI WRSQPVRLGVTFEAAVFCDISQGLS YTNWLM DSEGLPVSLPAVDTHRQTLILPSHTLEYGNYTALAKVQIEGSVVSNYCVG LEVRAQAPVSVISEGTHLFFSRTTSSPIVLRGTQSFDPDDPGATLSHSDDFS NRYHWE CATAGSPAHPCFDSSTAHLDAAPT VSF EAQWLSDSYDQFLVMLRVSSGGRNSSETR VFLSPYPDSAFRFVHISWVSFKDTFVNWNDELSLQAMCEDCSEIPNLSYSWDLFLVNA TEKNRIEVRVNTCEAPAEVTHSRAASDLSVIWKAAPNTCVGSFIELKPQFRRTCDVT HCPEGCVRRHLACLLGQLHKSSQLNLLPTEPGTADPDATTT PFSREPSVTLGQPATS APRGTPTEPMTGVYWI PPAGDSAVLGEAPEEGSLDLEPGPQSKGSLMTGRSERSQPTH SPDPHLSAKDTSFPGSGPSLSAEESPGDGDNLVDP SLSAGRAEPVLMIDWPKALLGRA		

	VFQGYSSSGITEQTVTIKPYSLSSGETYVLQVSVASKHGLLGKAQLYLTVPAPRDMA CQVQPHHGLEAHTVFSVFCMSGKPDFHYEFSYQIGNTSKHTLYHGRDTQYYFVLPAGE HLDNYKVMVSTEITDGKGSKVQPCTVVVTVLPRYHGNDCLGEDLYNSSLKNLSTLQLM GSYTEIRNYITVITRILSRLSKEDKTASCNQWSRIQDALISSVCRLAFVDQLGFMSAV LILKYTRALLAQGGFSGPFVIDKGVRLELIGLISRVEVSEQENSKEEVYRHEEGITV ISDLLLIGGVVGLNLYTCSSRRPINRQWLRKPVMMVEFGEDGLDNRRNKTTFVLLRDK VNLHQFTELSNPQESLQIEIEFSKPVTRAFVMLLVRFSEKPTPSDFLVKQIYFWDE SIVQIYIPAASQKDASVGYLSLLDADYDRKPPNRYLAKAVNYTVHFQWIRCLFWDKRE WKSERFSPQPGTSPEKVNC SYHRLAAAFALLRRKLKASFEVSDISKLQSHPENLLPSIF IMGSVILYGFLVAKSRQVDHHEKKKAGYIFLQEASLPGHQLYAVVIDTGFRAPASAPA QLGLLRKIRLWHDSRGPSPGWFISHVMVKELHTGQGWFPAQCWLSAGRHDGRVEREL TCLQGGGLGFRKLFYCKFTEYLEDFHVWLSVYSRPSRSSRYLHTPRLTVSFSLLCVYACL TALVAAGGQEQVRAIAFPYSSFQIRLHCGPFLPKSTKLTVLREKFKPGEASLAAGWP EKEQEGSARLSKVPATCPHGPVLLSSPFIAGEHAWRTTSFLLQEAPGSARVEPHSPLR GGAQTEAPHGGSEERRGLSRGLKQEGSEAQKNSESPVCLLSKYRQDRGRDTVEQQSGT QQWFGGTNAPVVKGPSALVELCSVGHLWDRFFGLQFGDRISSLQVCLMALGFAWKRA DNHFFTESLCEATRDLDELAERSWTRLPFSSSCSIPDCAGEVEKVLAARQQARHLRW AHPPSKAQLRGTRQRMRESRTRAALRDISMDILMLLLLLLCVIYGRFSQDEYSLNQAI RKEFTRNARNCLGGLRNIADWDWSLTLLDGLYPGGTPSARVPGAQPGALGGKCYLI GSSVIRQLKVFPRHLCKPPRPFSAIEDSIPTCSPEVGGPENPYLIDPENQNVTLNGP GGCGTREDCVLSLGRTRTEAHTALSRLRASMWIDRSTRAVSVHFTLYNPPTQLFTSVS LRVEILPTGSLVPSSLVESFSIFRSDSALQYHMLLPQVS
--	--

Further analysis of the NOV29a protein yielded the following properties shown in Table 29B.

Table 29B. Protein Sequence Properties NOV29a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 22 and 23

A search of the NOV29a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several

5 homologous proteins shown in Table 29C.

Table 29C. Geneseq Results for NOV29a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU14647	Novel bone marrow polypeptide #46 - Homo sapiens, 488 aa. [WO200157187-A2, 09-AUG-2001]	196..683 1..488	487/488 (99%) 487/488 (99%)	0.0
AAU29269	Human PRO polypeptide sequence #246 - Homo sapiens, 300 aa. [WO200168848-A2, 20-SEP-2001]	2046..2300 1..255	254/255 (99%) 255/255 (99%)	e-147
AAE03429				e-139

	protein HETDB76, SEQ ID NO: 112 - Homo sapiens, 561 aa. [WO200132675-A1, 10-MAY-2001]	1..240	240/240 (99%)	
AAU14741	Novel bone marrow polypeptide #140 - Homo sapiens, 142 aa. [WO200157187-A2, 09-AUG-2001]	478..618 2..142	140/141 (99%) 140/141 (99%)	5e-79
AAB41274	Human ORFX ORF1038 polypeptide sequence SEQ ID NO:2076 - Homo sapiens, 160 aa. [WO200058473-A2, 05-OCT-2000]	1621..1738 43..160	116/118 (98%) 116/118 (98%)	5e-66

In a BLAST search of public sequence databases, the NOV29a protein was found to have homology to the proteins shown in the BLASTP data in Table 29D.

Table 29D. Public BLASTP Results for NOV29a				
Protein Accession Number	Protein/Organism/Length	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96Q08	KIAA1879 PROTEIN - Homo sapiens (Human), 995 aa (fragment).	1449..1734 212..538	77/332 (23%) 136/332 (40%)	5e-17
CAB59175	SEQUENCE 3 FROM PATENT WO9518225 - Homo sapiens (Human), 1614 aa (fragment).	1621..1737 403..523	45/121 (37%) 67/121 (55%)	2e-16
CAB59174	SEQUENCE 1 FROM PATENT WO9518225 - Homo sapiens (Human), 4339 aa (fragment).	1621..1737 3128..3248	45/121 (37%) 67/121 (55%)	2e-16
O42181	PKD1 PROTEIN - Fugu rubripes (Japanese pufferfish) (Takifugu rubripes), 4578 aa.	308..795 2000..2473	120/517 (23%) 191/517 (36%)	2e-16
Q15141	POLYCYSTIC KIDNEY DISEASE 1 PROTEIN - Homo sapiens (Human), 4292 aa.	1621..1737 3161..3281	45/121 (37%) 67/121 (55%)	2e-16

PFam analysis predicts that the NOV29a protein contains the domains shown in the Table 29E.

Table 29E. Domain Analysis of NOV29a			
Pfam Domain	NOV29a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PKD: domain 1 of 2	371..454		0.14

		50/94 (53%)	
PKD: domain 2 of 2	456..536	24/93 (26%) 61/93 (66%)	3.7e-09
REJ: domain 1 of 1	592..710	39/144 (27%) 74/144 (51%)	0.0013
hormone3: domain 1 of 1	1221..1233	6/13 (46%) 13/13 (100%)	7.8
GPS: domain 1 of 1	1497..1544	13/54 (24%) 31/54 (57%)	0.18
PLAT: domain 1 of 1	1623..1684	15/69 (22%) 46/69 (67%)	2e-05

EXAMPLE 30.

The NOV30 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 30A.

Table 30A. NOV30 Sequence Analysis		
	SEQ ID NO: 139	3095 bp
NOV30a, CG59264-01 DNA Sequence	CTGGCCAGACCCTGCCTCCAGCCACCGAGGCACATTACCTGGGCCCAACAGATGTCCT GGCAGCGGGGCCACAGTATCCTCCTCATGCGTCCACTCTCTTCAAGCCCAGTGCAAAG AAAGCAGCGAATCCCAGCACCTGAAAAATACCCAGGAGGTCCTCGAACCCACTCTGGT AACTCTCAACCCCTCATCTGCGAGCACGTGGGGCTTGGGGCCACGTTTGTATGTTGGG GAGGGCCCTGTGTTTTGGGGGAATGCAGCTCCCTTCTGCAGAGATGGGAGGTGGTCAT GAAGCTGGACCGCAAGGCCTTGGCTCGCGATCCGCCCGCTTCAGCCTCCCAAAGTACT GGGATTACAAGAGTGAGCCACTGCGCCAGGCCTCAAACATCAATGTTGATACCTGTTT TCAAAGTACTGGAAGAAGGTAGGGGTAAGGACAAGGAACCGGGAGGAGTGAGGGCGT CACTGGGTTTCGGCGTCTGGCAAGCGGTTCACTGTCTGCTCCCTAGCAGCCGGCCTT CGGGTCGGGCGTCTCCGCCGGCTACTGCCGCTTCAGTTCTCCCGGTGTGGCCACGAGT CGGGTGAGTCTCGGTTCTGAAGACACTCAGCCGGGGATGCCAGAGCCTGTGGGCAGCCG TATCGAACACGCAGGGTATCGTGGCAATCCAGAACGCTTTTCTGAATGGGATAGTTTG AAAGAAGGGCAGGCGTCTTTGTGGCACGGTAGGAACTGGGATGCATTTCCGCCGCCA GAGAGAAATACATAAAATCTATTGAGCGAGCGCTTGTTAATTATATGTGCCTTCTGCT TATTATATGCCTACAGGTCACAGGAACCTCAAGTATTTACAGAATGATGCCAAGGAG CGCAGTTTCTACACCATCCCTCCCGATCAGACCACATCTATTCCAATCTCAGTCTCAA ACTCTTCCCCTTTTTCAACTCACATTAATGAATTCATAAGTACCTAGCACTGAGAAT AAAAAGCGAAAGCAGCATCCGTCTTCTCTACTCTCACACAGCTTACAGTCCACAGCAA AGCGCAAGTCATCAAACACACCTGGTTGGCAGTGCCAGATTCTGTCAGTGAGGGTCC AGGAAAGCTCTTGCCCTCTTGCCAGCAGCCGCAGTATCTCAACGGATGCCGTGCACC ATATTCCTGGATGCTGAAGACATGGCAGACTATGGGTGGCAGTACCAGAGCCAGGAC CAACGTCAAGGGTATCCCATCTGGGGCAAACCTCACTGTGTACCGGGGAGGAGGCTACG TGGTCCCCTTGTCAGGACTAGGCAACAAGAGCGAAACTCTGTCCCTGGCAAAAAAAA GAACACCTGGCTGGACGCCCTGACCAGAGCTGTGTTTGTGGAGTCCACTGTCTACAAC GCCAACGTCAACCTGTTCTGCATTGTACGCTGACGCTAGAGACCAGCGCTCTGGGTG GGTATTTTGAATTTCTTTCAAGAAATTCATAAATTTCTATCTAACTTTGGGGTCATT CGTGGTAGCGGCAGAGCTCATCTACTTCTCTCTCTCTCTACTACATTGTGGTGCAA GTGCTTGAATCCAGGAGGCACAGGTTGCACTATTTCTGCAGCAAGTGGAACCTTCTGG AGCTGGCCATCATCTGGCCAGCTGGAGCGCCCTGGCGGTGTTTGTGAAGAGGGCTGT CCTGGCCGAAAGGGACCTCCAGATCATTGAGACTGAGGGCGCTCTACCGAACTTCCAA GCTGTTCAAGGATCAACTATACAAATGAACAAATTATCCGCCTTCCTGGTACTCCTGT CCACAGTGAAGCTTTGGCATCTGCTCAGGTTGAATCCCAAATGAACATGATCACGGC AGCCCTACGCCGTGCCTGGGGCGACATTTCAAGGCTTTATGATTGTCATCCTTACCATG CTCCTGGCTTACTCCATCGCGGTAAGTATCTGCTTTGGGTGGAACTCCGTTCTTACA	

	AAACCCTCTTTGATGCGGCGGAGACGATGGTCAGCCTTCAGCTGGGAATCTTCAACTA CGAGGAGGTCCTGGACTATAGCCCAGTGCTTGGCTCCTTCCTCATTGGATCCCCACTG CACCTGGCCACATTTCTGTTTTTTTTTTTTTTTTTTTTTTTGGAGATGGAGTATGGCTA TTGCATCACACAGACAACCTTCATTAAAGGAAGACACCAAGACAGGAGCTGCTCAGGG GCCACTGGGCACTGCGGTTAGAAAGGGAGCGAGACGCTCTCTCAAAAAAAAAAAGAAAG AAAAAAAAATATTCTGGAAGAGGCAGGAGAATCACTTGAATCCGGGAGAGGGAGGTTGC AGTTCGAAAGGAAGCGAGAGGGAGCGAAAGGCAGAGGCACTATGTGCCGGGGCGTCAG TCTGCTTGTCAAGAAAAAATGATGGATGGGAAGAAAAAGATCACACATCAACATCAA CATATGAAAAGCATGAGAGACAATTTAAAAAAGAGGACCTCCCCTCACTGCATTTTC CCTCACCACAACCCCTCACTCCCAAGACTTTCCCAAAGATCAGGAACTAAGCCTGA GAGAAGCCAATGTGAGGACAACGTGGGTACGGGAGCCGGGAGCGGGCACTTGAAGCG GGTAGAGCCACAGGAGCCGGTTTGAAGTTTGTCTGTAAGGGTCATGAGAGGCGACCGAA GGATTTTAAGTGTAGGGGAGAAGAAGTTGAGGAGCCGGGATATCATTGGAAGGATCTT GGAGTCAGTGGCAGGCAAACAAAATTTAGGTAAGGACTCACAGGAAGAGGTGAGGCAG ATGGAGGAATTATCCAAGAGTGAAGTGCACAGATGGGAAACAGCCACAGGAACACCG TTGTGACTGTAGCATGCGGTACAAAGGGCCCTGAGTGCCAGCCTAAGCAACAGAGCAA GACTCAGTCTCAAAAAAAAAACAAACAAAAAATCCCTGGGCGTGGTGGCTCATGCCTG TAATCTCAACACTTTGGGAGAAAAATATATATATTTTTCCCCTTAAATTATCATGTTG CAGGCCGGGCACAGTGGCTCATGCCTGCAATCCCAGCACTTTGGGAGGCCAAGGCAGG CGGATCACCTGACGTAAGGAG		
	ORF Start: ATG at 52	ORF Stop: TAA at 2959	
	SEQ ID NO: 140	969 aa	MW at 108791.3kD
NOV30a, CG59264-01 Protein Sequence	MSWQRGHSILLMRPLSSSPVQRKQRI PAPEKYPPGGPRTHSGNSQPLICEHVGLGATFV CWGGPCVLGECSSLLQRWEVVMKLDKALARDPPASASQSTGITRVSHCARPQTSMLI PVFKVLEEGRGKDKPEPGGVEGVTGFRRLASGSAVCSLAAGLRVGRLLRLLPLQFSRCG HESGESRFEDTQPGMPEPVGSRIEHAGYRGNPERFSEWDSLKEGQASLWHGRNWDALS PPREKYIKSIALRVNYMCLLLIICLQVTGTSSISQNDAKERSFYTIIPDQTTSSIPIS VSNSSPFSTHINEFTKYLALRIKSESSIRLLYSHTAYSPPQQSASHQTHLVGSAQIRQV RVQESSCPLAQPPQYLNGCRAPYSLDAEDMADYGWQYQSQDQRQGYPIWGKLTVYRGG GYVVPLSRTRQQERNSVPGKKKNTWLDALTRAVFVESTVYNANVNLFCIVTLTLETS LGGYFEFLFKKFIFYLTLSFVVAELIYFLFLYYIVVQVLESRRHRLHYFCSKWN LLELAILASWSALAVFVKRAVLAERDLQIIETEGALPNFQAVQGSTIQMNKLSAFLV LLSTVKLWHLRLNPKMNMITAALRRAWGDISGFMIVILTMLLAYSIAVSICFGWKL SYKTLFDAAETMVSLLQGIIFYEEVLDYSPVLGSLIGSPLHLATFLFFFFFFFLRWS MAIASHRQPSLKEDTKTGAAQGPLGTAVRKGARRSLKKRKKKNILEEAGESLESGRG RLQFERKREGAKGRGTMCRGVSLLVKKKMDGKKKI THQHQMKSMDNLKKEDLPSL HFPSQPPLTPKTFPKDQETKPERSQCEDNVGHGSRERALEAGRATGAGLSFVVRVMRG DRRILSVGEKKLRSRDIIGRILESVAGKQNLGKDSQEEVRQMEELSKSEVHRWETAHR NTVVTVACGTKGPECQPKQSKTQSQKKQTKKIPGRGGSCL		

Further analysis of the NOV30a protein yielded the following properties shown in Table 30B.

