



(86) Date de dépôt PCT/PCT Filing Date: 2002/06/04
(87) Date publication PCT/PCT Publication Date: 2002/12/12
(85) Entrée phase nationale/National Entry: 2003/11/25
(86) N° demande PCT/PCT Application No.: US 2002/017559
(87) N° publication PCT/PCT Publication No.: 2002/099062
(30) Priorités/Priorities: 2001/06/04 (60/295,607) US;
2001/06/06 (60/296,404) US; 2001/06/06 (60/296,418) US;
2001/06/07 (60/296,575) US; 2001/06/11 (60/297,414) US;
2001/06/12 (60/297,567) US; 2001/06/12 (60/297,573) US;
2001/06/14 (60/298,285) US; 2001/06/15 (60/298,556) US;
2001/06/15 (60/298,528) US; 2001/06/18 (60/299,133) US;
2001/06/19 (60/299,230) US; 2001/06/21 (60/299,949) US;
...

(51) Cl.Int.⁷/Int.Cl.⁷ A61K 38/00, C07K 14/00

(71) Demandeur/Applicant:
CURAGEN CORPORATION, US

(72) Inventeurs/Inventors:
ANDERSON, DAVID W., US;
ZERHUSEN, BRYAN D., US;
LI, LI, US;
ZHONG, MEI, US;
CASMAN, STACIE J., US;
GERLACH, VALERIE L., US;
SHIMKETS, RICHARD A., US;
...

(74) Agent: GOUDREAU GAGE DUBUC

(54) Titre : NOUVEAUX ANTICORPS SE FIXANT A DES POLYPEPTIDES ANTIGENIQUES, ACIDES NUCLEIQUES
CODANT LES ANTIGENES ET MODES D'UTILISATION

(54) Title: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING THE
ANTIGENS, AND METHODS OF USE

(57) **Abrégé/Abstract:**

Disclosed herein are nucleic acid sequences that encode polypeptides. Also disclosed are 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, polypeptides, or antibodies, or fragments thereof.

- (30) **Priorités(suite)/Priorities(continued):** 2001/06/22 (60/300,177) US; 2001/06/28 (60/301,550) US; 2001/06/28 (60/301,530) US; 2001/07/03 (60/302,951) US; 2001/09/12 (60/318,771) US; 2001/09/25 (60/324,687) US; 2001/10/24 (60/339,266) US; 2001/11/16 (60/337,524) US; 2001/12/14 (60/341,143) US; 2002/02/21 (60/359,151) US; 2002/02/21 (60/358,643) US; 2002/02/28 (60/361,195) US; 2002/03/05 (60/361,964) US; 2002/04/10 (60/371,523) US; 2002/04/10 (60/371,346) US; 2002/06/03 (10/161,493) US
- (72) **Inventeurs(suite)/Inventors(continued):** GORMAN, LINDA, US; PENA, CAROL E. A., US; KEKUDA, RAMESH, US; PATTURAJAN, MEERA, US; SPYTEK, KIMBERLY A., US; LEITE, MARIO W., US; RASTELLI, LUCA, US; MACDOUGALL, JOHN R., US; TAUPIER, RAYMOND J., JR., US; GUO, XIAOJIA, US; MILLER, CHARLES E., US; SHENOY, SURESH G., US; HJALT, TORD, US; VOSS, EDWARD Z., US; BOLDOG, FERENC L., US; MALYANKAR, URIEL M., US; PADIGARU, MURALIDHARA, US; JI, WEIZHEN, US; SMITHSON, GLENNDA, US; EDINGER, SHLOMIT R., US; MILLET, ISABELLE, US; ELLERMAN, KAREN, US

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 December 2002 (12.12.2002)

PCT

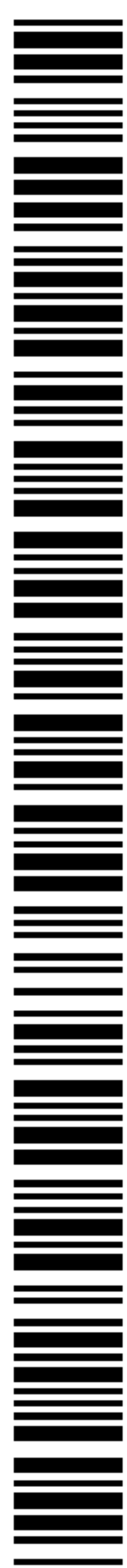
(10) International Publication Number
WO 02/099062 A3