Table 30B. Protein Sequence Properties NOV30a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3869 probability located in mitochondrial inner membrane; 0.3000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV30a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several

5 homologous proteins shown in Table 30C.

Table 30C. Geneseq Results for NOV30a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB68450	Amino acid sequence of a human PKD2 polypeptide - Homo sapiens, 968 aa. [US6228591-B1, 08-MAY-2001]	15..967 2..961	322/1001 (32%) 493/1001 (49%)	e-125
AAY78946	Polycystic kidney disease PKD2 amino acid sequence - Homo sapiens, 968 aa. [US6031088-A, 29-FEB-2000]	15..967 2..961	322/1001 (32%) 493/1001 (49%)	e-125
AAM51861	Murine polycystic kidney disease protein 2 - Mus musculus, 966 aa. [WO200177331-A1, 18-OCT-2001]	63..967 39..959	302/960 (31%) 467/960 (48%)	e-116
AAB68448	Amino acid sequence of an internal fragment of human PKD2 - Homo sapiens, 866 aa. [US6228591-B1, 08-MAY-2001]	423..967 295..859	188/580 (32%) 303/580 (51%)	2e-76
AAY70245	Human Polycystin-L protein - Homo sapiens, 805 aa. [WO200012046-A2, 09-MAR-2000]	217..807 75..668	164/616 (26%) 289/616 (46%)	2e-60

In a BLAST search of public sequence databases, the NOV30a protein was found to have homology to the proteins shown in the BLASTP data in Table 30D.

Table 30D. Public BLASTP Results for NOV30a				
Protein Accession Number	Protein/Organism/Length	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q13563	Polycystin 2 (Autosomal dominant polycystic kidney disease type II protein) (Polycystin) (R48321) - Homo sapiens (Human), 968 aa.	15..967 2..961	320/1001 (31%) 490/1001 (47%)	e-123
O35245	Polycystin 2 - Mus musculus (Mouse), 966 aa.	63..925 39..916	295/917 (32%) 454/917 (49%)	e-114
G02640	polycystic kidney disease protein 2 - human, 608 aa (fragment).	383..967 6..601	202/611 (33%) 321/611 (52%)	6e-92
Q9UP35	POLYCYSTIN-L - Homo sapiens (Human), 805 aa.	217..807 75..668	165/616 (26%) 290/616 (46%)	3e-60
Q9P0L9				3e-60

	2-LIKE PROTEIN - Homo sapiens (Human), 805 aa.	75..668	290/616 (46%)	
--	---	---------	---------------	--

PFam analysis predicts that the NOV30a protein contains the domains shown in the Table 30E.

Table 30E. Domain Analysis of NOV30a			
Pfam Domain	NOV30a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ion_trans: domain 1 of 1	489..688	41/233 (18%) 139/233 (60%)	2.4e-06

EXAMPLE 31.

5 The NOV31 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 31A.

Table 31A. NOV31 Sequence Analysis		
	SEQ ID NO: 141	2316 bp
NOV31a, CG59623-01 DNA Sequence	CCTGAGCCTCATTGGGGGGTCTCCCCCACGGGCCGGGCATGCTGCCCCCGGAAG GAACCCCTCTCCTCGCTCCCCCAGCGTCCACGCGGAGCATGAACATTGAGGATGGCG CGTGCCCGCGGCTCCCCGTGCCCCCGCTGCCGCCCGGTAGGATGTCTTGCCCCACG GGGCATTGCTCTTCTCTGGCTCTTCTCCCCACCCCTGGGGGCCGGTGGAGGTGGAGT GGCCGTGACGTCTGCCGCCGGAGGGGGCTCCCCGCCGGCCACCTCCTGCCCCGTGGCC TGCTCCTGCAGCAACCAGGCCAGCCGGGTGATCTGCACACGGAGAGACCTGGCCGAGG TCCCAGCCAGCATCCCGGTCAACACGCGGTACCTGAACCTGCAAGAGAACGGCATCCA GGTGATCCGGACGGACACGTTCAAGCACCTGCGGCACCTGGAGATTCTGCAGCTGAGC AAGAACCTGGTGCGCAAGATCGAGGTGGGCGCCTTCAACGGGCTGCCAGCCTCAACA CGCTGGAGCTTTTTGACAACCGGCTGACCACGGTGCCACGCAGGCCTTCGAGTACCT GTCCAAGCTGCGGGAGCTCTGGCTGCGGAACAACCCATCGAGAGCATCCCCTCCTAC GCCTTCAACCGCGTGCCCTCGCTGCGGCGCCTGGACCTGGGCGAGCTCAAGCGGCTGG AATACATCTCGGAGGCGGCCTTCGAGGGGCTGGTCAACCTGCGCTACCTCAACCTGGG CATGTGCAACCTCAAGGACATCCCCAACCTGACGGCCCTGGTGCGCCTGGAGGAGCTG GAGCTGTCGGGCAACCGGCTGGACCTGATCCGCCCGGGCTCCTTCCAGGGTCTCACCA GCCTGCGCAAGCTGTGGCTCATGCACGCCCAGGTAGCCACCATCGAGCGCAACGCCTT CGACGACCTCAAGTCGCTGGAGGAGCTCAACCTGTCCCACAACAACCTGATGTGCTG CCCCACGACCTCTTCACGCCCCCTGCACCGCCTCGAGCGCGTGACCTCAACCACAACC CCTGGCATTGCAACTGCGACGTGCTCTGGCTGAGCTGGTGGCTCAAGGAGACGGTGCC CAGCAACACGACGTGCTGCGCCCGCTGTATGCGCCCGCCGGCCTCAAGGGGCGCTAC ATTGGGGAGCTGGACCAGTCGCATTTACCTGCTATGCGCCCGTCATCGTGAGCCGC CCACGGACCTCAACGTCACCGAGGGCATGGCTGCCGAGCTCAAATGCCGCACGGGCAC CTCCATGACCTCCGTCAACTGGCTGACGCCAACGGCACCCCTCATGACCCACGGCTCC TACCGCGTGCGCATCTCCGTCTGCATGACGGCACGCTTAACCTTACCAACGTCACCG TGCAGGACACGGGCCAGTACACGTGCATGGTGACGAACTCAGCCGGCAACACCACCGC CTCGGCCACGCTCAACGTCTCGGCCGTGGACCCCGTGGCGGCCGGGGGCACCGGCAGC GGCGGGGGCGGCCCTGGGGGCAGTGGTGGTGTGGAGGGGGCAGTGGCGGCTACACCT ACTTCACCACGGTGACCGTGGAGACCTGGAGACGCAGCCCGGAGAGGAGGCCCTGCA GCCGCGGGGGACGGAGAAGGAACCGCCAGGGGCCACGACAGACGGTGTCTGGGGTGGG GGCCGGCCTGGGGACGCGGCCGGCCCTGCCTCGTCTTCTACCACGGCACCCGCCCGC GCTCCTCGCGGCCACGGAGAAGGCGTTACGGTGCCCATCACGGATGTGACGGAGAA CGCCCTCAAGGACCTGGACGACGTATGAAGACCACCAAAATCATCATCGGCTGCTTC GTGGCCATCACGTTTATGGCCGCGGTGATGCTCGTGGCCTTCTACAAGCTGCGCAAGC AGCACCAGCTCCACAAGCACCACGGGCCACGCGCACCGTGGAGATCATCAACGTGGA	

	GGACGAGCTGCCCCGCCCTCGGCCGTGTCCGTGGCCGCCGCGGCCGCCGTGGCCAGT GGGGGTGGTGTGGGCGGGGACAGCCACCTGGCCCTGCCCGCCCTGGAGCGAGACCACC TCAACCACCACCACTACGTGGCTGCCGCCTTCAAGGCGCACTACAGCAGCAACCCAG CGGCGGGGGCTGCGGGGGCAAAGGCCCGCCTGGCCTCAACTCCATCCACGAACCTCTG CTCTTCAAGAGCGGCTCCAAGGAGAACGTGCAAGAGACGCAGATCTGAGGCGGCGGGG CCGGGCGGGCGAGGGGCGTGGAGCCCCCACCAGGTCCCAGCCCGGGCGCAGC		
	ORF Start: ATG at 111	ORF Stop: TGA at 2250	
	SEQ ID NO: 142	713 aa	MW at 76433.0kD
NOV31a, CG59623-01 Protein Sequence	MARARGSPCPPLPPGRMSWPHGALLFLWLFSPPLGAGGGGVAVTSAAGGSGPPATSCP VACSCSNQASRVICTRRDLAEVPASIPVNTRYLNQENGIQVIRTDTFKHLRHLEILQ LSKNLVRKIEVGAFNGLPSLNTLELFDNRLTTVPTQAFEYLSKLRELWLRNNPIESIP SYAFNRVPSLRRLDLGELKRLEYISEAAFEGLVNLRYLNLMCNLKDIPNLTAIVRLE ELELSGNRLDLIRPGSFQGLTSLRKLWLMHAQVATIERNADFDDLKSLEELNLSHNNLM SLPHDLFTPLHRLERVHLNHNPNWHCNCVDLWLSWWLKETVPSNTTCCARCHAPAGLKG RYIGELDQSHFTCYAPVIVEPPTDLNVTEGMAAELKCRGTGTSMTSVNWLTPNGTLMTH GSYRVRISVLHDGTLNFTNVTVDGTGQYTCMVTNSAGNTTASATLNVSAVDPVAAGGT GSGGGGPGSGGGVGGSGGYTYFTTVTVETLETQPGEEALQPRGTEKEPPGPTTDGVW GGGRPGDAAGPASSSTTAPAPRSSRPTEKAFTVPITDVTENALKDLDVMMKTTKIIIG CFVAITFMAAVMLVAFYKLRKQHLKHGHPTRTVEIINVEDELPAASAVSVAAAAAV ASGGGVGGDSHLALPALERDHLNHHHYVAAAFKAHYSSNPSSGGGCGGKGPPGLNSIHE PLLFKSGSKENVQETQI		

Further analysis of the NOV31a protein yielded the following properties shown in Table 31B.

Table 31B. Protein Sequence Properties NOV31a	
PSort analysis:	0.7000 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.2000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 38 and 39

A search of the NOV31a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
5 homologous proteins shown in Table 31C.

Table 31C. Geneseq Results for NOV31a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE13006	Human leucine-rich repeat (LRR) family member protein - Homo sapiens, 713 aa. [WO200175105-A2, 11-OCT-2001]	1..713 1..713	712/713 (99%) 712/713 (99%)	0.0
AAU12355	Human PRO331 polypeptide sequence - Homo sapiens, 640 aa. [WO200140466-A2, 07-JUN-2001]	54..713 44..640	406/660 (61%) 485/660 (72%)	0.0

AAU00826	Human immune response protein PRO331 (UNQ292) - Homo sapiens, 640 aa. [WO200119991-A1, 22-MAR-2001]	54..713 44..640	406/660 (61%) 485/660 (72%)	0.0
AAB53089	Human angiogenesis-associated protein PRO331, SEQ ID NO:137 - Homo sapiens, 640 aa. [WO200053753-A2, 14-SEP-2000]	54..713 44..640	406/660 (61%) 485/660 (72%)	0.0
AAB65292	Human PRO331 protein sequence SEQ ID NO:501 - Homo sapiens, 640 aa. [WO200073454-A1, 07-DEC-2000]	54..713 44..640	406/660 (61%) 485/660 (72%)	0.0

In a BLAST search of public sequence databases, the NOV31a protein was found to have homology to the proteins shown in the BLASTP data in Table 31D.

Table 31D. Public BLASTP Results for NOV31a				
Protein Accession Number	Protein/Organism/Length	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD10322	SEQUENCE 1 FROM PATENT WO0175105 - Homo sapiens (Human), 713 aa.	1..713 1..713	712/713 (99%) 712/713 (99%)	0.0
Q9NT99	HYPOTHETICAL 45.1 KDA PROTEIN - Homo sapiens (Human), 422 aa (fragment).	216..637 1..422	422/422 (100%) 422/422 (100%)	0.0
T46266	hypothetical protein DKFZp761A179.1 - human, 421 aa (fragment).	216..636 1..421	421/421 (100%) 421/421 (100%)	0.0
Q9HCJ2	KIAA1580 PROTEIN - Homo sapiens (Human), 640 aa (fragment).	54..713 44..640	406/660 (61%) 485/660 (72%)	0.0
Q9HBW1	BRAIN TUMOR ASSOCIATED PROTEIN NAG14 - Homo sapiens (Human), 653 aa.	42..713 34..653	381/672 (56%) 475/672 (69%)	0.0

PFam analysis predicts that the NOV31a protein contains the domains shown in the Table 31E.

Table 31E. Domain Analysis of NOV31a			
Pfam Domain	NOV31a Match Region	Identities/ Similarities for the Matched Region	Expect Value

LRRNT: domain 1 of 1	56..85	13/31 (42%) 21/31 (68%)	4.8e-06
LRR: domain 1 of 9	87..110	6/25 (24%) 18/25 (72%)	1.1
LRR: domain 2 of 9	111..134	9/25 (36%) 17/25 (68%)	0.38
LRR: domain 3 of 9	135..158	8/25 (32%) 19/25 (76%)	0.074
LRR: domain 4 of 9	159..182	10/25 (40%) 18/25 (72%)	0.013
LRR: domain 5 of 9	183..207	7/26 (27%) 19/26 (73%)	42
LRR: domain 6 of 9	208..229	8/25 (32%) 17/25 (68%)	1.5
LRR: domain 7 of 9	230..253	12/25 (48%) 20/25 (80%)	0.0068
LRR: domain 8 of 9	254..277	5/25 (20%) 16/25 (64%)	70
LRR: domain 9 of 9	278..301	14/25 (56%) 20/25 (80%)	0.00088
LRRCT: domain 1 of 1	311..362	19/54 (35%) 36/54 (67%)	6.2e-05
ig: domain 1 of 1	378..438	15/65 (23%) 41/65 (63%)	2.2e-07

EXAMPLE 32.

The NOV32 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 32A.

Table 32A. NOV32 Sequence Analysis		
	SEQ ID NO: 143	1206 bp
NOV32a, CG59247-01 DNA Sequence	AAACTGACCACAAAAGGGCTAACGGAATTTTtaggggatgataaatatgttcaacatc TTGACCATGGTGGCTGGCTGTGCCCTGGCCCTGGTGATGGTTGTCCAGCTGGGGCAGC AGATATTAATGTGCCAGGCAGTGCTGGCAGGTGAAGCCCCGAGTGGACCCTGTAGATC GGATGGAGACCACGTGGAGTACCACTACAGCAAGGCCATGCCACTCATCTTCATGAGC AGTATGCTCTGCAGCGGCACCATGCTGATACTCACTGGGCTGGATGCCACCTTGAGG TGCACCGCAGCAAGGAGACGCACATCATCCCTTGTGTGCTGGCCATGCACCAGGCCTG GTCCAAGTCTGGCCACAAGAACTACAGCTGGACAAGGCGGGCGTGACTGACGAGGTG CTGGACATTGCCATGCAGGCCTTCATCCTGGAGGTGATCTCTAAGCAAAGGGAGCCAG CCCATGTGCTCTCCAACAAGGACCACTTCAGACTCAAGTCCCTTGAATCACTGGTCTA CCTGTACACCTGTTCTCCAGCTCCAAGTTCCTGCTTATGGCTCAGGACAGCCATGTC TCCATGCACTCCTTGATCACATGCAAGGTCACTATTGCAGGCTTCGACCTCAGCAGCT ATGGCAACTGCCTCACCAAGTGGAACAAGGCCATAGAGGTGATGTACACCCAGTGCAT GGAGGTGGGCAAGGACAAGTGCCTGCTCGTGTACTACAAGGAAGTGGTGCTGCCTAGG AGCTTCCTCAGACTCATCCCAGACCATCTCGGCATCACCTGGAGCAACACTGTCCTCC	

	ACCATCAAGACCTCACTGGCAAGTGGAATGGCATCTCCCTGTCTAAGATCCAGTGGTC CATGGATGAGGTCATCAAGCCTGTGAACCTGGAAGTGCTCTCCAAGTGGACTCACCAC ATCCCTGGGGACATGGTGCCAGACATGGCCCAGATTGTCCCATGCTGGCTCAGCTTAG CCATGACCCCTATGCAAATACCCTCCCCAACCCCACTTCCACTATAGCAACCCTGAC CCCCATCATCATCAGTAACGCACACCAAGTAAGGGACTATAAAACACCAGTCAATCTG AAAGGATATTTTCAGGTGAACCAGAATAGCACCTCCTCCCCTTAGGAAGCTCATGAT TTCCAGATCTCTGCAAATGGCTTTGTTGCCCAAAGAGAAGAACT		
	ORF Start: ATG at 38	ORF Stop: TGA at 1157	
	SEQ ID NO: 144	373 aa	MW at 41568.3kD
NOV32a, CG59247-01 Protein Sequence	MINMFNILTMVAGCALALVMVVQLGQQILMCQAVLAGEAPSGPCRSDGDHVEYHYSKA MPLIFMSSMLCSGTMLILTGLDAHLEVHRSKETHIIPCVLAMHQAWSKSGHKKLQLDK AGVTDEVLDIAMQAFILEVISKQREPAHVLSNKHDFRLKSLESLVYLSHLFSSSKFL MAQDSHVSMHSLITCKVTIAGFDLSSYGNCLTKWNKAI EVMYTQCM EVGKDKCLLVYY KELVLP RSFLRLIPDHLGITWSNTVLHHQDLTGKWNGLSLSKIQWSMDEVIKPVNLEV LSKWTHHIPGDMVPDMAQIVPCWLSLAMTPMQIPSPPTSTIATLTPIIISNAHQVRD YKTPVNLKGYFQVNQNSTSSHLGSS		

Further analysis of the NOV32a protein yielded the following properties shown in Table 32B.