(51) International Patent Classification ⁷ : A61K 38/00 , C07K 14/00		US	60/337,524 (CIP)
		Filed on	16 November 2001 (16.11.2001)
		US	60/296,404 (CIP)
(21) International Application Number: PCT/US02/17559		Filed on	6 June 2001 (06.06.2001)
		US	60/296,418 (CIP)
(22) International Filing Date: 4 June 2002 (04.06.2002)		Filed on	6 June 2001 (06.06.2001)
		US	60/296,575 (CIP)
(25) Filing Language: English		Filed on	7 June 2001 (07.06.2001)
		US	60/359,151 (CIP)
(26) Publication Language: English		Filed on	21 February 2002 (21.02.2002)
		US	60/297,414 (CIP)
(30) Priority Data:		Filed on	11 June 2001 (11.06.2001)
60/295,607	4 June 2001 (04.06.2001)	US	60/297,573 (CIP)
60/296,418	6 June 2001 (06.06.2001)	US	Filed on
60/296,404	6 June 2001 (06.06.2001)	US	12 June 2001 (12.06.2001)
60/296,575	7 June 2001 (07.06.2001)	US	60/341,143 (CIP)
60/297,414	11 June 2001 (11.06.2001)	US	Filed on
60/297,573	12 June 2001 (12.06.2001)	US	14 December 2001 (14.12.2001)
60/297,567	12 June 2001 (12.06.2001)	US	60/297,567 (CIP)
60/298,285	14 June 2001 (14.06.2001)	US	Filed on
60/298,556	15 June 2001 (15.06.2001)	US	12 June 2001 (12.06.2001)
60/298,528	15 June 2001 (15.06.2001)	US	60/318,771 (CIP)
60/299,133	18 June 2001 (18.06.2001)	US	Filed on
60/299,230	19 June 2001 (19.06.2001)	US	12 September 2001 (12.09.2001)
60/299,949	21 June 2001 (21.06.2001)	US	60/298,285 (CIP)
60/300,177	22 June 2001 (22.06.2001)	US	Filed on
60/301,550	28 June 2001 (28.06.2001)	US	14 June 2001 (14.06.2001)
60/301,530	28 June 2001 (28.06.2001)	US	60/298,528 (CIP)
60/302,951	3 July 2001 (03.07.2001)	US	Filed on
60/318,771	12 September 2001 (12.09.2001)	US	15 June 2001 (15.06.2001)
60/324,687	25 September 2001 (25.09.2001)	US	60/299,133 (CIP)
60/339,266	24 October 2001 (24.10.2001)	US	Filed on
60/337,524	16 November 2001 (16.11.2001)	US	18 June 2001 (18.06.2001)
60/341,143	14 December 2001 (14.12.2001)	US	60/299,230 (CIP)
60/359,151	21 February 2002 (21.02.2002)	US	Filed on
60/358,643	21 February 2002 (21.02.2002)	US	19 June 2001 (19.06.2001)
60/361,195	28 February 2002 (28.02.2002)	US	60/358,643 (CIP)
60/361,964	5 March 2002 (05.03.2002)	US	Filed on
60/371,523	10 April 2002 (10.04.2002)	US	21 February 2002 (21.02.2002)
60/371,346	10 April 2002 (10.04.2002)	US	60/299,949 (CIP)
10/161,493	3 June 2002 (03.06.2002)	US	Filed on
(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:		Filed on	21 June 2001 (21.06.2001)
US	60/300,177 (CIP)	US	60/361,964 (CIP)
Filed on	22 June 2001 (22.06.2001)	Filed on	5 March 2002 (05.03.2002)
US	60/295,607 (CIP)	US	60/361,195 (CIP)
Filed on	4 June 2001 (04.06.2001)	Filed on	28 February 2002 (28.02.2002)
		US	60/371,523 (CIP)
		Filed on	10 April 2002 (10.04.2002)
		US	60/301,530 (CIP)
		Filed on	28 June 2001 (28.06.2001)
		US	60/371,346 (CIP)
		Filed on	10 April 2002 (10.04.2002)

[Continued on next page]

(54) Title: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING THE ANTIGENS, AND METHODS OF USE

(57) Abstract: Disclosed herein are nucleic acid sequences that encode polypeptides. Also disclosed are 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, polypeptides, or antibodies, or fragments thereof.



WO 02/099062 A3

WO 02/099062 A3



US 60/301,550 (CIP)
 Filed on 28 June 2001 (28.06.2001)
 US 60/302,951 (CIP)
 Filed on 3 July 2001 (03.07.2001)
 US 60/339,266 (CIP)
 Filed on 24 October 2001 (24.10.2001)
 US 10/161,493 (CIP)
 Filed on 3 June 2002 (03.06.2002)

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

(72) **Inventors; and**

(75) **Inventors/Applicants** (*for US only*): **ANDERSON, David, W.** [US/US]; 85 Montoya Drive, Branford, CT 06405 (US). **ZERHUSEN, Bryan, D.** [US/US]; 337 Monticello Drive, Branford, CT 06405 (US). **LI, Li** [CN/US]; 56 Jerimoth Drive, Branford, CT 06405 (US). **ZHONG, Mei** [CA/US]; 45 Harrison Avenue, Apartment 1B, Branford, CT 06405 (US). **CASMAN, Stacie, J.** [US/US]; 17 Peck Street, North Haven, CT 06473 (US). **GERLACH, Valerie, L.** [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). **SHIMKETS, Richard, A.** [US/US]; 5 Indian Meadows Drive, Guilford, CT 06437 (US). **GORMAN, Linda** [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). **PENA, Carol, E., A.** [US/US]; 604 Orange Street, Number 2, New Haven, CT 06511 (US). **KEKUDA, Ramesh** [IN/US]; 1213 Avalon Valley Drive, Danbury, CT 06810 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street, Number 1, New Haven, CT 06511 (US). **LEITE, Mario, W.** [US/US]; 756 Orange Avenue, Milford, CT 06460 (US). **RASTELLI, Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). **MACDOUGALL, John, R.** [CA/US]; 117 Russell Street, Hamden, CT 06517 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **GUO, Xiaojia** [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). **MILLER, Charles, E.** [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US). **SHENOY, Suresh, G.** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **HJALT, Tord** [SE/US]; 514 Main Street, Apartment 30, East Haven, CT 06512 (US). **VOSS, Edward, Z.** [US/US]; 123 Knollwood Drive, Wallingford, CT 06492 (US). **BOLDOG, Ferenc, L.** [HU/US]; 1687 Hartford Turnpike,

North Haven, CT 06473 (US). **MALYANKAR, Uriel, M.** [IN/US]; 229 Branford Road, Number 330, Branford, CT 06405 (US). **PADIGARU, Muralidhara** [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). **Ji, Weizhen** [CN/US]; 3 Business Park Drive, Room 101, Branford, CT 06405 (US). **SMITHSON, Glenda** [US/US]; 125 Michael Drive, Guilford, CT 06435 (US). **EDINGER, Shlomit, R.** [US/US]; 766 Edgewood Avenue, New Haven, CT 06515 (US). **MILLET, Isabelle** [FR/US]; 74 Carrington Avenue, Milford, CT 06460 (US). **ELLERMAN, Karen** [US/US]; 87 Montoya Drive, Branford, CT 06405 (US).

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

(81) **Designated States** (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, 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.

(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).

Published:

- *with international search report*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) **Date of publication of the international search report:**
 20 February 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 190

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 190

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

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

**NOVEL ANTIBODIES THAT BIND TO ANTIGENIC
POLYPEPTIDES, NUCLEIC ACIDS ENCODING THE ANTIGENS,
AND METHODS OF USE**

5

FIELD OF THE INVENTION

The present invention relates to novel antibodies that bind immunospecifically to
10 antigenic polypeptides, wherein the polypeptides have characteristic properties related to
biochemical or physiological responses in a cell, a tissue, an organ or an organism. The
novel polypeptides are gene products of novel genes, or are specified biologically active
fragments or derivatives thereof. Methods of use of the antibodies encompass procedures
for diagnostic and prognostic assay of the polypeptides, as well as methods of treating
15 diverse pathological conditions.

BACKGROUND OF THE INVENTION

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

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

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

Many pathological conditions involve dysregulation of expression of important effector proteins. In certain classes of pathologies the dysregulation is manifested as elevated or excessive synthesis and secretion of protein effectors. In a clinical setting a subject may be suspected of suffering from a condition brought on by elevated or excessive levels of a protein effector of interest.

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

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

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of nucleic acid sequences encoding novel polypeptides. The novel nucleic acids and polypeptides are referred to herein as NOVX, or NOV1, NOV2, NOV3, *etc.*, nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as “NOVX” nucleic acid or polypeptide sequences.

In one aspect, the invention provides an isolated polypeptide comprising a mature form of a NOVX amino acid. The polypeptide can be, for example, a NOVX amino acid sequence or a variant of a NOVX amino acid sequence, wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than

15% of the amino acid residues in the sequence are so changed. The invention also includes fragments of any of NOVX polypeptides. In another aspect, the invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof.

5 Also included in the invention is a NOVX polypeptide that is a naturally occurring variant of a NOVX sequence. In one embodiment, the variant includes an amino acid sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a NOVX nucleic acid sequence. In another embodiment, the NOVX polypeptide is a variant polypeptide described therein, wherein any amino acid specified in the chosen
10 sequence is changed to provide a conservative substitution.

In another aspect, invention provides a method for determining the presence or amount of the NOVX polypeptide in a sample by providing a sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the NOVX polypeptide, thereby determining
15 the presence or amount of the NOVX polypeptide in the sample.

In yet another aspect, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide in a mammalian subject by measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in the
20 sample of the first step to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease. 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 aspect, the invention includes pharmaceutical compositions that include
25 therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or an antibody specific for a NOVX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective amount of this pharmaceutical composition.

30 In still another aspect, the invention provides the use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease that is associated with a NOVX polypeptide.

In a further aspect, the invention provides a method for modulating the activity of a NOVX polypeptide by contacting a cell sample expressing the NOVX polypeptide with

antibody that binds the NOVX polypeptide in an amount sufficient to modulate the activity of the polypeptide.

The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. In a preferred embodiment, the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant. In another embodiment, the nucleic acid encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant. In another embodiment, the nucleic acid molecule differs by a single nucleotide from a NOVX nucleic acid sequence. In one embodiment, the NOVX nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 73, or a complement of the nucleotide sequence. In one embodiment, the invention provides a nucleic acid molecule wherein the nucleic acid includes the nucleotide sequence of a naturally occurring allelic nucleic acid variant.

Also included in the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein. The invention is also directed to host cells transformed with a vector comprising any of the nucleic acid molecules described above.

In yet another aspect, the invention provides for a method for determining the presence or amount of a nucleic acid molecule in a sample by contacting a sample with a probe that binds a NOVX nucleic acid and determining the amount of the probe that is bound to the NOVX nucleic acid. For example the NOVX nucleic acid may be a marker for cell or tissue type such as a cell or tissue type that is cancerous.