Table 32B. Protein Sequence Properties NOV32a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1279 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 28 and 29

A search of the NOV32a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several

5 homologous proteins shown in Table 32C.

Table 32C. Geneseq Results for NOV32a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV32a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM93565	Human polypeptide, SEQ ID NO: 3341 - Homo sapiens, 377 aa. [EP1130094-A2, 05-SEP-2001]	10..373 10..377	240/376 (63%) 273/376 (71%)	e-123
AAM93219	Human polypeptide, SEQ ID NO: 2626 - Homo sapiens, 377 aa. [EP1130094-A2, 05-SEP-2001]	10..373 10..377	240/376 (63%) 273/376 (71%)	e-123
AAY69421	Amino acid sequence of human TPST-2 polypeptide - Homo sapiens, 377 aa. [WO9965712-A2, 23-DEC-1999]	10..373 10..377	240/376 (63%) 273/376 (71%)	e-123

AAY84306	A human tyrosylprotein sulfotransferase 2 (TPST-2) polypeptide - Homo sapiens, 377 aa. [WO200014250-A1, 16-MAR-2000]	10..373 10..377	240/376 (63%) 273/376 (71%)	e-123
AAY06625	Human tyrosylprotein sulfotransferase TPST-2 - Homo sapiens, 377 aa. [WO9938980-A2, 05-AUG-1999]	10..373 10..377	240/376 (63%) 273/376 (71%)	e-123

In a BLAST search of public sequence databases, the NOV32a protein was found to have homology to the proteins shown in the BLASTP data in Table 32D.

Table 32D. Public BLASTP Results for NOV32a				
Protein Accession Number	Protein/Organism/Length	NOV32a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O60704	Protein-tyrosine sulfotransferase 2 (EC 2.8.2.20) (Tyrosylprotein sulfotransferase-2) (TPST-2) - Homo sapiens (Human), 377 aa.	10..373 10..377	240/376 (63%) 273/376 (71%)	e-122
O88856	Protein-tyrosine sulfotransferase 2 (EC 2.8.2.20) (Tyrosylprotein sulfotransferase-2) (TPST-2) - Mus musculus (Mouse), 376 aa.	10..373 10..376	237/375 (63%) 270/375 (71%)	e-121
O70281	Protein-tyrosine sulfotransferase 1 (EC 2.8.2.20) (Tyrosylprotein sulfotransferase-1) (TPST-1) - Mus musculus (Mouse), 370 aa.	10..324 10..332	159/326 (48%) 213/326 (64%)	2e-78
O60507	Protein-tyrosine sulfotransferase 1 (EC 2.8.2.20) (Tyrosylprotein sulfotransferase-1) (TPST-1) - Homo sapiens (Human), 370 aa.	10..363 10..368	168/369 (45%) 226/369 (60%)	2e-78
Q9VYB7	Probable protein-tyrosine sulfotransferase (EC 2.8.2.20) (Tyrosylprotein sulfotransferase) (TPST) - Drosophila melanogaster (Fruit fly), 385 aa.	46..324 57..333	131/280 (46%) 182/280 (64%)	5e-68

PFam analysis predicts that the NOV32a protein contains the domains shown in the Table 32E.

Table 32E. Domain Analysis of NOV32a			
Pfam Domain	NOV32a Match Region	Identities/	Expect Value

		for the Matched Region	
Sulfotransfer: domain 1 of 1	36..313	38/311 (12%) 150/311 (48%)	7.9

EXAMPLE 33.

The NOV33 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 33A.

Table 33A. NOV33 Sequence Analysis			
	SEQ ID NO: 145	1240 bp	
NOV33a, CG59430-01 DNA Sequence	ACCAAGGACCCCAGAGGATGGAGGCCTCTCGGTGGTGGCTGCTGGTCACTGTGCTCAT GGCTGGGGCTCATTGTGTGGCCCTGGTTGACCAAGAAGCTTCTGATCTCATCCATTCT GGCCCCCAGGACAGCAGCCCTGGGCCTGCCCTGCCACAAAATCTCTGTGAGCA ACATAGACTTTGCCTTCAAGCTCTACAGACAGTTGGCTTTGAACGCCCCCGGGGAGAA CATTCTCTTCTCCCCAGTGAGCATCTCCCTGGCCTTGGCCATGCTTTCTTGGGGGGCC CCAGTGGCCAGCAGGACCCAACTCCTGGAGGGCCTGGGGTTACCCCTCACCGTGGTG CTGAGGAGGAGATCCAGGAAGGCTTCTGGGATCTGCTGATCAGGCTCCGTGGGCAGGG TCCCCGGCTCCTCCTGACCATGGACCAGCGCAGGTTTCAGCGGCCTGGGCGCGAGGGCC AACCAGAGCCTAGAGGAGGCCCAAAACACATTGACGAATATACAGAGCAGCAGACCC AGGGGAAGCTCGGGGCCTGGGAGAAGGACCTCGGCAGTGAAACCACAGCGGTTCTGGT GAATCACATGCTCCTCAGAGCTGAGTGGATGAAGCCCTTTGACTCACGTGCCACCAGC CCAAAGGAGTTCTTTGTAGATGAGCACAGCGCTGTGTGGGTGCCCATGATGAAGGAGA AGGCCAGCCACCGCTTCTGCACGACCGTGAGCTGCAATGCTCTGTGCTGCGGATGGA CCACGCTGGGAACACCACCACCTTCTTCATCTTCCCCAACAGGGGCAAGATGAGGCAG CTGGAAGATGCCCTGCTGCCTGAAACACTGATTAAGTGGGACAGTCTGCTCAGGCTCG ATTTCCACTTCCCCAAATTTTCCATTTCTAGAACCTGCAGACTGGAGATGCTCCTCCC AAAAGTGACTGTGGGTGGAGGCTTCCCTGGGCAGCCTGGACTGAACATTTCTAAAGTA AGTTGGGGATGGTGTGTTTCAAGGGCCTCTCATAAGGCCATGATGACGCTGGATGAGA GGGGCTCTGAAGCTGCTGCAGCCACCAGCATTGAGCTCACCCCTGGGCCTCGCCCAGA CCTTGACTTCCCACCCACTCTGGGCACTGAGTTCAGTCGGCCCTTCTGGTGATGACT TTCCACACGGAAACAGGAAGCATGCTTTTTTCTGGAGAAGATTGTAAACCCACTGGGAT AACGCCCCCTCAGACATGCTGG		
	ORF Start: ATG at 18	ORF Stop: TAA at 1218	
	SEQ ID NO: 146	400 aa	MW at 44726.0kD
NOV33a, CG59430-01 Protein Sequence	MEASRWLLVTVLMAHAHCVALVDQEASDLIHSGPQDSSPGPALPCHKISVSNIDFAF KLYRQLALNAPGENILFSPVSI SLALAMLSWGAPVASRTQLLEGLGFTLTVP EEEIQ EGFWDLLIRLRGQGPRLLLLTMDQRRFSGLGARANQSLEEAQKHIDEYTEQQQTQGLGA WEKDLGSETTAVLVNHMLLRAEWMKPFDSRATSPKEFFVDEHSAVWVPMMEKASHRF LHDRELQCSVL RMDHAGNTTTFIFPNRGKMRQLEDALLPETLIKWDSLRLDFHFPK FSISRTCRLEMLLPKVTVGGGFPGQPLNISKVSWGVCVQRASHKAMMTLDERGSEAA AATSIQLTPGPRPDLD FPPTLGTEFSRPFLVMTFHTETGSMLFLEKIVNPLG		

Further analysis of the NOV33a protein yielded the following properties shown in

5 Table 33B.

Table 33B. Protein Sequence Properties NOV33a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1700 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)

SignalP analysis:	Likely cleavage site between residues 20 and 21
-------------------	---

A search of the NOV33a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 33C.

Table 33C. Geneseq Results for NOV33a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV33a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB68434	Amino acid sequence of human serpin protease Zserp11 - Homo sapiens, 366 aa. [WO200138534-A2, 31-MAY-2001]	1..400 1..366	352/400 (88%) 356/400 (89%)	0.0
AAO13910	Human polypeptide SEQ ID NO 27802 - Homo sapiens, 495 aa. [WO200164835-A2, 07-SEP-2001]	7..399 80..493	164/431 (38%) 234/431 (54%)	4e-67
AAG73736	Human colon cancer antigen protein SEQ ID NO:4500 - Homo sapiens, 446 aa. [WO200122920-A2, 05-APR-2001]	7..399 31..444	164/431 (38%) 234/431 (54%)	4e-67
AAY28643	Human serine protease inhibitor from cDNA clone HETDK50 - Homo sapiens, 422 aa. [WO9940183-A1, 12-AUG-1999]	7..399 7..420	164/431 (38%) 234/431 (54%)	4e-67
AAB74691	Human protease and protease inhibitor PPIM-24 - Homo sapiens, 422 aa. [WO200110903-A2, 15-FEB-2001]	7..399 7..420	163/431 (37%) 233/431 (53%)	1e-66

In a BLAST search of public sequence databases, the NOV33a protein was found to have homology to the proteins shown in the BLASTP data in Table 33D.

Table 33D. Public BLASTP Results for NOV33a				
Protein Accession Number	Protein/Organism/Length	NOV33a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC42686	SEQUENCE 1 FROM PATENT WO0138534 - Homo sapiens (Human), 366 aa.	1..400 1..366	352/400 (88%) 356/400 (89%)	0.0
Q96BZ5				5e-66

	PROTEIN - Homo sapiens (Human), 427 aa.	9..424	235/423 (55%)	
P29622	Kallistatin precursor (Kallikrein inhibitor) (Protease inhibitor 4) - Homo sapiens (Human), 427 aa.	8..398 9..424	161/423 (38%) 235/423 (55%)	1e-65
P05544	Contrapsin-like protease inhibitor 3 precursor (CPI-23) (Serine protease inhibitor 1) (SPI-1) - Rattus norvegicus (Rat), 413 aa.	12..398 9..412	151/412 (36%) 222/412 (53%)	3e-60
S08102	serine proteinase inhibitor 1 - rat, 403 aa.	36..398 22..402	145/388 (37%) 213/388 (54%)	4e-60

PFam analysis predicts that the NOV33a protein contains the domains shown in the Table 33E.

Table 33E. Domain Analysis of NOV33a			
Pfam Domain	NOV33a Match Region	Identities/ Similarities for the Matched Region	Expect Value
serpin: domain 1 of 3	46..137	48/93 (52%) 74/93 (80%)	8.9e-31
serpin: domain 2 of 3	154..306	68/168 (40%) 105/168 (62%)	6.5e-34
serpin: domain 3 of 3	332..398	31/71 (44%) 53/71 (75%)	1.1e-14

EXAMPLE 34.

The NOV34 clone was analyzed, and the nucleotide and predicted polypeptide
5 sequences are shown in Table 34A.

Table 34A. NOV34 Sequence Analysis		
	SEQ ID NO: 147	1026 bp
NOV34a, CG59305-01 DNA Sequence	ATGCCACAGCCCAGGGGAGGCCAGCCTGCCTGGCAGCTGACACCCAGCCCTCCCCCA GCTCCCGGATAATGAGCACCCATGTGGCAGGCCTGGGCCTGGACAAGATGAAGCTGGG CAATCCCCAGTCCTTCTGGACCAGGAGGAGGCAGATGACCAGCAGCTGCTGGAACCA GAGGCGTGGAAGACCTACACCGAGCGCCGCAATGCCCTGCGTGAGTTCTTGACCTCGG ACCTGAGTCCGCACCTGCTCAAGCGCCACCACGCCCGCATGCAGCTGCTGCGTAAGTG CTCCTACTACATCGAGGTCCTGCCCAAGCACCTGGCCCTGGGCGACCAGAACCCGCTG GTGCTGCCTAGCGCCTTGTTCCAGCTCATCGACCCCTGGAAGTTCCAGCGCATGAAGA AGGTGGGCACAGCTCAGACCAAGATCCAGCTCCTGCTGCTCGGGACCTGTTGGAACA GCTCGACCATGGCCGTGCTGAGCTGGATGCCCTGCTCCGGTCGCCAGACCCACGGCCC TTCCTGGCCGACTGGGCGCTGGTGGAGCGGCGGCTGGCGGACGTGTCGGCCGTCATGG ACAGCTTCCTGACCATGATGGTGCCGGGGCGGCTACACGTCAAGCACCGCCTGGTGTC TGATGTCAGTGCCACCAAGATCCCGCACATCTGGCTCATGCTGAGCACCAAGATGCCT GTCGTGTTTGACCGAAAGGCGTCGGCGGCTCACCAGGACTGGGCCCCGGCTGCGCTGGT TCGTCACCATCCAGCCAGCCACATCGGAGCAGTATGAGTTGCGCTTCAGGCTGCTGGA CCCGCGGACACAGCAGGAGTGCGCCAGTGTTGGCGTCATCCCCGTGGCTGCCTGCACC	

	TTCGACGTCCGAAACCTGCTGCCCAACCGATCCTATAAGTTCACCATCAAGAGGGGCCG AGACCTCCACGCTGGTGTACGAGCCCTGGAGGGACAGCCTCACCCTGCACACCAAGCC GGAGCCCCTGGAGGGGGCCCGCCCTCAGCCACTCTGTCTGA		
	ORF Start: ATG at 1	ORF Stop: TGA at 1024	
	SEQ ID NO: 148	341 aa	MW at 38993.7kD
NOV34a, CG59305-01 Protein Sequence	MPQPRGGQPAWQLTPSPPPSSRIMSTHVAGLGLDKMKLGNPQSFLDQEEADDQQLLEP EAWKTYTERRNALREFLTSDLSPHLLKRHHARMQLLRKCSYYIEVLPHKHLALGDQNPL VLPSALFQLIDPWKFQRMKKVGTAQTKIQLLLLGLDLEQLDHGRAELDALLRSPDPRP FLADWALVERRLADVSAVMDSFLTMMVPGRLHVKHRLVSDVSATKI PHIWLMLSTKMP VVFDRKASAAHQDWARLRWFVTIQPATSEQYELRFRLLDPRTQQECAQCGVIPVAECT FDVRNLLPNRSYKFTIKRAETSTLVYEPWRDSLTLHTKPEPLEGPALSHSV		
	SEQ ID NO: 149	1026 bp	
NOV34b, CG59305-02 DNA Sequence	ATGCCACAGCCCAGGGGAGGCCAGCCTGCCTGGCAGCTGACACCCAGCCCTCCCCCA GCTCCCGGATAATGAGCACCCATGTGGCAGGCCTGGGCCTGGACAAGATGAAGCTGGG CAATCCCCAGTCCTTCTGGACCAGGAGGAGGCAGATGACCAGCAGCTGCTGGAACCA GAGGCGTGGAAGACCTACACCGAGCGCCGCAATGCCCTGCGTGAGTTCCTGACCTCGG ACCTGAGTCCGCACCTGCTCAAGCGCCACCACGCCCGCATGCAGCTGCTGCGTAAGTG CTCCTACTACATCGAGGTCCTGCCCAAGCACCTGGCCCTGGGCGACCAGAACCCGCTG GTGCTGCCTAGCGCCTTGTTCCAGCTCATCGACCCCTGGAAGTTCAGCGCATGAAGA AGGTGGGCACAGCTCAGACCAAGATCCAGCTCCTGCTGCTCGGGGACCTGTTGGAACA GCTCGACCATGGCCGTGCTGAGCTGGATGCCCTGCTCCGGTCGCCAGACCCACGGCCC TTCCTGGCCGACTGGGCGCTGGTGGAGCGGCGGCTGGCGGACGTGTCGGCCGTGTCATGG ACAGCTTCCTGACCATGATGGTGCCGGGGCGGCTACACGTCAAGCACCGCCTGGTGTC TGATGTCAGTGCCACCAAGATCCCGCACATCTGGCTCATGCTGAGCACCAAGATGCCT GTCGTGTTTGACCGAAAGGCGTCGGCGGCTCACCAGGACTGGGCCCCGGCTGCGCTGGT TCGTCACCATCCAGCCAGCCACATCGGAGCAGTATGAGTTGCGCTTCAGGCTGCTGGA CCCGCGGACACAGCAGGAGTGCGCCAGTGTGGCGTCATCCCCGTGGCTGCCTGCACC TTCGACGTCCGAAACCTGCTGCCCAACCGATCCTATAAGTTCACCATCAAGAGGGCCG AGACCTCCACGCTGGTGTACGAGCCCTGGAGGGACAGCCTCACCCTGCACACCAAGCC GGAGCCCCTGGAGGGGGCCCGCCCTCAGCCACTCTGTCTGA		
	ORF Start: ATG at 1	ORF Stop: TGA at 1024	
	SEQ ID NO: 150	341 aa	MW at 38993.7kD
NOV34b, CG59305-02 Protein Sequence	MPQPRGGQPAWQLTPSPPPSSRIMSTHVAGLGLDKMKLGNPQSFLDQEEADDQQLLEP EAWKTYTERRNALREFLTSDLSPHLLKRHHARMQLLRKCSYYIEVLPHKHLALGDQNPL VLPSALFQLIDPWKFQRMKKVGTAQTKIQLLLLGLDLEQLDHGRAELDALLRSPDPRP FLADWALVERRLADVSAVMDSFLTMMVPGRLHVKHRLVSDVSATKI PHIWLMLSTKMP VVFDRKASAAHQDWARLRWFVTIQPATSEQYELRFRLLDPRTQQECAQCGVIPVAECT FDVRNLLPNRSYKFTIKRAETSTLVYEPWRDSLTLHTKPEPLEGPALSHSV		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 34B.