In yet a further aspect, the invention provides a method for determining the presence of or predisposition to a disease associated with altered levels of a nucleic acid molecule in a first mammalian subject, wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

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

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

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

10 Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In 15 the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting.

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

20

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 25 referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table 1 provides a summary of the NOVX nucleic acids and their encoded polypeptides.

30

**TABLE 1. NOVX Polynucleotide and Polypeptide Sequences and
Corresponding SEQ ID Numbers**

	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1a	CG100653-01	1	2	Cadherin Associated Protein-like
2a	CG100689-01	3	4	Leucine Rich Repeat-like
3a	CG100760-01	5	6	Leucine Rich Repeat-like
4a	CG100851-02	7	8	Leukocyte Surface Antigen CD53-like
5a	CG101068-01	9	10	Claudin-9-like
6a	CG101231-01	11	12	Integral Membrane Protein Isoform 2-like
6b	CG101231-02	13	14	Integral Membrane Protein Isoform 2-like
7a	CG101362-01	15	16	Prion Protein-like
8a	CG101458-01	17	18	Von Willebrand Domain Containing Protein-like
9a	CG101475-01	19	20	Plasma Membrane Protein-like
9b	CG101475-02	21	22	Plasma Membrane Protein-like
10a	CG101772-01	23	24	XAGE-like
11a	CG102532-01	25	26	Emerin-like
12a	CG102575-01	27	28	ATPase-like
12b	CG102575-02	29	30	ATPase-like
13a	CG102615-01	31	32	Mat8 (Mammary Tumor 8 kDa) Protein-like
13b	CG102615-04	33	34	Mat8 (Mammary Tumor 8 kDa) Protein-like
14a	CG102646-01	35	36	High Affinity Proline Permease-like
15a	CG102878-01	37	38	Transmembrane-like
15b	CG102878-02	39	40	Transmembrane-like
16a	CG103459-01	41	42	Peptide/Histidine Transporter-like
17a	CG104210-01	43	44	Type III Membrane Protein-like
17b	CG104210-02	45	46	Type III Membrane Protein-like
17c	272249075	47	48	Type III Membrane Protein-like
18a	CG104251-01	49	50	Type III Membrane Protein-like
19a	CG104934-01	51	52	Phospholipid-Transporting ATPase IH-like
20a	CG105463-01	53	54	Meningioma-Expressed Antigen 6/11 (MEA6) (MEA11)-like
20b	CG105463-02	55	56	Meningioma-Expressed Antigen 6/11 (MEA6) (MEA11)-like
21a	CG105491-01	57	58	Serine Protease-like
22a	CG105954-01	59	60	Neurofascin Precursor-like
23a	CG105963-01	61	62	Cadherin-like
24a	CG105973-01	63	64	Integrin Alpha 8-like
24b	CG105973-02	65	66	Integrin Alpha 8-like
25a	CG106915-01	67	68	Nogo Receptor Isoform-1-like
26a	CG106924-01	69	70	Nogo Receptor Isoform-2-like
26b	210062144	71	72	Nogo Receptor Isoform-2-like
27a	CG106942-01	73	74	NRAMP-like Membrane Protein
28a	CG107513-01	75	76	Syntaxin Domain Containing Protein-like
29a	CG107533-02	77	78	Tumor Necrosis Factor-like
30a	CG107562-01	79	80	Leucine-Rich Repeat Type III

				Transmembrane-like
30b	CG107562-02	81	82	Leucine-Rich Repeat Type III Transmembrane-like
30c	210086373	83	84	Leucine-Rich Repeat Type III Transmembrane-like
30d	210086403	85	86	Leucine-Rich Repeat Type III Transmembrane-like
30e	210086422	87	88	Leucine-Rich Repeat Type III Transmembrane-like
31a	CG108184-01	89	90	Transmembrane Protein Tm7-like
31b	CG108184-02	91	92	Transmembrane Protein Tm7-like
31c	CG108184-03	93	94	Transmembrane Protein Tm7-like
32a	CG108238-01	95	96	Sialic Acid Binding Immunoglobulin-like
33a	CG108695-01	97	98	OB binding protein (SIGLEC)-like
34a	CG109505-01	99	100	Aldehyde Dehydrogenase-like
35a	CG109742-01	101	102	Latent Transforming Growth Factor Beta Binding Protein 3-like
35b	207639410	103	104	Latent Transforming Growth Factor Beta Binding Protein 3-like
35c	207639427	105	106	Latent Transforming Growth Factor Beta Binding Protein 3-like
35d	207639438	107	108	Latent Transforming Growth Factor Beta Binding Protein 3-like
35e	207639448	109	110	Latent Transforming Growth Factor Beta Binding Protein 3-like
36a	CG109844-01	111	112	C4B-Binding Protein-like
37a	CG110014-02	113	114	Colon Carcinoma kinase 4-like
37b	CG110014-03	115	116	Colon Carcinoma kinase 4-like
37c	CG110014-04	117	118	Colon Carcinoma kinase 4-like
38a	CG110187-01	119	120	Alpha C1-like Protocadherin
38b	CG110187-03	121	122	Alpha C1-like Protocadherin
39a	CG110205-01	123	124	Disintegrin-like / Metalloprotease (Reprolysin Type) with Thrombospondin Type I Motif-like
39b	CG110205-02	125	126	Disintegrin-like / Metalloprotease (Reprolysin Type) with Thrombospondin Type I Motif-like
39c	207756942	127	128	Disintegrin-like / Metalloprotease (Reprolysin Type) with Thrombospondin Type I Motif-like
39d	207756946	129	130	Disintegrin-like / Metalloprotease (Reprolysin Type) with Thrombospondin Type I Motif-like
39e	207756950	131	132	Disintegrin-like / Metalloprotease (Reprolysin Type) with Thrombospondin Type I Motif-like
39f	207756966	133	134	Disintegrin-like / Metalloprotease (Reprolysin Type) with Thrombospondin Type I Motif-like
40a	CG110242-01	135	136	Ebnerin-like
40b	207728344	137	138	Ebnerin-like
40c	207728348	139	140	Ebnerin-like
40d	207728354	141	142	Ebnerin-like
40e	207728365	143	144	Ebnerin-like
41a	CG99598-01	145	146	Endosomal Glycoprotein Precursor-like