Table 34B. Comparison of NOV34a against NOV34b and NOV34c.		
Protein Sequence	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV34b	1..341 1..341	328/341 (96%) 328/341 (96%)

Further analysis of the NOV34a protein yielded the following properties shown in Table 34C.

Table 34C. Protein Sequence Properties NOV34a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.4466 probability located in microbody (peroxisome); 0.2245 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV34a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 34D.

Table 34D. Geneseq Results for NOV34a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found				

In a BLAST search of public sequence databases, the NOV34a protein was found to have homology to the proteins shown in the BLASTP data in Table 34E.

Table 34E. Public BLASTP Results for NOV34a				
Protein Accession Number	Protein/Organism/Length	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BVV2	HYPOTHETICAL 36.5 KDA PROTEIN - Homo sapiens (Human), 318 aa.	24..341 1..318	318/318 (100%) 318/318 (100%)	0.0
Q9D9W3	1700026M20RIK PROTEIN - Mus musculus (Mouse), 163 aa.	66..173 2..109	89/108 (82%) 96/108 (88%)	6e-46

PFam analysis predicts that the NOV34a protein contains the domains shown in the Table 34F.

Table 34F. Domain Analysis of NOV34a			
Pfam Domain	NOV34a Match Region	Identities/ Similarities for the Matched Region	Expect Value
fn3: domain 1 of 1	231..312	10/87 (11%) 52/87 (60%)	5.9

EXAMPLE 35.

The NOV35 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 35A.

Table 35A. NOV35 Sequence Analysis			
	SEQ ID NO: 151	1610 bp	
NOV35a, CG59547-01 DNA Sequence	CGTCTTCGGGACGCGCCCGCTCTTCGCCTTTCGCTGCAGTCCGTCGATTTCTTTCTCC AGGAAGAAAAATGGCATCCGTTGCAGTTGATCCACAACCGAGTGTGGTGACTCGGGTG GTCAACCTGCCCTTGGTGAGCTCCACGTATGACCTCATGTCTCAGCCTATCTCAGTA CAAAGGACCAGTATCCCTACCTGAAGTCTGTGTGTGAGATGNCAGAGAACGGTGTGAA GACCATCACCTCCGTGGCCATGACCAGTGCTCTGCCCATCATCCAGAAGCTAGAGCCG CAAATTGCAGTTGCCGATACCTATGCCTGTAAGGGGCTAGACAGGATTGAGGAGAGAC TGCCTATTCTGAATCAGCCATCAACTCAGATTGTTGCCAATGCCAAAGGCGCTGTGAC TGGGGCAAAGATGCTGTGACGACTACTGTGACTGGGGCCAAGGATTCTGTNGCCAGC ACGATCACAGGGGTGATGGACAAGACCAAAGGGGCAGTGACTGGCAGTGTGGAGAAGA CCAAGTCTGTGGTCAGTGGCAGCATTAAACACAGTCTTGGGGAGTCGGATGATGCAGCT CGTGAGCAGTGGCGTAGAAAATGCACTACCAAATCAGAGCTGTTGGTAGAACAGTAC CTCCCTCTCACTGAGGAAGAACTAGAAAAAGAAGCAAAAAAAGTTGAAGGATTTGATC TGGTTCAGAAGCCAAGTTATTATGTTAGACTGGGATCCCTGTCTACCAAGCTTCACTC CCGTGCCTACCAGCAGGCTCTCAGCAGGGTTAAAGAAGCTAAGCAAAAAAGCCAACAG ACCATTTCTCAGCTCCATTCTACTGTTACCTGATTGAATTTGCCAGGAAGAATGTGT ATAGTGCCAATCAGAAAATTCAGGATGCTCAGGATAAGCTCTACCTCTCATGGGTAGA GTGGAAAAGGAGCATTGGATATGATGATACTGATGAGTCCCACTGTGCTGAGCACATT GAGTCACGTACTCTTGCAATTGCCCGCAACCTGACTCAGCAGCTCCAGACCACGTGCC ACACCCTCCTGTCCAACATCCTTTGTGTACCACAGAACATCCCCATCATTTTTTTGCA AAAGGGGGTGATGGCAGGCGACATCTACTCAGTGTTCCGGAATGCTGCCTCCTTTAA GAAGTGTCTGACAGCCTCCTCACTTCTAGCAAGGGGCAGCTGCAGAAAATGAAGGAAT CTTTAGATGACGTGATGGATTATCTTGTTTACAAAACGCCCTAAACTGGCTGGTAGG TCCCTTTTATCCTCAGCTGACTGAGTCTCAGAATGCTCAGGACCAAGGTGCAGAGATG GACAAGAGCAGCCAGGAGACCCAGCGATCTGAGCATAAACTCATTAAACCTGCCCCCT ATCACTAGTGCATGCTGTGGCCAGACAGATGACACCTTTTGTATGTTGAAATTAAC TGCTAGGCAACCCTAAATTGGGAAGCAAGTAGCTAGTATAAAGGCCCTCAATTGTAGT TGTTTCCAGCTGAATTAAGAGCTTTAAAGTTTCTGGCATTAGCAGATGATTTCTGTTC ACCTGGTAAGAAAAGAATGATAGGCTTGTCAGAGCCTATAGCCA		
	ORF Start: ATG at 69	ORF Stop: TAA at 1380	
	SEQ ID NO: 152	437 aa	MW at 48148.2kD
NOV35a, CG59547-01 Protein Sequence	MASVAVDPQPSVVTRVNLPLVSSSTYDLMSSAYLSTKDQYPYLKSVCEMXENGVKITIT SVAMTSALPIIQKLEPQIAVADTYACKGLDRIEERLPILNQPSTQIVANAKGAVTGAK DAVTTTVTGAKDSVASTITGVMDKTKGAVTGSVEKTKSVVSGSINTVLGSRMMQLVSS GVENALTKSELLVEQYLPLETEEELEKEAKKVEGFDLVQKPSYYVRLGSLSTKLHSRAY QQALSRVKEAKQKSQQTISQLHSTVHLIEFARKNVYSANQKIQDAQDKLYLSWVEWKR SIGYDDTDESHCAEHIESRTLAIARNLTQQLQTTCHTLLSNILCVPQNI PHHFLQKGV MAGDIYSVFRNAASFKEVSDSLTSSKGQLQKMKESLDDVMDYLVYKTPLNWLVGPFY PQLTESQNAQDQGAEMDKSSQETQRSEHKTH		

Further analysis of the NOV35a protein yielded the following properties shown in

5 Table 35B.

Table 35B. Protein Sequence Properties NOV35a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)

SignalP analysis:	No Known Signal Sequence Predicted
-------------------	------------------------------------

A search of the NOV35a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 35C.

Table 35C. Geneseq Results for NOV35a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY99534	Human adipocyte-specific differentiation-related protein ADRP - Homo sapiens, 437 aa. [WO200031532-A1, 02-JUN-2000]	1..300 138..437	289/300 (96%) 289/300 (96%)	e-161
AAW53264	Human adipocyte-specific differentiation-related protein - Homo sapiens, 437 aa. [US5739009-A, 14-APR-1998]	1..300 138..437	289/300 (96%) 289/300 (96%)	e-161
AAB58800	Breast and ovarian cancer associated antigen protein sequence SEQ ID 508 - Homo sapiens, 250 aa. [WO200055173-A1, 21-SEP-2000]	51..300 1..250	238/250 (95%) 238/250 (95%)	e-133
AAW06798	Murine p154 - Mus sp, 425 aa. [US5541068-A, 30-JUL-1996]	1..298 138..423	231/298 (77%) 256/298 (85%)	e-125
AAR45151	Sequence of mouse adipocyte polypeptide (ap) p154 - Acomys cahirinus, 425 aa. [US5268295-A, 07-DEC-1993]	1..298 138..423	231/298 (77%) 256/298 (85%)	e-125

In a BLAST search of public sequence databases, the NOV35a protein was found to have homology to the proteins shown in the BLASTP data in Table 35D.

Table 35D. Public BLASTP Results for NOV35a				
Protein Accession Number	Protein/Organism/Length	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC09025	SEQUENCE 22 FROM PATENT WO0031532 - Homo sapiens (Human), 437 aa.	1..300 138..437	289/300 (96%) 289/300 (96%)	e-160
Q9BSC3				e-160

	RELATED PROTEIN - Homo sapiens (Human), 437 aa.	138..437	289/300 (96%)	
Q99541	Adipophilin (Adipose differentiation-related protein) (ADRP) - Homo sapiens (Human), 437 aa.	1..300 138..437	287/300 (95%) 287/300 (95%)	e-159
Q9TUM6	Adipophilin (Adipose differentiation-related protein) (ADRP) - Bos taurus (Bovine), 450 aa.	1..282 138..419	239/282 (84%) 258/282 (90%)	e-132
Q9MZE5	ADIPOSE DIFFERENTIATION-RELATED PROTEIN - Sus scrofa (Pig), 404 aa (fragment).	1..267 138..404	231/267 (86%) * 246/267 (91%)	e-127

PFam analysis predicts that the NOV35a protein contains the domains shown in the Table 35E.

Table 35E. Domain Analysis of NOV35a			
Pfam Domain	NOV35a Match Region	Identities/ Similarities for the Matched Region	Expect Value
SPX: domain 1 of 1	29..153	24/347 (7%) 80/347 (23%)	8
perilipin: domain 1 of 1	1..259	166/411 (40%) 247/411 (60%)	7.4e-89

EXAMPLE 36.

The NOV36 clone was analyzed, and the nucleotide and predicted polypeptide
5 sequences are shown in Table 36A.

Table 36A. NOV36 Sequence Analysis			
	SEQ ID NO: 153	355 bp	
NOV36a, CG58508-01 DNA Sequence	ACCTCTTTGCCACCAATACCATGAAGCTCTGCGTGACTGTCCTGTCTCTCCTCGTGCT AGTAGCTGCCTTCTGCTCTCTAGCACTCTCAGCACCAATGGGCTCAGACCCTCCCACC GCCTGCTGCTTTTCTTACACCGCGAGGAAGCTTCCTCACAACCTTTGTGGTAGATTACT ATGAGACCAGCAGCCTCTGCTCCAGCCAGCTGTGGTATTCCAAACCAAAAGAGGCAA GCAAGTCTGCGCTGACCCAGTGAGTCCTGGGTCCAGGAGTACGTGTATGACCTGGAA CTGAAGTGAAGCTGCTCAGAGACAGGACAGTCACGCAGAGCTTCATGGTATTGGTGGCA AAGAGGT		
	ORF Start: ATG at 21	ORF Stop: TGA at 297	
	SEQ ID NO: 154	92 aa	MW at 10146.6kD
NOV36a, CG58508-01 Protein Sequence	MKLCVTVLSLLVLVAAFCSLALSAPMGSDPPTACCFSTARKLPHNFVVDYYETSSLC SQPAVVFQTKRGKQVCADPSESWSVQYVYDLELN		
	SEQ ID NO: 155	355 bp	

NOV36b, CG58508-02 DNA Sequence	<u>ACCTCTTTGCCACCAATACCATGAAGCTCTGCGTGACTGTCTGTCTCTGAGCAGCTC</u> <u>AGTTCAGTTCCAGGTCATACACGTACTCCTGGACCCAGGACTCACTGGGGTCAGCGCA</u> <u>GACTTGCTTGCCCTCTTTTGGTTTGAATACCACAGCCGGCTGGGAGCAGAGGCTGCTG</u> <u>GTCTCATAGTAATCTACCACAAAGTTGCGAGGAAGCTTCCTCGCGGTGTAAGAAAAGC</u> <u>AGCAGGCGGTGGGAGGGTCTGAGCCCATTGGTGCTGAGAGTGCTAGAGAGCAGAAGGC</u> <u>AGCTACTAGCACGAGGAGAGACAGGACAGTCACGCAGAGCTTCATGGTATTGGTGGCA</u> <u>AAGAGGT</u>		
	ORF Start: ATG at 21	ORF Stop: TAG at 297	
	SEQ ID NO: 156	92 aa	MW at 10149.6kD
NOV36b, CG58508-02 Protein Sequence	MKLCVTVLSSLVLVA AFCSPALSAPMGSDPPTACCF SYTARKLPRNFVVDYYETSSLC SQPAVV FQTKRGKQVCADPSES VWQEY VYDLELN		
	SEQ ID NO: 157	219 bp	
NOV36c, 170072532 DNA Sequence	<u>GGATCCGCACCAATGGGCTCAGACCCTCCCACCGCCTGCTGCTTTTCTTACACCGCGA</u> <u>GGAAGCTTCCTCGCAACTTTGTGGTAGATTACTATGAGACCAGCAGCCTCTGCTCCCA</u> <u>GCCAGCTGTGGTATTCCAAACCAAAGAGGCAAGCAAGTCTGCGCTGACCCAGTGAG</u> <u>TCCTGGGTCCAGGAGTACGTGTATGACCTGGA ACTGAACCTCGAG</u>		
	ORF Start: GGA at 1	ORF Stop:	
	SEQ ID NO: 158	73 aa	MW at 8175.1kD
NOV36c, 170072532 Protein Sequence	GSAPMGSDPPTACCF SYTARKLPRNFVVDYYETSSLC SQPAVV FQTKRGKQVCADPSE SWVQEY VYDLELNLE		
	SEQ ID NO: 159	219 bp	
NOV36d, 170072551 DNA Sequence	<u>GGATCCGCACCAATGGGCTCAGACCCTCCCACCGCCTGCTGCTTTTCTTACACCGCGA</u> <u>GGAAGCTTCCTCGCAACTTTGTGGTAGATTACTATGAGACCAGCAGCCTCTGCTCCCA</u> <u>GCCAGCTGTGGTATTCCAAACCAAAGAAGCAAGCAAGTCTGTGCTGATCCCAGTGAA</u> <u>TCCTGGGTCCAGGAGTACGTGTATGACCTGGA ACTGAACCTCGAG</u>		
	ORF Start: GGA at 1	ORF Stop:	
	SEQ ID NO: 160	73 aa	MW at 8205.1kD
NOV36d, 170072551 Protein Sequence	GSAPMGSDPPTACCF SYTARKLPRNFVVDYYETSSLC SQPAVV FQTKRSKQVCADPSE SWVQEY VYDLELNLE		
	SEQ ID NO: 161	219 bp	
NOV36e, 170072555 DNA Sequence	<u>GGATCCGCACCAATGGGCTCAGACCCTCCCACCGCTTGCTGCTTTTCTTACACCGCGA</u> <u>GGAAGCTTCCTCGCAACTTTGTGGTAGATTACTATGAGACCAGCAGCCTCTGCTCCCA</u> <u>GCCAGCTGTGGTATTCCAAACCAAAGAAGCAAGCAAGTCTGTGCTGATCCCAGTGAA</u> <u>TCCTGGGTCCAGGAGTACGTGTATGACCTGGA ACTGAACCTCGAG</u>		
	ORF Start: GGA at 1	ORF Stop:	
	SEQ ID NO: 162	73 aa	MW at 8205.1kD
NOV36e, 170072555 Protein Sequence	GSAPMGSDPPTACCF SYTARKLPRNFVVDYYETSSLC SQPAVV FQTKRSKQVCADPSE SWVQEY VYDLELNLE		
	SEQ ID NO: 163	301 bp	
NOV36f, CG58508-03 DNA Sequence	<u>CAGCCTCACCTCTGAGAAAACCTCTTTTCCACCAATACC</u> <u>ATGAAGCTCTGCGTGACTG</u> <u>TCCTGTCTCTCCTCATGCTAGTAGCTGCCTTCTGCTCTCCAGCGCTCTCAGCCAGCTG</u> <u>TGGTATTCCAAACCAAAGAAGCAAGCAAGTCTGTGCTGAT</u> <u>CCCAGTGAATCCTGGGT</u> <u>CCAGGAGTACGTGTATGACCTGGA ACTGAACTGAGCTGCTCAGAGACAGGAAGTCTTC</u>		

	AGGGAAGGTCACCTGAGCCCGGATGCTTCTCCATGAGACACATCTCCTCCATACTCAG GACTCCTCTCA		
	ORF Start: ATG at 40	ORF Stop: TGA at 154	
	SEQ ID NO: 164	38 aa	MW at 3940.8kD
NOV36f, CG58508-03 Protein Sequence	MKLCVTVLSELLMLVAAFCSALSASCGIPNQKKQASLC		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 36B.