Table 1 indicates the homology of NOVX polypeptides to known protein families. Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table 1 will be useful in therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table 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 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 domains that are characteristic of, other members of such protein families. Details of the sequence relatedness and domain analysis for each NOVX are presented in Example A.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit diseases associated with the protein families listed in Table 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 C. Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the detection of a variety of diseases with differential expression in normal vs. diseased tissues, *e.g.* detection of a variety of cancers.

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

NOVX clones

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

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

The NOVX genes and their corresponding encoded proteins are useful for preventing, treating or ameliorating medical conditions, *e.g.*, by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

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

In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 73; (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 73, 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 73; (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 73 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 73; (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 73 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 73; (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 73, 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 73 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 73; (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 73 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 73; 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 73 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-
5 encoding nucleic acids (*e.g.*, NOVX mRNA's) 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
10 acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

A NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally
15 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, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host
20 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
25 methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-
30 translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probes", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- 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 utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, *etc.*). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

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

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

As used herein, the term “oligonucleotide” refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue.

Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NO:2 n -1, wherein n is an integer between 1-73, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence SEQ ID NO:2 n -1, wherein n is an integer between 1-73, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of a NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence of SEQ ID NO:2 n -1, wherein n is an integer between 1-73, is one that is sufficiently complementary to the nucleotide sequence of SEQ ID NO:2 n -1, wherein n is an integer between 1-73, that it can hydrogen bond with little or no mismatches to the nucleotide sequence of SEQ ID NO:2 n -1, wherein n is an integer between 1-73, thereby forming a stable duplex.

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

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, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

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

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

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those

sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

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

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73; or an anti-sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73; or of a naturally occurring mutant of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73.

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 further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor.

5 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.

10 "A polypeptide having a biologically-active portion of a NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, that encodes a polypeptide having a NOVX
15 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
20 nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence of SEQ ID NO:2*n*, wherein *n*
25 is an integer between 1-73.

In addition to the human NOVX nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, 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
30 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.

5 Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from any one of the human SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, 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
10 to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

 Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the
15 nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to
20 describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

 Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a
25 probe using methods well known in the art for nucleic acid hybridization and cloning.

 As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures
30 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 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.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65 °C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50 °C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to any one of the sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Reinhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55 °C, followed by one or more washes in 1X SSC, 0.1% SDS at 37 °C. Other conditions of moderate stringency that may be used are well-known within the art. See, e.g., Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Krieger, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions

are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40 °C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50 °C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* 78: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, thereby leading to changes in the amino acid sequences of the encoded NOVX proteins, without altering the functional ability of said NOVX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1-73. 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 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 any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, 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 of SEQ ID NO:2*n*, wherein *n* is an integer between 1-73. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1-73; more preferably at least about 70% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1-73; still more preferably

at least about 80% homologous to SEQ ID NO:2 n , wherein n is an integer between 1-73; even more preferably at least about 90% homologous to SEQ ID NO:2 n , wherein n is an integer between 1-73; and most preferably at least about 95% homologous to SEQ ID NO:2 n , wherein n is an integer between 1-73.

5 An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein of SEQ ID NO:2 n , wherein n is an integer between 1-73, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:2 n -1, wherein n is an integer between 1-73, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

10 Mutations can be introduced into any of SEQ ID NO:2 n -1, wherein n is an integer between 1-73, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar
15 side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and
20 aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such
25 as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of any one of SEQ ID NO:2 n -1, wherein n is an integer between 1-73, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

30 The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that

may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

5 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).

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

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
 15 nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, 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
 20 molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1-73, or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO:2*n*-1,
 25 wherein *n* is an integer between 1-73, 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
 30 molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

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

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(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
5 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
10 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
15 sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

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

25 **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
30 nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a

complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX cDNA disclosed herein (*i.e.*, any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. See, *e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

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

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. See, *e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleotide bases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomer can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNA 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*).

5 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
10 DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleotide bases, and orientation (*see, Hyrup, et al., 1996. supra*). The synthesis of PNA-DNA chimeras can be performed as described
15 in Hyrup, *et al., 1996. supra* and Finn, *et al., 1996. Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g., Mag, et al., 1989. Nucl Acid Res* 17: 5973-5988.
20 PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g., Finn, et al., 1996. supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g., Petersen, et al., 1975. Bioorg. Med. Chem. Lett.* 5: 1119-11124.

 In other embodiments, the oligonucleotide may include other appended groups such
25 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
30 (*see, e.g., Krol, et al., 1988. BioTechniques* 6:958-976) or intercalating agents (*see, e.g., Zon, 1988. Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1-73. The invention also includes a
5 mutant or variant protein any of whose residues may be changed from the corresponding residues shown in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1-73, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, a NOVX variant that preserves NOVX-like function includes any variant
10 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
15 conservative substitution as defined above.

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

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

10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably
5 less than about 5% of the volume of the NOVX protein preparation.

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

15 Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1-73) 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
20 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
25 the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1-73. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1-73, and retains the functional activity of the protein of SEQ ID NO:2*n*, wherein *n* is an integer
30 between 1-73, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between

1-73, and retains the functional activity of the NOVX proteins of SEQ ID NO:2 n , wherein n is an integer between 1-73.

Determining Homology Between Two or More Sequences

5 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
10 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
15 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
20 sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence of SEQ ID NO:2 n -1, wherein n is an integer between 1-73.

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
25 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
30 comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent

identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

5 The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1-73, whereas a "non-NOVX polypeptide" refers to a
10 polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically-
15 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
20 with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

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

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

30 In yet another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the

surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX Agonists and Antagonists

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

limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

5 Variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a
10 variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods
15 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.
20 Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.*, Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

25 In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per
30 molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with 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
5 libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of
10 vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. *See, e.g., Arkin and Yourvan, 1992. Proc.*
15 *Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

NOVX Antibodies

The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that
20 contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} , F_{ab}' and $F_{(ab)2}$ fragments, and an F_{ab} expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule.
25 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
30 fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An

antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence of SEQ ID NO:2 n , wherein n is an integer between 1-73, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

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

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

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

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

Polyclonal Antibodies

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

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen 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).

5 **Monoclonal Antibodies**

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs)
10 of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

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

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

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

10 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
15 radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

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

25 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.

30 The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into

expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

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

human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

Human Antibodies

5 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
10 hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et
15 al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by
20 introducing human immunoglobulin loci into transgenic animals, *e.g.*, mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806;
25 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al. (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

30 Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light

immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the 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

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

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

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

least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., *Methods in Enzymology*, 121:210 (1986).

5 According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced
10 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.

15 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
20 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
25 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
30 Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling *in vitro* to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

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

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

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

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted

cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

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

Immunoconjugates

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

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, *Phytolacca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the

tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as 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 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.

Immunoliposomes

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

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

Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention

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 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 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 of the subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice

of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York.

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

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

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

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, *e.g.*, films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and γ ethyl-L-glutamate, non-

degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON 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
5 over 100 days, certain hydrogels release proteins for shorter time periods.

ELISA Assay

An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be
10 polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, F_{ab} or F_{(ab)2}) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of
15 indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological
20 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
25 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;
30 "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.

NOVX Recombinant Expression Vectors and Host Cells

5 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
10 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
15 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
20 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
25 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
30 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.

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
5 TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

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

15 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: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

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

In yet another embodiment, a nucleic acid of the invention is expressed in
25 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
30 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

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

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

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and

"recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

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

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

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

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

medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

Transgenic NOVX Animals

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

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences, *i.e.*, any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be

operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, 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.

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

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

 Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit
30 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 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
5 (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion
10 media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles
15 such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated
20 into the compositions.

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

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

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

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is

applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

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

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

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

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

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used

herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and
5 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,
10 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
15 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.

20 **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
25 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
30 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
5 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,
10 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
15 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
20 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.

25 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
30 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 second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

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

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

with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

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

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

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

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as

to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and
5 micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the
10 non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes
15 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
20 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
25 molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity
30 associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of

expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

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

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

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

Detection Assays

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

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, 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.

5 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.

10 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
15 individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see*, Verma, *et al.*, HUMAN
20 CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

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

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line
30 through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease.

5 Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

10 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
15 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
20 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
25 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
30 humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes.

Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If coding sequences, such as those of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

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

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

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

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

5 **Diagnostic Assays**

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the
10 presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-73, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500
15 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
20 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
25 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
30 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.

5 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.

10 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.

15 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
20 suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

Prognostic Assays

25 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
30 subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder

associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

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

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

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

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

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

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing

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

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence.

Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g., Naeve, et al., 1995. Biotechniques* 19: 448), including sequencing by mass spectrometry (*see, e.g., PCT International Publication No. WO 94/16101; Cohen, et al., 1996. Adv. Chromatography* 36: 127-162; and Griffin, *et al., 1993. Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g., Myers, et al., 1985. Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with 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.

5 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
10 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.
15 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
20 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA
25 fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

30 In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of

high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g., Rosenbaum and Reissner, 1987. Biophys. Chem. 265: 12753.*

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

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

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

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein.

However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

5 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, 10 Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.) 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 15 drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such 20 pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the 25 response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering 30 the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of

oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome pregnancy zone protein precursor enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

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

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described

herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical 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) nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention

or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with
5 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
10 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,
15 immunocytochemistry, *etc.*) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

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

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention

involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX

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

Stimulation of NOVX activity is desirable *in situations* in which NOVX is
20 abnormally downregulated and/or in which increased NOVX activity has 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).