Table 36B. Comparison of NOV36a against NOV36b through NOV36f.		
Protein Sequence	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV36b	15..92 15..92	76/78 (97%) 76/78 (97%)
NOV36c	23..92 2..71	69/70 (98%) 69/70 (98%)
NOV36d	23..92 2..71	68/70 (97%) 68/70 (97%)
NOV36e	23..92 2..71	68/70 (97%) 68/70 (97%)
NOV36f	1..27 1..27	23/27 (85%) 24/27 (88%)

Further analysis of the NOV36a protein yielded the following properties shown in Table 36C.

Table 36C. Protein Sequence Properties NOV36a	
PSort analysis:	0.8200 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Likely cleavage site between residues 24 and 25

- 5 A search of the NOV36a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 36D.

Table 36D. Geneseq Results for NOV36a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV36a Residues/	Identities/ Similarities	Expect Value

		Residues	Matched Region	
AAR36770	MIP-1beta - Homo sapiens, 92 aa. [WO9309799-A, 27-MAY-1993]	1..92 1..92	89/92 (96%) 90/92 (97%)	9e-48
AAB15789	Human chemokine MIP1beta SEQ ID NO: 20 - Homo sapiens, 92 aa. [WO200042071-A2, 20-JUL-2000]	1..92 1..92	88/92 (95%) 89/92 (96%)	6e-47
AAW82717	Human Act-2 protein - Homo sapiens, 92 aa. [WO9854326-A1, 03-DEC-1998]	1..92 1..92	88/92 (95%) 89/92 (96%)	6e-47
AAW76225	Human chemokine MIP-1beta domain protein fragment - Homo sapiens, 92 aa. [WO9838212-A2, 03-SEP-1998]	1..92 1..92	88/92 (95%) 89/92 (96%)	6e-47
AAW76223	Human chemokine MIP-1beta domain protein from clone MPB-X - Homo sapiens, 331 aa. [WO9838212-A2, 03-SEP-1998]	1..92 1..92	88/92 (95%) 89/92 (96%)	6e-47

In a BLAST search of public sequence databases, the NOV36a protein was found to have homology to the proteins shown in the BLASTP data in Table 36E.

Table 36E. Public BLASTP Results for NOV36a				
Protein Accession Number	Protein/Organism/Length	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P13236	Small inducible cytokine A4 precursor (Macrophage inflammatory protein 1-beta) (MIP-1-beta) (T-cell activation protein 2) (ACT-2) (PAT 744) (H400) (SIS-gamma) (Lymphocyte activation gene-1 protein) (LAG-1) (HC21) (G-26 T lymphocyte-secreted protein) - Homo sapiens (Human), 92 aa.	1..92 1..92	88/92 (95%) 89/92 (96%)	2e-46
P46632	Small inducible cytokine A4 precursor (Macrophage inflammatory protein 1-beta) (MIP-1-beta) (Immune activation protein 2) (ACT-2) - Oryctolagus cuniculus (Rabbit), 92 aa.	1..92 1..92	75/92 (81%) 84/92 (90%)	7e-39
P50230	Small inducible cytokine A4 precursor (Macrophage inflammatory protein 1-beta) (MIP-1-beta) - Rattus norvegicus (Rat), 92 aa.	1..92 1..92	71/92 (77%) 81/92 (87%)	3e-38

P14097	Small inducible cytokine A4 precursor (Macrophage inflammatory protein 1-beta) (MIP-1-beta) (H400 protein) (SIS-gamma) (ACT2) - Mus musculus (Mouse), 92 aa.	1..92 1..92	69/92 (75%) 82/92 (89%)	1e-36
CAA01323	HUMAN ACT-2 SYNTHETIC GENE PROTEIN - synthetic construct, 74 aa (fragment).	19..92 1..74	69/74 (93%) 69/74 (93%)	2e-35

PFam analysis predicts that the NOV36a protein contains the domains shown in the Table 36F.

Table 36F. Domain Analysis of NOV36a			
Pfam Domain	NOV36a Match Region	Identities/ Similarities for the Matched Region	Expect Value
IL8: domain 1 of 1	24..89	25/70 (36%) 60/70 (86%)	2.6e-32

EXAMPLE 37.

The NOV37 clone was analyzed, and the nucleotide and predicted polypeptide
5 sequences are shown in Table 37A.

Table 37A. NOV37 Sequence Analysis		
	SEQ ID NO: 165	5285 bp
NOV37a, CG59819-01 DNA Sequence	GGCCGGGGGAGGGGGCCGGACCGCGCGCGACCGGTGCGGCCCCGCTGGGGCCCCGCGATG GCGGGGGCCTGGCTCAGGTGGGGGCTCCTGCTCTGGGCAGGGCTCCTCGCGTCTCGG CGCACGGCCGGCTGCGGAGGATCACCTACGTGGTGCACCCGGGCCCCGGCCTGGCAGC CGGCGCCTTGCCCTGAGCGGGCCCCCGCGTTTCGCGGACATTCAACGTCGCGCTCAAC GCCAGGTACAGCCGAGCTCGGCGGCTGCCGGCGCCCCAGCCGTGCCTCCCCGGGG TCCCCCTCGGAGAGGACCCGGCGCACGAGCAAGCCGGGCGGCGCGGCCCTGCAGGGGCT CAGACCGCCGCGCCGCGCCGCGGAGCCTGCGCGTCCCGCGGTCCCGGGCGGGCAG CTCCACCCCAATCCCGGCGGCCACCCGGCAGCCGCCCGTTACCAAACAAGGCAGGC AAGTTGTGCGCTCCAAGGTGCCGAGGAGACCCAGAGCGGCGGAGGCTCTAGGCTGCA GGTTCACCAGAAGCAGCAGCTGCAGGGGGTCAATGTCTGTGGAGGGCGGTGCTGTCAT GGCTGGAGTAAGGCCCTGGCTCCCAGAGGTGCACCAAACCTAGCTGTGTTCCGCCAT GTCAGAATGGAGGGATGTGTCTCCGGCCACAACCTGTGTGTGTAAACCAGGGACCAA GGGCAAAGCCTGTGAAACAATAGCTGCCAGGACACCTCGTCACCAGTCTTTGGAGGG CAGAGTCCTGGGGCTGCTTCCTCGTGGGGCCCTCCTGAGCAAGCAGCAAAGCATACTT CATCTAAGAAGGCAGACACTCTACCAAGAGTCAGCCCTGTGGCCAGATGACCTTAAC CCTCAAGCCGAAGCCTTCAGTGGGACTCCCCAGCAGATACATTCTCAAGTGACTCCT CTTTCTTCCAGAGCGTGGTTATTCACCATGGCCAGACCCAGGAATACGTGCTCAAGC CCAAGTACTTTCCAGCCCAGAAGGGGATTTAGGAGAACAGTCCACTGAAGGTTCTTT CCCTTTAAGATATGTGCAGGATCAAGTTGCGGCACCTTTTCAGCTGAGTAACCACACT GGCCGCATCAAGGTGGTCTTTACTCCGAGCATCTGTAAAGTGACCTGCACCAAGGGCA GCTGTGAGAACAGCTGTGAGAAGGGGAACACCACCACTCTCATTAGTGAGAATGGTCA TGCTGCCGACACCCTGACGGCCACGAACCTCCGAGTGGTAATTTGCCATCTTCCATGT ATGAATGGTGGCCAGTGCAAGTTCAAGGGACAAATGTCAGTGCCCTCAAATTTACAG GAAACTTTGTGAGATCCAGTCCATGGTGCCAGCGTGCCTAAACTTTATCAGCATTC CCAGCAGCCAGGCAAGGCGTTGGGGACGCATGTCATCCATTCAACACATACCTTGCCT CTGACCGTGACTAGCCAGCAAGGAGTCAAAGTGAAATTTCTCCTAACATAGTCAATA TCCATGTGAAACATCCTCCTGAAGCTTCCGTCCAGATACATCAGGTTTCAAGAATTGA	

TGGCCCAACAGGCCAGAAGACAAAAGAAGCTCAACCAGGCCAATCCCAAGTCTCGTAC
 CAAGGGCTTCCTGTCCAGAAGACCCAGACCATAACATCCACATACTCCCACCAGCAGG
 TCATTCTCAGTCTACCCCGTGGCTGCTAAGACACAGCTTGGCCGGTGCTTCCAGGA
 AACCATTGGGTACAGTGTGGCAAAGCGCTCCCTGGCCTTTCAAAGCAAGAGGACTGC
 TGTGGAAGTGTGGGTACCTCCTGGGGCTTTAACAAATGCCAGAAATGCCCCAAGAAAC
 CATCTTATCATGGATACAACCAAATGATGGAATGCCTACCGGGTTATAAGCGGGTTAA
 CAACACCTTTTGCCAAGATATTAATGAATGTCAGCTACAAGGTGTATGCCCTAATGGT
 GAGTGTTTGAATACCATGGGCAGCTATCGATGTACCTGCAAAATAGGATTTGGGCCGG
 ATCCTACCTTTTCAAGTTGTGTTCTGTATCCCCCTGTGATCTCGGAAGAGAAAGGGCC
 CTGTTACCGACTTGTCTAGTTCTGGAAGACAGTGTATGTACCCTCTGTCTGTTCACCTC
 ACCAAGCAGCTCTGCTGTTGTAGTGTGGGCAAGGCCTGGGGCCCACACTGTGAGAAAT
 GTCCCCTTCCAGGCACAGCTGCTTTTAAGGAAATCTGTCTGGTGAATGGGTTATAC
 GGTTCCTGGCGTTCATAGACGCAGGCCAATCCATCACCATGTAGGTAAAGGACCTGTA
 TTTGTCAAGCCAAAGAACACTCAACCTGTTGCTAAAAGTACTCATCCTCCACCTCTCC
 CAGCCAAGGAAGAGCCAGTGGAGGCCCTGACCTTCTCCGGGAACACGGGCCAGGAGT
 GGCGGAGCCAGAAGTGGCAACTGCACCCCTGAAAAGGAAATACCTTCATTGGATCAA
 GAGAAAACCAAACCTTGAGCCTGGTCAACCCAGCTGTCTCCAGGCATTTCCGCTATTC
 ATCTGCATCCACAGTTTCCAGTAGTGATTGAAAAACATCACCTCCTGTGCCTGTTGA
 AGTAGCTCCTGAAGCTTCTACGTCTAGTGCCAGCCAAGTGATTGCTCCTACTCAAGTG
 ACAGAAATCAATGAATGTACTGTGAACCCTGATATCTGTGGAGCAGGACACTGCATTA
 ACCTACCAGTGAGATATACCTGTATATGCTACGAGGGCTACAGGTTCAAGTGAACAACA
 GAGGAAATGTGTGTATATTGATGAGTGTACTCAGGTCCAACACCTCTGCTCCAGGGC
 CGCTGTGAAAACACCGAGGGAAGTTTCTTGTGCATTTGCCCAGCAGGATTTATGGCCA
 GTGAGGAGGGTACTAACTGCATAGATGTTGACGAATGCCTGAGGCCGGACGTCTGTGG
 GGAGGGGCACTGTGTCAATACTGTGGGGCCTTCCGGTGTGAATACTGTGACAGCGGG
 TACCGCATGACTCAGAGAGGCCGTTGTGAGGATATTGATGAATGTTTGAATCCAAGCA
 CTTGTCCAGATGAGCAGTGTGTGAATTCTCCTGGATCTTACCAGTGCGTTCCTGAC
 AGAAGGATTCCGAGGCTGGAATGGACAGTGCCTTGATGTGGACGAGTGCCTGGAACCA
 AACGTCTGCGCAAATGGTGATTGTTCCAACCTTGAAGGCTCCTACATGTGTTTATGCC
 ACAAGGCTATACCCGGACTCCGGACCACAAGCACTGTAGAGATATTGATGAATGTCA
 GCAAGGGAATCTATGTGTAAACGGGCAGTGCAAAAATACCGAGGGCTCCTTCAGGTGC
 ACCTGTGGACAGGGGTACCAGCTGTGCGCAGCTAAAGACCAGTGTGAAGACATTGATG
 AATGCCAGCACCGTCATCTCTGTGCTCATGGGCAGTGCAGGAACACTGAGGGGCTCTTT
 TCAATGTGTGTGTGACCAGGGTTACAGAGCATCTGGGCTTGGAGACCACTGTGAAGAT
 ATCAATGAATGCTTGGAGGACAAGAGTGTGTTGCCAGAGAGGAGACTGCATTAATACTG
 CAGGGTCCTATGATTGTACTTGTCCGGATGGATTTTCAAGTAGATGACAATAAAACATG
 TCAAGATATTAATGAATGTGAACATCCAGGGCTCTGTGGTCCGCAAGGGGAGTGCCTA
 AACACAGAGGGTTCTTTCCATTGTGTCTGCCAGCAGGGTTTCTCAATCTCTGCAGATG
 GCCGTACGTGTGAAGATATTGATGAATGTGTAAACAACACTGTTTGTGACAGTCACGG
 GTTTTGTGACAATAACAGCTGGCTCCTTCCGCTGCCTCTGTTATCAGGGCTTTCAAGCC
 CCACAGGATGGGCAAGGGTGTGTGGATGTGAATGAATGTGAAGTGTGCTCAGTGGGGTGT
 GTGGTGAAGCCTTCTGTGAAAACGTGGAAGGGTCTTCTGTGCGTGTGTGCTGATGA
 AAACCAAGAGTACAGCCCCATGACTGGGCAGTGCCGCTCCCGGACCTCCACAGATTTA
 GATGTAGATGTAGATCAACCCAAAGAAGAAAGAAAGAAATGCTACTATAATCTCAATG
 ACGCCAGTCTCTGTGATAATGTGTTGGCCCCCAATGTCACGAAACAAGAATGCTGCTG
 TACATCAGGCGTGGGATGGGGAGATAACTGCGAAATCTTCCCCTGCCCGGTCTTGGGA
 ACTGCTGAGTTCACTGAAATGTGTCCAAAGGGAAAGGTTTTGTGCCTGCTGGAGAAT
 CATCTTCTGAAGCTGGTGGTGAGAACTATAAAGATGCAGATGAATGCCTACTTTTTGG
 ACAAGAAATCTGCAAAAATGGTTTTCTGTTTGAACACTCGGCCTGGGTATGAATGCTAC
 TGTAAGCAAGGGACGTACTATGATCCTGTGAAACTGCAGTGCTTTGATATGGATGAAT
 GTCAAGACCCCAAGTAGTTGTATTGATGGCCAGTGTGTTAATACAGAGGGCTCTTACAA
 CTGCTTCTGTACTACCCCATGGTCCTGGATGCGTCAGAAAAAAGATGTATACGACCG
 GCTGAGTCAAACGAACAAATAGAAGAACTGATGTCTACCAAGATTTGTGCTGGGAAC
 ATCTGAGTGATGAATACGTGTGTAGCCGGCCTCTTGTGGGCAAGCAGACAACGTACAC
 TGAGTGCTGCTGTCTGTATGGAGAGGCCTGGGCGATGCAGTGTGCCCTCTGCCCCCTG
 AAGGATTCAGATGACTATGCTCAGCTGTGTAACATCCCGTGACGGGACGCCGGCAGC
 CATATGGACGGGACGCCTTGGTTGACTTCAGTGAACAGTATACTCCAGAAGCCGATCC
 CTACTTCATCCAAGACCGTTTTCTAAATAGCTTTGAGGAGTTACAGGCTGAGGAATGC
 GGCATCCTCAATGGATGTGAAAATGGTCGCTGTGTGAGGGTCCAGGAAGGTTACACCT
 GCGATTGCTTTGATGGGTATCACTTGGATACTGCCAAGATGACCTGTTTCGATGTAAA
 TGAATGCGATGAGTTGAACAACCGGATGTCTCTCTGCAAGAATGCCAAGTGCATTAAC