25 **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
30 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
5 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
10 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,
15 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
20 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.

25

EXAMPLES

Example A: Polynucleotide and Polypeptide Sequences, and Homology Data

5

Example 1.

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

Table 1A. NOV1 Sequence Analysis		
	SEQ ID NO: 1	3430 bp
NOV1a, CG100653-01 DNA Sequence	GGGCTGCAGGAATCCCCCACAGAGGGAGCATGACTTCGGCAACTTCACCTATCATTC TGAAATGGGACCCCAAAGTTTGGAAATCCGGACGCTAACAGTGGAAAGGCTGTTGGA GCCACTTGTTACACAGGTGACTACACTTGTCAACACAAGCAACAAAGGCCCATCTGGT AAAAAGAAAGGGAGGTCAAAGAAAGCCCATGTACTAGCTGCCTCTGTAGAGCAAGCCA CTCAGAATTTCTTGAAAAGGGTGAACAGATCGCTAAGGAGAGTCAAGATCTCAAAGA AGAGTTGGTGGCTGCTGTAGAGGATGTGCGCAAACAAGGTGAGACGATGCGGATCGCC TCCTCCGAGTTTGCAGATGACCCCTTGCTCGTCGGTAAAGCGCGGCACCATGGTACGGG CGGCAAGGGCTTTGCTCTCCGCGGTGACACGCTTACTCATCCTGGCGGACATGGCAGA TGTCATGAGACTTTTATCCCATCTGAAAATTGTGGAAGAGGCCCTGGAAGCTGTCAA AATGCTACAAATGAGCAAGACCTTGCAAACCGTTTAAAGAGTTTGGGAAAAGATGG TGAAACTTAACATATGTAGCAGCAAGAAGACAACAGGAGCTGAAGGATCCTCACTGTCG GGATGAGATGGCAGCCGCCGAGGGGCTCTGAAGAAGAATGCCACAATGCTGTACACG GCCTCTCAAGCATTCTCCGCCACCCAGATGTCGCCGCTACGAGAGCCAACCGAGATT ATGTGTTCAAACAAGTCCAGGAGGCCATCGCCGGCATCTCCAATGCTGCTCAAGCTAC CTCGCCCACTGACGAAGCCAAGGGCCACACGGGCATCGGCGAGCTGGCTGCGGCTCTT AATGAGTTTGACAATAAGATTATCCTGGACCCCATGACGTTCAAGCAGGAGGAGGTTCC GGCCGTCCCTGGAGGAGAGGCTGGAGAGCATCATCAGCGGCGCAGCGCTGATGGCCGA CTCCTCCTGCACGCGAGACGACCGGCGCGAGAGGATCGTGGCGGAGTGCAACGCCGTG CGGCAGGCGCTCCAGGACCTGCTCAGCGAGTACATGAATAATACTGGAAGGAAAGAAA AAGGAGATCCTCTCAACATTGCGATTGATAAGATGACTAAGAAAACAAGAGATCTAAG GAGACAGCTTCGGAAAGCAGTGATGGATCACATATCTGACTCTTTCCTGGAACCAAT GTTCTTTGCTAGTTCTCATTTGAGGCTGCAAAGAGCGGAAATGAAAAGGAAGTGAAAG AATATGCCCAAGTTTTCGTGAGCATGCCAACAACTGGTAGAGGTTGCCAATTTGGC CTGTTCCATCTCCAACAATGAAGAAGGGGTGAAATTAGTTCCGATGGCAGCCACCCAG ATTGACAGCCTGTGTCCCCAGGTCATCAATGCCGCTCTGACACTGGCTGCCCGGCCAC AGAGCAAAGTTGCTCAGGATAACATGGACGTCTTCAAAGACCAGTGGGAGAAGCAGGT CCGAGTGTGACAGAGGCCGTGGATGACATCACCTCAGTGGATGACTTCCTCTCTGTC TCAGAAAATCACATCTTGGAGGATGTGAACAAGTGTGTGATAGCCCTCCAAGAGGGCG ATGTGGACACTCTGGACCGGACTGCAGGGGCCATCAGGGGCCGGGCAGCTCGAGTCAT ACACATCATCAATGCTGAGATGGAGAATATGAAGCTGGGGTTTATACTGAGAAGGTG TTGGAAGCTACAAAATTGCTTTCTGAAACAGTGATGCCACGCTTCGCTGAACAAGTAG AGGTTGCCATTGAAGCCCTGAGTGCCAACGTTCTCAACCGTTTGAGGAGAATGAGTT CATCGATGCCTCTCGCCTGGTGTATGATGGCGTTTCGGGACATCAGAAAGGCTGTGCTG ATGATCAGGACCCCAAGAACTAGAGGATGATTCTGACTTTGAGCAGGAAGATTATG ATGTGCGTAGAGGGACAAGTGTTCAGACTGAGGATGACCAGCTCATTGCAGGGCAGAG CGCACGGGCCATCATGGCGCAACTACCGCAGGAGGAGAAGGCAAAAATAGCTGAGCAG GTGGAGATATTCCATCAAGAGAAAAGCAAGCTGGATGCAGAAGTGGCCAAATGGGACG ACAGCGGCAATGATATCATTGTACTGGCCAAGCAGATGTGTATGATCATGATGAAAT GACAGACTTCACAAGAGGCAAAGGCCCATTTGAAAAATACATCTGATGTCATTAATGCT GCCAAGAAAATTGCCGAAGCAGGTTCTCGAATGGACAAATTAGCTCGTGCTGTGGCTG ATCAGCTGGACAGTGCCACATCGCTTATCCAGGCAGCTAAAAACCTGATGAATGCTGT TGTCCTCACGGTGAAAGCATCCTATGTGGCCTCAACCAAATACCAGAAGGTCTATGGG ACAGCAGCTGTCAACTCACCTGTTGTGCTTGGAAAGATGAAGGCTCCAGAGAAGAAGC	