	ACCGATGGTTCCTACAAGTGTGTTGTGTCTGCCAGGCTACGTGCCTTCTGACAAGCCAA ACTACTGCACTCCGTTGAATACCGCCTTGAATTTAGAGAAAGACAGTGACCTGGAGTG AAACAGAATCTACATAACCTAAGCCCATATACTCTGCACTGTGTAAAGGAAAAGGGAG AAATGTA		
	ORF Start: ATG at 56	ORF Stop: TGA at 5219	
	SEQ ID NO: 166	1721 aa	MW at 186900.6kD
NOV37a, CG59819-01 Protein Sequence	MAGAWLRWGLLLWAGLLASSAHGRLRRITYVVHPPGLAAGALPLSGPPRSRTFNVAL NARYSRSSAAAGAPSRASPGVPSERTRRTSKPGGAALQGLRPPPPPPPEPARPAVPGG QLHPNPGGHPAAAPFTKQGRQVVRSKVPQETQSGGSRQLQVHQKQQLQGVNVCGRCC HGWSKAPGSQRCTKPSCVPPCQNGGMCLRPQLCVCKPGTKGKACETIAAQDTSSPVFG GQSPGAASSWGPPEQAAKHTSSKKADTLPRVSPVAQMTLTLKPKPSVGLPQQIHSQVT PLSSQSVVIHHGQTQEYVLKPKYFPAQKGISGEQSTEGSFPLRYVDQVAAPFQLSNH TGRIKVVFTPSICKVTCTKGSCQNSCEKGNTTTLISENGHAADTLTATNFRVVI CHLP CMNGGQCSSRDKCQCPPNFTGKLCQIPVHGASVPKLYQHSQQPGKALGTHVIHSTHTL PLTVTSQQGVKVKFPPNIVNIHVKHPPPEASVQIHQVSRIDGPTGQKTKEAQPGQSQVS YQGLPVQKTQTIHSTYSHQQVI PHVYPVAAKTQLGRCFQETIGSQCGKALPGLSKQED CCGTVGTSWGFNKCQKCPKPSYHGYNQMMELPGYKRVNNTFCQDINECQLQGVCPN GECLNTMGSYRCTCKIGFGPDPTFSSCVPDPPISEEKGPCYRLVSSGRQCMYPLSVH LTKQLCCCSVGKAWGPHCEKCPLPGTAAFKEICPGGMGYTVSGVHRRRPIHHHVKGKP VFVKPKNTQPVAKSTHPPPLPAKEEPVEALTF SREHGPVGAEP EVATAPPEKEI PSLD QEKTKLEPGQPQLSPGISAIHLHPQFPVVEIKTSPPVPVEVAPEASTSSASQVIAPTQ VTEINECTVNPDICGAGHCINLPVRYTCICYEGYRFSEQQRKCVYIDECTQVQHLC SQ GRCENTEGSFLCICPAGFMASEEGTNCIDVDECLRPDVCGEHCVNTVGAFRC EYCD S GYRMTQRGRCEDIDECLNPSTCPDEQCVNSPGSYQCVPCTEGFRGWNGQCLDVDECL E PNVCANGDCSNLEGSYMC SCHKGYTRTPDHKHCRDIDECQQGNLCVNGQCKNTEGSFR CTCGQGYQLSAAKDQCEDIDECQHRHLCAHGQCRNTEGSFQCVCDQGYRASGLGDHCE DINECLEDKSVCQRGDCINTAGSYDCTCPDGFQLDDNKTCQDINECEHPGLCGPQGE C LNTEGSFHCVCQQGFSISADGRTCEDIDECVNNTVCD SHGFCDNTAGSFRCLCYQGFQ APQDGGQCVDVNECELLSGVCGEAFCENVEGSFLCVCADENQEYSPMTGQCRSRTSTD LDVDVDQPKEEKKECYYNLNDASLCDNVLPVNTKQECCCTSGVGWGDNCEIFPCPVL GTAEFTEMCPKGKGFVPAGESSEAGGENYKDADECLLFGQEICKNGFCLNTRPGYEC YCKQGTYYDPVKLQCFDMDECQDPSSCIDGQCVNTEGSYNCFCTHPMVLDASEKRCIR PAESNEQIEETDVYQDLCWEHLSDEYVCSRPLVGKQTTYTECCCLYGEAWAMQCALCP LKDSDDYAQLCNI PVTGRRQPYGRDALVDFSEQYTPEADPYFIQDRFLNSFEELQAE E CGILNGCENGRCVRVQEGYTCDCFDGYHLDTAKMTCFDVNECDELNNRMSLCKNAKCI NTDGSYKCLCLPGYVPSDKPNYCTPLNTALNLEKDSLE		
	SEQ ID NO: 167	5126 bp	
NOV37b, CG59819-02 DNA Sequence	GGCCGGGGGAGGGGGCCGGACCGCGCGCGACCGGTGCGGCCCGCTGGGGCCCGCGATG GCGGGGGCCTGGCTCAGGTGGGGGCTCCTGCTCTGGGCAGGGCTCCTCGCGTCTCGG CGCACGGCCGGCTGCGGAGGATCACCTACGTGGTGCACCCGGGCCCCGGCCTGGCAGC CGGCGCCTTGCCCCCTGAGCGGGCCCCCGCGTTGCGGGACATTCAACGTGCGGCTCAAC GCCAGGTACAGCCGAGCTCGGCGGCTGCCGGCGCCCCAGCCGTGCCTCCCCGGGG TCCCCCTCGGAGAGGACCCGGCGCACGAGCAAGCCGGGCGGCGGGCCCTGCAGGGGCT CAGACCGCCGCCGCCGCCGCCGCCGAGCCTGCGCGTCCCGCGGTCCCCGGCGGGCAG CTCCACCCCAATCCCGGCGGCCACCCGGCAGCCGCCCCGTTACCAAACAAGGCAGGC AAGTTGTGCGCTCCAAGGTGCCGAGGAGACCCAGAGCGGCGGAGGCTCTAGGCTGCA GGTTCACCAGAAGCAGCAGCTGCAGGGGGTCAATGTCTGTGGAGGGCGGTGCTGTCAT GGCTGGAGTAAGGCCCCCTGGCTCCCAGAGGTGCACCAAACCTAGCTGTGTTCCGCCAT GTCAGAATGGAGGGATGTGTCTCCGGCCACAACCTCTGTGTGTGTAAACCAGGGACCAA GGGCAAAGCCTGTGAAACAATAGCTGCCAGGACACCTCGTCACCACTCTTTGGAGGG CAGAGTCCTGGGGCTGCTTCCTCGTGGGGCCCTCCTGAGCAAGCAGCAAAGCATACTT CATCTAAGAAGGCAGACACTCTACCAAGAGTCAGCCCTGTGGCCAGATGACCTTAAC CCTCAAGCCGAAGCCTTCAGTGGGACTCCCCAGCAGATACATTCTCAAGTGACTCCT CTTTCTTCCCAGAGCGTGGTTATTACCATGGCCAGACCCAGGAATACGTGCTCAAGC CCAAGTACTTTCCAGCCCAGAAGGGGATTTAGGAGAACAGTCCACTGAAGGTTCTTT CCCTTTAAGATATGTGCAGGATCAAGTTGCGGCACCTTTTCAGCTGAGTAACCACACT GGCCGCATCAAGGTGGTCTTTACTCCGAGCATCTGTAAAGTGACCTGCACCAAGGGCA GCTGTGAGAACAGCTGTGAGAAGGGGAACACCACCACTCTCATTAGTGAGAATGGTCA		

TGCTGCCGACACCCTGACGGCCACGAACTTCCGAGTGGAATTTGCCATCTTCCATGT
ATGAATGGTGGCCAGTGCAGTTCAAGGGACAAATGTCAGTGCCCTCCAAATTTACACAG
GAAAACTTTGTGAGATCCAGTCCATGGTGCCAGCGTGCCATAAACTTTATCAGCATTC
CCAGCAGCCAGGCAAGGCGTTGGGGACGCATGTCATCCATTCAACACATACCTTGCCT
CTGACCGTGACTAGCCAGCAAGGAGTCAAAGTGAAATTTCTCCTAACATAGTCAATA
TCCATGTGAAACATCCTCCTGAAGCTTCCGTCCAGATACATCAGGTTTCAAGAATTGA
TGGCCCAACAGGCCAGAAGACAAAAGAAGCTCAACCAGGCCAATCCCAAGTCTCGTAC
CAAGGGCTTCCTGTCCAGAAGACCCAGACCATACTCCACATACTCCCACCAGCAGG
TCATTCTCAGTCTACCCCGTGGCTGCTAAGACACAGCTTGGCCGGTGCTTCCAGGA
AACCATTGGGTACAGTGTGGCAAAGCGCTCCCTGGCCTTTCAAAGCAAGAGGACTGC
TGTGGAACGTGGGTACCTCCTGGGGCTTTAAACAAATGCCAGAAATGCCCCAAGAAAC
CATCTTATCATGGATACAACCAATGATGGAATGCCTACCGGGTTATAAGCGGGTTAA
CAACACCTTTTGCCAAGATATTAATGAATGTCAGCTACAAGGTGTATGCCCTAATGGT
GAGTGTGTTGAATACCATGGGCAGCTATCGATGTACCTGCAAAATAGGATTTGGGCCGG
ATCCTACCTTTTCAAGTTGTGTTCTGATCCCCCTGTGATCTCGGAAGAGAAAGGGCC
CTGTTACCGACTTGTGAGTCTGGAAGACAGTGTATGCACCCTCTGTCTGTTACCTC
ACCAAGCAGCTCTGCTGTTGTAGTGTGGGCAAGGCTGGGGCCCACACTGTGAGAAAT
GTCCCCTTCCAGGCACAGCCAAGGAAGAGCCAGTGGAGGCCCTGACCTTCTCCCGGA
ACACGGGCCAGGAGTGGCGGAGCCAGAAGTGGCAACTGCACCCCTGAAAAGGAAATA
CCTTCATTGGATCAAGAGAAAACCAACTTGAGCCTGGTCAACCCAGCTGTCTCCAG
GCATTTCCGCTATTATCTGCATCCACAGTTTCCAGTAGTGATTGAAAAACATCACC
TCCTGTGCCTGTTGAAGTAGCTCCTGAAGCTTCTACGTCTAGTGCCAGCCAAGTGATT
GCTCCTACTCAAGTGACAGAAATCAATGAATGTACTGTGAACCCTGATATCTGTGGAG
CAGGACACTGCATTAACCTACCAGTGAGATATACCTGTATATGCTACGAGGGCTACAG
GTTTCAGTGAACAACAGAGGAAATGTGTGTATATTGATGAGTGTACTCAGGTCCAACAC
CTCTGCTCCCAGGGCCGCTGTGAAAACACCGAGGGAAGTTTCTTGTGCATTTGCCAG
CAGGATTTATGGCCAGTGAGGAGGGTACTAACTGCATAGATGTTGACGAATGCCTGAG
GCCGGACGTCTGTGGGGAGGGGCACTGTGTCAATACTGTGGGGGCCTTCCGGTGTGAA
TACTGTGACAGCGGGTACCGCATGACTCAGAGAGGCCGTTGTGAGGATATTGATGAAT
GTTTGAATCCAAGCACTTGTCCAGATGAGCAGTGTGTGAATTCTCCTGGATCTTACCA
GTGCGTTCCCTGCACAGAAGGATTCCGAGGCTGGAATGGACAGTGCCTTGATGTGGAC
GAGTGCCTGGAACCAACGTCTGCGCAAATGGTGATTGTTCCAACCTGAAGGCTCCT
ACATGTGTTTATGCCACAAAGGCTATACCCGGACTCCGGACCACAAGCACTGTAGAGA
TATTGATGAATGTCAGCAAGGGAATCTATGTGTAAACGGGCAGTGCAAAAATACCGAG
GGCTCCTTCAGGTGCACCTGTGGACAGGGGTACCAGCTGTGCGCAGCTAAAGACCAGT
GTGAAGACATTGATGAATGCCAGCACCGTCACTCTCTGTGCTCATGGGCAGTGCAGGAA
CACTGAGGGCTCTTTTCAATGTGTGTGTGACCAGGGTTACAGAGCATCTGGGCTTGA
GACCACTGTGAAGATATCAATGAATGCTTGGAGGACAAGAGTGTGTTGCCAGAGAGGAG
ACTGCATTAATACTGCAGGGTCTATGATTGTACTTGTCCGGATGGATTTACAGCTAGA
TGACAATAAAACATGTCAAGATATTAATGAATGTGAACATCCAGGGCTCTGTGGTCCG
CAAGGGGAGTGCCTAAACACAGAGGGTTCTTTCCATTGTGTCTGCCAGCAGGGTTTCT
CAATCTCTGCAGATGGCCGTACGTGTGAAGATATTGATGAATGTGTAAACAACACTGT
TTGTGACAGTCAAGGGTTTTGTGACAATACAGCTGGCTCCTTCCGCTGCCTCTGTTAT
CAGGGCTTTCAAGCCCCACAGGATGGGCAAGGGTGTGTGGATGTGAATGAATGTGAAC
TGCTCAGTGGGGTGTGTGGTGAAGCCTTCTGTGAAAACGTGGAAGGGTCTTCTCTGTG
CGTGTGTGCTGATGAAAACCAAGAGTACAGCCCCATGACTGGGCAGTGCCGCTCCCGG
ACCTCCACAGATTTAGATGTAGATGTAGATCAACCCAAAGAAGAAAAGAAAGAAATGCT
ACTATAATCTCAATGACGCCAGTCTCTGTGATAATGTGTTGGCCCCCAATGTACGAA
ACAAGAATGCTGCTGTACATCAGGCGTGGGATGGGGAGATAACTGCGAAATCTTCCCC
TGCCCGGTCTTGGGAACTGCTGAGTTCACTGAAATGTGTCCCAAAGGGAAAGGTTTTG
TGCTGCTGGAGAATCATCTTCTGAAGCTGGTGGTGAGAACTATAAAGATGCAGATGA
ATGCCTACTTTTTGGACAAGAAATCTGCAAAAATGGTTTTCTGTTTGAACACTCGGCCT
GGGTATGAATGCTACTGTAAGCAAGGGACGTACTATGATCCTGTGAACTGCAGTGCT
TTGATATGGATGAATGTCAAGACCCAGTAGTTGTATTGATGGCCAGTGTGTTAATAC
AGAGGGCTCTTACAACCTGCTTCTGTACTCACCCCATGGTCCTGGATGCGTCAGAAAAA
AGATGTATACGACCGGCTGAGTCAAACGAACAAATAGAAGAACTGATGTCTACCAAG
ATTTGTGCTGGGAACATCTGAGTGTGAATACGTGTGTAGCCGGCCTCTTGTGGGCAA
GCAGACAACGTACACTGAGTGCTGCTGTCTGTATGGAGAGGCCTGGGCGATGCAGTGT
GCCCTCTGCCCCCTGAAGGATTGAGTACTATGCTCAGCTGTGTAACATCCCCGTGA
CGGGACGCCGGCAGCCATATGGACGGGACGCCTTGGTTGACTTCAGTGAACAGTATAC
TCCAGAAGCCGATCCCTACTTCATCCAAGACCGTTTTCTAAATAGCTTTGAGGAGTTA

	CAGGCTGAGGAATGCGGCATCCTCAATGGATGTGAAAATGGTCGCTGTGTGAGGGTCC AGGAAGGTTACACCTGCGATTGCTTTGATGGGTATCACTTGGATACTGCCAAGATGAC CTGTTTCGATGTAAATGAATGCGATGAGTTGAACAACCGGATGTCTCTCTGCAAGAAT GCCAAGTGCATTAACACCGATGGTTCCTACAAGTGTGTGTCTGCCAGGCTACGTGC CTTCTGACAAGCCAACTACTGCACTCCGTTGAATACCGCCTTGAATTTAGAGAAAGA CAGTGACCTGGAGTGAAACAGAATCTACATAACCTAAGCCCATATACTCTGCACTGTG TAAAGGAAAAGGGAGAAATGTA		
	ORF Start: ATG at 56	ORF Stop: TGA at 5060	
	SEQ ID NO: 168	1668 aa	MW at 181174.9kD
NOV37b, CG59819-02 Protein Sequence	MAGAWLRWGLLLWAGLLASSAHGRLRRITYVVHPGPGLAAGALPLSGPPRSRTFNVAL NARYSRSSAAAGAPSRASPGVPSERTRRTSKPGGAALQGLRPPPPPPPEPARPAVPGG QLHPNPGGHPAAAPFTKQGRQVVRSKVPQETQSGGSRQLQVHQKQQLQGVNVCGRCC HGWSKAPGSQRCTKPSVPPCQNGMCLRPQLCVCKPGTKGKACETIAAQDTSSPVFG GQSPGAASSWGPPEQAAKHTSSKKADTLPRVSPVAQMTLTLKPKPSVGLPQQIHSQVT PLSSQSVVIHHGQTQEYVLKPKYFPAQKGISGEQSTEGSFPLRYVDQVAAPFQLSNH TGRIKVVFTPSICKVTCTKGSCQNSCEKGNNTTLISENGHAADTLTATNFRVVIChLP CMNGGQCSSRDKCQCPPNFTGKLCQIPVHGASVPKLYQHSQQPGKALGTHVIHSTHTL PLTVTSQQGVKVKFPPNIVNIHVKHPPPEASVQIHQVSRIDGPTGQKTKEAQPGQSQVS YQGLPVQKTQTIHSTYSHQQVIPHVYPVAAKTQLGRCFQETIGSQCGKALPGLSKQED CCGTVGTSWGFNKCQKCPKKPSYHGYNQMMELPGYKRVNNTFCQDINECQLQGVCPN GECLNTMGSYRCTCKIGFGPDPTFSSCVPDPPIVSEEKGPCYRLVSSGRQCMHPLSVH LTKQLCCCSVGKAWGPHCEKCPPLPGTAKEEPVEALTFSREHGPVGAPEPEVATAPPEKE IPSLDQEKTKLEPGQPQLSPGISAIHLHPQFPVVEIKTSPPVPVEVAPEASTSSASQV IAPTQVTEINECTVNPDICGAGHCINLPVRYTCICYEGYRFSEQQRKCVYIDECTQVQ HLCSQGRCENTEGSFLCICPAGFMASEEGTNCIDVDECLRPDVCGEGHCVNTVGAFCR EYCDSGYRMTQRGRCEDIDECLNPSTCPDEQCVNSPGSYQCVPCTEGFRGWNQCLDV DECLPNVCANGDCSNLEGSYMCSCCHKGYTRTPDHKHCRIDECQQGNLCVNGQCKNT EGSFRCTCGQGYQLSAAKDQCEDIDECQHRHLCAHGQCRNTEGSFQCVCDDQGYRASGL GDHCEDINECLEDKSVQCRGDCINTAGSYDCTCPDGFQLDDNKTCQDINECEHPGLCG PQGECLNTEGSFHCVCQQGFSISADGRTCEDIDECVNNTVCDSHGFCDNTAGSFRCLC YQGFQAPQDQGQGVNNECELLSGVCGEAFCEVNEGSFLCVCADENQEYSPMTGQCRS RTSTDLDVDVDQPKKEKKECYYNLNDASLCDNVLPVNTKQECCTSGVGWGDNCEIF PCPVLGTAEFTMCPKKGFPAGESSEAGGENYKDADECLLFGQEICKNGFCLNTR PGYECYCKQGTYYDPVKLQCFDMDECQDPSSCIDGQCVNTEGSYNCFCTHPMVLDAE KRCIRPAESNEQIEETDVYQDLCEHLSDYVCSRPLVGKQTTYTECCCLYGEAWAMQ CALCPLKDSDDYAQLCNIPTVGRRQPYGRDALVDFSEQYTPEADPYFIQDRFLNSFEE LQAECEGILNGCENGRCVRVQEGYTCDCFDGYHLD TAKMTCFDVNECDELNNRMSLCK NAKCINTDGSYKCLCLPGYVPSDKPNYCTPLNTALNLEKDSLE		
	SEQ ID NO: 169	6074 bp	
NOV37c, CG59819-03 DNA Sequence	GGCCGGGGGAGGGGGCCGGACCGCGCGCGACCGGTGCGGCCCGCTGGGGCCCGCGATG GCGGGGGCCTGGCTCAGGTGGGGGCTCCTGCTCTGGGCAGGGCTCCTCGCGTCTCTCGG CGCACGGCCGGCTGCGGAGGATCACCTACGTGGTGCACCCGGGCCCCGGCCTGGCAGC CGGCGCCTTGCCCCGTGAGCGGGCCCCCGCGTTGCGGGACATTCAACGTGCGGCTCAAC GCCAGGTACAGCCGAGCTCGGCGGCTGCCGGCGCCCCAGCCGTGCCTCCCCGGGG TCCCCCTCGGAGAGGACCCGGCGCACGAGCAAGCCGGGCGGCGCGGCCCTGCAGGGGCT CAGACCGCCGCCGCCGCCGCCGCCGAGCCTGCGCGTCCCGCGGTCCCCGGCGGGCAG CTCCACCCCAATCCCGGGCGGCCACCCGGCAGCCGCCCGTTACCAAACAAGGCAGGC AAGTTGTGCGCTCCAAGGTGCCGAGGAGACCCAGAGCGGCGGAGGCTCTAGGCTGCA GGTTCACCAGAAGCAGCAGCTGCAGGGGGTCAATGTCTGTGGAGGGCGGTGCTGTCAT GGCTGGAGTAAGGCCCCCTGGCTCCCAGAGGTGCACCAAACCTAGCTGTGTTCCGCCAT GTCAGAATGGAGGGATGTGTCTCCGGCCACAACCTGTGTGTGTAAACCAGGGACCAA GGGCAAAGCCTGTGAAACAATAGCTGCCCAGGACACCTCGTCACCAGTCTTTGGAGGG CAGAGTCCTGGGGCTGCTTCCTCGTGGGGCCCTCCTGAGCAAGCAGCAAAGCATACTT CATCTAAGAAGGCAGACACTCTACCAAGAGTCAGCCCTGTGGCCAGATGACCTTAAC CCTCAAGCCGAAGCCTTCAGTGGGACTCCCCCAGCAGATACATTCTCAAGTGA CTCT CTTTCTTCCCAGAGCGTGGTTATTACCATGGCCAGACCCAGGAATACGTGCTCAAGC CCAAGTACTTTCCAGCCCAGAAGGGGATTTTCAAGAGAACAGTCCACTGAAGGTTCTTT CCCTTTAAGATATGTGCAGGATCAAGTTGCGGCACCTTTTCACTGAGTAACCACACT		

GGCCGCATCAAGGTGGTCTTTACTCCGAGCATCTGTAAAGTGACCTGCACCAAGGGCA
 GCTGTCAGAACAGCTGTGAGAAGGGGAACACCACCCTCTCATTAGTGAGAATGGTCA
 TGCTGCCGACACCCTGACGGCCACGAACTTCCGAGTGGTAATTTGCCATCTTCCATGT
 ATGAATGGTGGCCAGTGCAGTTCAAGGGACAAATGTCAGTGCCCTCCAAATTTACAG
 GAAACTTTGTCAGATCCCAGTCCATGGTGCCAGCGTGCCTAAACTTTATCAGCATTC
 CCAGCAGCCAGGCAAGGCGTTGGGGACGCATGTCATCCATTCAACACATACCTTGCCCT
 CTGACCGTGACTAGCCAGCAAGGAGTCAAAGTGAAATTTCCCTCCTAACATAGTCAATA
 TCCATGTGAAACATCCTCCTGAAGCTTCCGTCCAGATACATCAGGTTTCAAGAATTGA
 TGGCCCAACAGGCCAGAAGACAAAAGAAGCTCAACCAGGCCAATCCCAAGTCTCGTAC
 CAAGGGCTTCCCTGTCCAGAAGACCCAGACCATAATTCCACATACTCCCACCAGCAGG
 TCATTCCCTCACGTCTACCCCGTGGCTGCTAAGACACAGCTTGGCCGGTGCTTCCAGGA
 AACCATTGGGTACAGTGTGGCAAAGCGCTCCCTGGCCTTTCAAAGCAAGAGGACTGC
 TGTGGAAGTGTGGGTACCTCCTGGGGCTTTAACAAATGCCAGAAATGCCCCAAGAAAC
 CATCTTATCATGGATACAACCAATGATGGAATGCCTACCGGGTTATAAGCGGGTTAA
 CAACACCTTTTGCCAAGATATTAATGAATGTCAGCTACAAGGTGTATGCCCTAATGGT
 GAGTGTTTGAATACCATGGGCAGCTATCGATGTACCTGCAAAATAGGATTTGGGCCGG
 ATCCTACCTTTTCAAGTTGTGTTCCCTGATCCCCCTGTGATCTCGGAAGAGAAAGGGCC
 CTGTTACCGACTTGTCAAGTCTGGAAGACAGTGTATGTACCCTCTGTCTGTTACCTC
 ACCAAGCAGCTCTGCTGTTGTAGTGTGGGCAAGGCTGGGCCACACTGTGAGAAATGTC
 CCCTTCCAGGCACAGCTGCTTTTAAGGAAATCTGTCCTGGTGGAAATGGGTATACGGT
 TTCTGGCGTTCATAGACGCAGGCCAATCCATCACCATGTAGGTAAAGGACCTGTATTT
 GTCAAGCCAAAGAACACTCAACCTGTTGCTAAAAGTACTCATCCTCCACCTCTCCCAG
 CCAAGGAAGAGCCAGTGGAGGCCCTGACCTTCTCCCGGGAACACGGGGCCAGGAGTGC
 GGAGCCAGAAGTGGCAACTGCACCCCTGAAAAGGAAATACCTTCATTGGATCAAGAG
 AAAACCAAACCTTGAGCCTGGTCAACCCCAAGCTGTCTCCAGGCATTTCCGCTATTATC
 TGCATCCACAGTTTCCAGTAGTGATTGAAAAACATCACCTCCTGTGCCTGTTGAAGT
 AGCTCCTGAAGCTTCTACGTCTAGTGCCAGCCAAGTGATTGCTCCTACTCAAGTGACA
 GAAATCAATGAATGTACTGTGAACCCTGATATCTGTGGAGCAGGACACTGCATTAACC
 TACCAGTGAGATATACCTGTATATGCTACGAGGGCTACAGGTTCAAGTGAACAACAGAG
 GAAATGTGTGGATATTGATGAGTGTACTCAGGTCCAACACCTCTGCTCCCAGGGCCGC
 TGTGAAAACACCGAGGGAAGTTTCTTGTGCATTTGCCCAGCAGGATTTATGGCCAGTG
 AGGAGGGTACTAACTGCATAGATGTTGACGAATGCCTGAGGCCGGACGTCTGTGGGGA
 GGGGCACTGTGTCAATACTGTGGGGGCCCTTCCGGTGTGAATACTGTGACAGCGGGTAC
 CGCATGACTCAGAGAGGCCGTTGTGAGGATATTGATGAATGTTTGAATCCAAGCACTT
 GTCCAGATGAGCAGTGTGTGAATTCTCCTGGATCTTACCAGTGCGTTCCCTGCACAGA
 AGGATTCCGAGGCTGGAATGGACAGTGCCTTGATGTGGACGAGTGCCTGGAACCAAC
 GTCTGCGCAAATGGTGATTGTTCCAACCTTGAAGGCTCCTACATGTGTTTATGCCACA
 AAGGCTATACCCGACTCCGGACCACAAGCACTGTAGAGATATTGATGAATGTCAGCA
 AGGGAATCTATGTGTAAACGGGCAGTGCAAAAATACCGAGGGCTCCTTCAGGTGCACC
 TGTGGACAGGGGGTTACCAGCTGTGCGCAGCTAAAGACCAGTGTGAAGACATTGATG
 AATGCCAGCACCGTCATCTCTGTGCTCATGGGCAGTGCAGGAACACTGAGGGCTCTTT
 TCAATGTGTGTGTGACCAGGGTTACAGAGCATCTGGGCTTGGAGACCACTGTGAAGAT
 ATCAATGAATGCTTGGAGGACAAGAGTGTGTTGCCAGAGAGGAGACTGCATTAATACTG
 CAGGGTCCTATGATTGTACTTGTCCGGATGGATTTTCAAGTAGATGACAATAAAACATG
 TCAAGATATTAATGAATGTGAACATCCAGGGCTCTGTGGTCCACAAGGGGAGTGCCTA
 AACACAGAGGGTTCTTTCCATTGTGTCTGCCAGCAGGGTTTCTCAATCTCTGCAGATG
 GCCGTACGTGTGAAGATGTGAATGAATGTGAACCTGCTCAGTGGGGTGTGTGGTGAAGC
 CTTCTGTGAAAACGTGGAAGGGTCTTCCCTGTGCGTGTGTGCTGATGAAAACCAAGAG
 TACAGCCCCATGACTGGGCAGTGCCGCTCCCGGACCTCCACAGATTTAGATGTAGATG
 TAGATCAACCCAAAGAAGAAAAGAAAGAAATGCTACTATAATCTCAATGACGCCAGTCT
 CTGTGATAATGTGTTGGCCCCCAATGTCACGAAACAAGAATGCTGCTGTACATCAGGC
 GCGGGATGGGGAGATAACTGCGAAATCTTCCCTGCCCCGGTCTTGGGAACTGCTGAGT
 TCACTGAAATGTGTCCCAAAGGGAAAGGTTTTGTGCCTGCTGGAGAATCATCTTCTGA
 AGCTGGTGGTGAGAACTATAAAGATGCAGATGAATGCCTACTTTTTGGACAAGAAATC
 TGCAAAAATGGTTTCTGTTTGAACACTCGGCCTGGGTATGAATGCTACTGTAAGCAAG
 GGACGTACTATGATCCTGTGAACTGCAGTGCTTTGATATGGATGAATGTCAAGACCC
 CAGTAGTTGTATTGATGGCCAGTGTGTTAATACAGAGGGCTCTTACAACCTGCTTCTGT
 ACTCACCCCATGGTCCCTGGATGCGTCAGAAAAAGATGTATACGACCGGCTGAGTCAA
 ACGAACAAATAGAAGAACTGATGTCTACCAAGATTTGTGCTGGGAACATCTGAGTGA
 TGAATACGTGTGTAGCCGGCCTCTTGTGGGCAAGCAGACAACGTACACTGAGTGCTGC
 TGTCTGTATGGAGAGGCCTGGGGCATGCAGTGTGCCCTCTGCCCCCTGAAGGATTGAG

	ATGACTATGCTCAGCTGTGTAACATCCCCGTGACGGGACGCCGGCAGCCATATGGACG GGACGCCTTGTTGACTTCAGTGAACAGTATACTCCAGAAGCCGATCCCTACTTCATC CAAGACCGTTTTCTAAATAGCTTTGAGGAGTTACAGGCTGAGGAATGCGGCATCCTCA ATGGATGTGAAAATGGTCGCTGTGTGAGGGTCCAGGAAGGTTACACCTGCGATTGCTT TGATGGGTATCACTTGGATACGGCCAAGATGACCTGTGTGATGTAAATGAATGCGAT GAGTTGAACAACCGGATGTCTCTCTGCAAGAATGCCAAGTGCATTAACACCGATGGTT CCTACAAGTGTGTTGTGTCTGCCAGGCTACGTGCCTTCTGACAAGCCAACTACTGCAC TCCGTTGAATACCGCCTTGAATTTAGAGAAAGACAGTGACCTGGAGTGAACAGAATC TACATAACCTAAGCCCATATACTCTGCACTGTGTAAAGGAAAAGGGAGAAATGTATTA TACTTGAGACATTGCACCTACCCCGGAAGGCTGGAAATACGGAAACAGCATGGAGTTG CAAGTCCTCTGAAGACAATGAGAGGATTTAGGATGAGCCCGATAGGTGTGGCAGACCA AATGGACATTTCTCTAAAAAACAGTATATATAGTCTGTTCATATGTAAATTTCAATG GAAGAGAGGTGGAACAGTGCTGTTATTTTAAACAGAAGGTTGTATTATTATGTTGTTT TGTTTTTTTACTATTGCTTGATTAAATTTGGCATTAAATAGTGGTGGAAATATTTTA TATAATTTTCATTTTTTGGTTGTGCAGTTCCTTGGCTACTGTTTTTCTTTTACTTCAG TTTTTTAAAAATCTCAAATGAAAAAGTCTTCGATACAATATTGTTAAGCTGTATTATA AGTATTGTTACACAGGGTTATGCAATTCCCGGCCTGGAGCATTTTTTGAAATTCAAATT GTCTGTCTGTGGAGCAGGCAGTGATTTTGTTCCAAACCTTTGTATACACATTTGGAG AAAAGTACTTTATATTTTCAGTGTTTGTCTGATTTTAATGTCCGTTCTTAGCCAAGC TGCTAGCAGGTGTTAATTGGATCCCTTTCCTTCACTGAAATGGAAGAGTTTATAAGCT TACGTTAGTATTGTAATATGTAAAGTAAGCCCAACAAAAATTTTTAAAAATTTGATGA TCCCCAATATATCTACCATTGTATGTTAAATAAATCACCATTTTTGTAGAAAAAATTC TACCTGAGAGTAATTGTCAATGAGTACATGTGTATAAGTTGTATCCCACTCTCCCCAC TTTTATCTTTTCCAGTGGTCTTCTGTTAATGTAGTGTCTTTTACAAGTTAATCATTA ATTTGTTAGATCTTGTTATGGGCTAAAAAATAAAAAAAAAAAAAA		
	ORF Start: ATG at 56	ORF Stop: TGA at 5093	
	SEQ ID NO: 170	1679 aa	MW at 182193.4kD
NOV37c, CG59819-03 Protein Sequence	MAGAWLRWGLLLWAGLLASSAHGRLRRITYVHPGPGLAAGALPLSGPPRSRTFNVAL NARYSRSSAAAGAPSRASPGVPSERTRRTSKPGGAALQGLRPPPPPPPEPARPAVPGG QLHPNPGGHPAAAPFTKQGRQVRSKVPQETQSGGSRQLQVHQKQQLQGVNVCGRCC HGWSKAPGSQRCTKPSVPPCQNGGMCLRPQLCVCKPGTKGKACETIAAQDTSSPVFG GQSPGAASSWGPPEQAAKHTSSKKADTLPRVSPVAQMTLTLKPKPSVGLPQQIHSQVT PLSSQSVVIHHGQTQEYVLKPKYFPAQKGISGEQSTEGSFPLRYVQDQVAAPFQLSNH TGRIKVVFTPSICKVTCTKGSCQNSCEKGNTTTTLISENGHAADTLTATNFRVVIHCLP CMNGGQCSSRDKCQCPPNFTGKLCQIPVHGASVPKLYQHSQQPGKALGTHVIHSTHTL PLTVTSQQGVKVKFPPNIVNIHVKHPPPEASVQIHQVSRIDGPTGQKTKEAQPQSQVS YQGLPVQKTQTIHSTYSHQQVIHVYPVAAKTQLGRCFQETIGSQCGKALPGLSKQED CCGTVGTSWGFNKCQKCPKPSYHGYNQMMELPGYKRVNNTFCQDINECQLQGVCPN GECLNTMGSYRCTCKIGFGPDPTFSSCVPDPVPISEEKGPCYRLVSSGRQCMYPLSVH LTKQLCCCSVGKAGPHCEKCPLPGTAAAFKEICPGMGYTVSGVHRRRPIHHHVKGKPV FVKPKNTQPVAKSTHPPPLPAKEEPVEALTFSREHGARSAEPEVATAPPEKEIPSLDQ EGTKLEPGQPQLSPGISAIHLHPQFPVIEKTSPPVPVEVAPEASTSSASQVIAPTQV TEINECTVNPDICGAGHCINLPVRYTCICYEGYRFSEQQRKCDIDECTQVQHLCSQG RCENTECSFLCICPAGFMASEEGTNCIDVDECLRPDVCGEHCVNTVGAFRCEYCDSG YRMTQRGRCEDIDECLNPSTCPDEQCVNSPGSYQCVPCTEGFRGWNQCLDVECLEP NVCANGDCSNLEGSYMCCHKGYTRTPDHKHCARDIDECQQGNLCVNGQCKNTEGSFRC TCGQGGYQLSAAKDQCEDIDECQHRHLCAHGQCRNTEGSFQCVCDQGYRASGLGDHCE DINECLEDKSVQRGDCINTAGSYDCTCPDGFQLDDNKTCQDINECEHPGLCGPQGEC LNTEGSFHCVCQQGFSISADGRTCEDVNECELLSGVCGEAFCEVNEGSFLCVCADENQ EYSPMTGQCRSRTSTDLDVDVDQPKKEKKECYYNLNDASLCDNVLPVNTKQECCTS GAGWGDNCEIFPCPVLGTAEFTMCPKGKGFVPAGESSEAGGENYKDADECLLFGQE ICKNGFCLNTRPGYECYCKQGTYYDPVKLQCFDMDECQDPSSCIDGQCVNTEGSYNCF CTHPMVLDASEKRCIRPAESNEQIEETDVYQDLWEHLSDEYVCSRPLVGKQTTYTEC CCLYGEAWGMQCALCPLKDSDDYAQLCNIPTVGRRQPYGRDALVDFSEQYTPADPYF IQDRFLNSFEELQAECEGILNGCENGRVVRVQEGYTCDCFDGYHLDATAKMTCDVDNEC DELNNRMSLCKNAKCINTDGSYKCLCLPGYVPSDKPNYCTPLNTALNLEKSDLE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 37B.

Table 37B. Comparison of NOV37a against NOV37b through NOV37c.		
Protein Sequence	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV37b	19..1721 19..1668	1561/1703 (91%) 1562/1703 (91%)
NOV37c	19..1721 19..1679	1565/1704 (91%) 1565/1704 (91%)

Further analysis of the NOV37a protein yielded the following properties shown in Table 37C.

Table 37C. Protein Sequence Properties NOV37a	
PSort analysis:	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 24 and 25

A search of the NOV37a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several

5 homologous proteins shown in Table 37D.

Table 37D. Geneseq Results for NOV37a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR22461	Masking protein high polymer unit precursor MPU-P - Rattus rattus, 1712 aa. [JP04066597-A, 02-MAR-1992]	1..1721 1..1712	1525/1721 (88%) 1603/1721 (92%)	0.0
AAR14584	TGF beta 1 binding protein encoded by clone BPA 13 - Homo sapiens, 1355 aa. [WO9113152-A, 05-SEP-1991]	342..1721 16..1355	1324/1380 (95%) 1326/1380 (95%)	0.0
AAR53089	Human masking protein subunit hMPU-P - Homo sapiens, 845 aa. [JP06092995-A, 05-APR-1994]	742..1586 1..845	841/845 (99%) 841/845 (99%)	0.0
AAR53086	Human masking protein subunit hMPU-1 - Homo sapiens, 756 aa. [JP06092995-A, 05-APR-1994]	832..1586 2..756	752/755 (99%) 752/755 (99%)	0.0
AAR53087				0.0

	hMPU-2 - Homo sapiens, 752 aa. [JP06092995-A, 05-APR-1994]	1..752	749/752 (99%)	
--	---	--------	---------------	--

In a BLAST search of public sequence databases, the NOV37a protein was found to have homology to the proteins shown in the BLASTP data in Table 37E.

Table 37E. Public BLASTP Results for NOV37a				
Protein Accession Number	Protein/Organism/Length	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q00918	Latent transforming growth factor beta binding protein 1 precursor (Transforming growth factor beta-1 binding protein 1) (TGF-beta1-BP- 1) (Transforming growth factor beta-1 masking protein, large subunit) - Rattus norvegicus (Rat), 1712 aa.	1..1721 1..1712	1536/1721 (89%) 1611/1721 (93%)	0.0
O88349	LATENT TGF BETA BINDING PROTEIN - Mus musculus (Mouse), 1713 aa.	1..1720 1..1712	1523/1721 (88%) 1603/1721 (92%)	0.0
P22064	Latent transforming growth factor beta binding protein 1 precursor (Transforming growth factor beta-1 binding protein 1) (TGF-beta1-BP- 1) - Homo sapiens (Human), 1394 aa.	342..1721 16..1394	1369/1380 (99%) 1370/1380 (99%)	0.0
O35806	LATENT TGF-BETA BINDING PROTEIN-2 LIKE PROTEIN - Rattus norvegicus (Rat), 1764 aa.	72..1705 75..1760	710/1748 (40%) 937/1748 (52%)	0.0
Q14767	LATENT TRANSFORMING GROWTH FACTOR-BETA-BINDING PROTEIN-2 (LTBP-2) - Homo sapiens (Human), 1821 aa.	74..1706 87..1818	693/1810 (38%) 919/1810 (50%)	0.0

PFam analysis predicts that the NOV37a protein contains the domains shown in the Table 37F.

Table 37F. Domain Analysis of NOV37a			
Pfam Domain	NOV37a Match Region	Identities/ Similarities for the Matched Region	Expect Value
EGF: domain 1 of 18	191..218	15/47 (32%) 19/47 (40%)	0.0056

EGF: domain 2 of 18	403..430	15/47 (32%) 23/47 (49%)	0.00014
wap: domain 1 of 1	385..433	12/57 (21%) 30/57 (53%)	9.3
TB: domain 1 of 4	566..609	15/48 (31%) 41/48 (85%)	6e-13
EGF: domain 3 of 18	630..665	14/47 (30%) 27/47 (57%)	1e-05
Keratin_B2: domain 1 of 1	578..717	40/180 (22%) 64/180 (36%)	1.5
TB: domain 2 of 4	687..728	25/47 (53%) 40/47 (85%)	1.1e-21
Arthro_defensin: domain 1 of 1	874..901	9/37 (24%) 18/37 (49%)	8.4
EGF: domain 4 of 18	877..913	15/47 (32%) 27/47 (57%)	1.8e-05
EGF: domain 5 of 18	919..955	15/47 (32%) 27/47 (57%)	6e-05
granulin: domain 1 of 2	942..957	6/16 (38%) 12/16 (75%)	0.57
EGF: domain 6 of 18	961..996	15/47 (32%) 22/47 (47%)	7.9
EGF: domain 7 of 18	1002..1036	13/47 (28%) 27/47 (57%)	50
EGF: domain 8 of 18	1042..1077	15/47 (32%) 24/47 (51%)	0.00066
EGF: domain 9 of 18	1083..1118	16/47 (34%) 30/47 (64%)	0.00019
EGF: domain 10 of 18	1124..1159	14/47 (30%) 28/47 (60%)	0.00026
EGF: domain 11 of 18	1165..1200	14/47 (30%) 26/47 (55%)	0.0071
EGF: domain 12 of 18	1206..1242	13/47 (28%) 27/47 (57%)	0.00073
granulin: domain 2 of 2	1226..1244	10/19 (53%) 15/19 (79%)	20
EGF: domain 13 of 18	1248..1284	13/47 (28%) 25/47 (53%)	0.00063
EGF: domain 14 of 18	1290..1327		0.0037

		27/47 (57%)	
TB: domain 3 of 4	1357..1400	24/47 (51%) 36/47 (77%)	2e-18
EGF: domain 15 of 18	1428..1465	14/47 (30%) 27/47 (57%)	0.014
EGF: domain 16 of 18	1471..1506	15/47 (32%) 29/47 (62%)	1.2e-05
TB: domain 4 of 4	1534..1576	18/47 (38%) 40/47 (85%)	8.6e-18
EGF: domain 17 of 18	1625..1660	16/47 (34%) 26/47 (55%)	0.0004
EGF: domain 18 of 18	1666..1705	16/49 (33%) 31/49 (63%)	5.8e-06

EXAMPLE 38.

The NOV38 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 38A.

Table 38A. NOV38 Sequence Analysis			
	SEQ ID NO: 171	1034 bp	
NOV38a, CG59685-01 DNA Sequence	GCGGCCGCCCCGGCGGCTCCTGGAACCCCGGTTTCGCGGCGATGCCAGCCACCCCAGCG AAGCCGCCGAGTTTCACTGCTTGGATAATTTGAAAGTACAATAGTTGGTTTCCCTGTC CACCCGCCCCACTTCGCTTGCCATCACAGCACGCCTATCGGATGTGAGAGGAGAAGTC CCGCTGCTCGGGCACTGTCTATATACGCCTAACACCTACATATATTTTAAAAACATTA AATATAATTAACAATCAAAGAAAGAGGAGAAAGGAAGGGAAGCATTACTGGGTTACT ATGCACTTGCGACTGATTTCTTGGCTTTTATCATTTTGAACCTTTATGGAATACATCG GCAGCCAAAACGCCTCCCGGGGAAGGCGCCAGCGAAGAATGCATCCTAACGTTAGTCA AGGCTGCCAAGGAGGCTGTGCAACATGCTCAGATTACAATGGATGTTTGTGTCATGTAAG CCCAGACTATTTTTTGCTCTGGAAAGAATTGGCATGAAGCAGATTGGAGTATGTCTCT CTTCATGTCCAAGTGGATATTATGGAACCTCGATATCCAGATATAAATAAGTGTACAAG TAAGTGCCACACGAAAAGCTGACTGTGATACCTGTTTCAACAAAAATTTCTGCACA AAATGTAAAAGTGGATTTTACTTACACCTTGAAAGTGCCTTGACAATTGCCCAGAAG GGTTGGAAGCCAACAACCATACTATGGAGTGTGTGTCAGTTCAGTGCCTGTGAGGTCAG TGAATGGAATCCTTGAGTCCATGCACGAAGAAGGGAACATGTGGCTTCAAAAGA GGGACTGAAACACGGGTCCGAGAAATAATACAGCATCCTTCAGCAAAGGGTAACCTGT GTCCCCCAACAAATGAGACAAGAAAGTGTACAGTGCAAAGGAAGAAGTGTGAGAAGGG AGAACGAGGTACAATCATAATAACAAAATGTGCTTGTGTTTGAATCCTCATAATCTGTTG CATTTTTCATTTTATTTCTTATGAAACACTTGGCATTATCTTTCATGC		
	ORF Start: ATG at 291	ORF Stop: TGA at 1008	
	SEQ ID NO: 172	239 aa	MW at 27062.1kD
NOV38a, CG59685-01 Protein Sequence	MHLRLISWLFIIILNFM EYIGSQNASRRRQRRMHPNVSQGCQGGCATCS DYNGCL SCK PRLFFALERIGMKQIGVCLSSCP SGYYGTRYPDINKCTSKCPHEKADCDTCFNKNFCT KCKSGFYHLGLKCLDNCP EGGLEANNHTMECVSSVHCEVSEWNPWPCTKKGKTCGFKR GTETRVREIIQHPSAKGNLCPPTNETRKCTVQRKKCQKGERGTIIITKACLNPHNLL HFSFYFL		
	SEQ ID NO: 173	585 bp	
NOV38b,	GGATCCCAAAACGCCTCCCGGGGAAGGCGCCAGCGAAGAATGCATCCTAACGTTAGTC AAGGCTGCCGAGGAGGCTGTGCAACATGCTCAGATTACAATGGATGTTTGTGTCATGTAA		

DEMANDE OU BREVET VOLUMINEUX

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

CECI EST LE TOME 1 DE 3
CONTENANT LES PAGES 1 À 215

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 3
CONTAINING PAGES 1 TO 215

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

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

We claim:

1. 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 86;
 - 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 86, 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) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86;
 - 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 86, 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).
2. The polypeptide of claim 1 that is a naturally occurring allelic variant of the sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86.
3. The polypeptide of claim 2, wherein the allelic variant comprises an amino acid sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 2n, wherein n is an integer between 1 and 86.
4. The polypeptide of claim 1 that is a variant polypeptide described therein, wherein any amino acid specified in the chosen sequence is changed to provide a conservative substitution.

5. A pharmaceutical composition comprising the polypeptide of claim 1 and a pharmaceutically acceptable carrier.
6. A kit comprising in one or more containers, the pharmaceutical composition of claim 5.
7. 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 is the polypeptide of claim 1.
8. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:
 - (a) providing the sample;
 - (b) introducing the sample to an antibody that binds immunospecifically to the polypeptide; and
 - (c) determining the presence or amount of antibody bound to the polypeptide, thereby determining the presence or amount of polypeptide in the sample.
9. A method for determining the presence of or predisposition to a disease associated with altered levels 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 amount of the polypeptide in the sample of step (a) 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.
10. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:
 - (a) introducing the polypeptide to the agent; and

- (b) determining whether the agent binds to the polypeptide.

11. The method of claim 10 wherein the agent is a cellular receptor or a downstream effector.

12. 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 devoid of the substance, the substance is identified as a potential therapeutic agent.

13. A method for screening for a modulator of activity or of latency or predisposition to a pathology associated with the polypeptide of claim 1, the 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 the test animal recombinantly expresses the polypeptide of claim 1;
- b) measuring the activity of the polypeptide in the test animal after administering the compound of step (a); and
- c) 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 claim 1.

14. The method of claim 13, wherein the test animal is a recombinant test animal that expresses a test protein transgene or expresses the transgene under the control of a promoter at an

increased level relative to a wild-type test animal, and wherein the promoter is not the native gene promoter of the transgene.

15. A method for modulating the activity of the polypeptide of claim 1, the method comprising introducing a cell sample expressing the polypeptide of the claim with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

16. 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.

17. The method of claim 16, wherein the subject is a human.

18. 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 86, or a biologically active fragment thereof.

19. 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 86;
- 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 86, 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 86;
 - 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 86, 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 86, 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
 - f) the complement of any of the nucleic acid molecules.
20. The nucleic acid molecule of claim 19, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant.
21. The nucleic acid molecule of claim 19 that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.
22. The nucleic acid molecule of claim 19, 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 86.
23. The nucleic acid molecule of claim 19, wherein the 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 86;
 - 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 86, 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 86; 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 86, 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.

24. The nucleic acid molecule of claim 19, 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 86, or a complement of the nucleotide sequence.

25. The nucleic acid molecule of claim 19, wherein the nucleic acid molecule comprises 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.

26. A vector comprising the nucleic acid molecule of claim 19.

27. The vector of claim 26, further comprising a promoter operably linked to the nucleic acid molecule.

28. A cell comprising the vector of claim 27.

29. A method for determining the presence or amount of the nucleic acid molecule of claim 19 in a sample, the method comprising:

- (a) providing the sample;

- (b) introducing the sample to a probe that binds to the nucleic acid molecule; and
- (c) 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.

30. The method of claim 29 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

31. The method of claim 30 wherein the cell or tissue type is cancerous.

32. A method for determining the presence of or predisposition to a disease associated with altered levels of the nucleic acid molecule of claim 19 in a first mammalian subject, the method comprising:

- a) measuring the amount of the nucleic acid in a sample from the first mammalian subject; and
- b) 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.