	CCCTTGTGAAGAGAGAGAAAAGCCTGAAGAATTCCAGACACGAGTTCGACGAGGTTCTCA GAAGAAACACATTTTCGCCTGTACAGGCTTTAAGTGAATTCAAAGCAATGGATTCTTTC TAGGACGATAGGTTTTTAACAAGAAAGCTTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT TTTAATTCCATTTTTGTATGCATACCTGCCAGCTCGTATGCCTCTGGCATGGGGAAAT TAAGGGAACAGTGTCTGTTTGCATGTAAGATGAGATGAGATCAATACTACTGATCCAT CTGTAGCCTGGGAAGGAGACAGGACATTCTGTACTAAGGTGGCACAGAGCTGTCCTT TGCAACATTCTCATAAAATTGGGCACAGAGTTCGCATTGGCGCAATATTTATGGGAGT GGGAGGGATGGGGAAAATAAACTTAACTCTACAAAAGCAAACCTAATGCATGCAAGA ATCATTAGGTGGCAGGTATATGCATAAGTGAAGAAATCTGGAAGTGAATGGTAGAAC ATAAACTTGTATTGCTTCTGTTTCAGTGCAAAAATGTACTAGCCAATACGCTTAAGT GTGTGGCCCATGAATTGAACAATTTAACCTTGAAGTCTATATCCGTGATATTATGTCTG ATTTTAACTGAGGGGAAATTAAGTAGTCCAGCCTAAAATGCTTCTTTTAATCTGCAT TCTGTTTCTCTCTTAGTGTGTCATTACTAGTGATCATGTTTTTTTCCCCCTTTAA TGAAAACAATAAACATCTATTTGAGACAATTAATAATCCTTCTGGGGGCACTGGAAGCA CAATACGGTGACCAATCTTGCTTTCATTTTTTTTTCTTTTAAATTTGAACCATGATTT TGCTAGAAATAGAAGGCCAGTGGTGGAAATATTAGAGGGAAGGAACTGACAACGTGT GAAAGTTA		
	ORF Start: ATG at 31	ORF Stop: TAG at 2611	
	SEQ ID NO: 2	860 aa	MW at 95525.9kD
NOV1a, CG100653-01 Protein Sequence	MTSATSPIILKWDPKSLEIRTLTVERLLEPLVTQVTTLVNTSNKGPSGKKKGRSKKAH VLAASVEQATQNFLEKGEQIAKESQDLKEELVAAVEDVRKQGETMRIASSEFADDPSC SVKRGTMVRAARALLSAVTRLLILADMADVMRLLSHLKIVEEALAEAVKNATNEQDLAN RFKEFGKKMVKLNYVAARRQQELKDPHCRDEMAAARGALKKNATMLYTASQAFLRHPD VAATRANRDYVFKQVQEAIAGISNAAQATSPDEAKGHTGIGELAAALNEFDNKIILD PMTFSEARFRPSLEERLESIIISGAALMADSSCTRDDRERIVAECNAVRQALQDLLSE YMNNTGRKEKGDPLNIAIDKMTKKTRDLRRQLRKAVMDHISDSFLETNPVLLVLIEAA KSGNEKEVKEYAQVFREHANKLVEVANLACSI SNNEEGVKLVMAATQIDSLCPQVIN AALTLAARPQSKVAQDNMDVFKDQWEKQVRVLTEAVDDITSVDDFLSVSENHILEDVN KCVIALQEGDVDTLDRTAGAIRGRAARVIHIINAEMENYEAGVYTEKVLKLEATKLLSET VMPRFAEQVEVAIEALSANVPQPFENEFFIDASRLVYDGVDRDIRKAVLMIRTPPELED DSDFEQEDYDVRGTSVQTEDDQLIAGQSARAIMAQLPQEEKAKIAEQVEIFHQEKSK LDAEVAKWDDSGNDIIVLAKQMCMIMMEMTDFTRGKGPLKNTSDVINAAKKIAEAGSR MDKLARAVADQLDSATSLIQAANKLMNAVVLTVKASYVASTKYQKVYGTAAVNSPVVS WKMKAPEKKPLVKREKPEEFQTRVRRGSQKKHISPVALSEFKAMDSF		

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

Table 1B. Protein Sequence Properties NOV1a	
PSort analysis:	0.3600 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 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 NOV1a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 1C.

Table 1C. Geneseq Results for NOV1a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR58778	Neural alpha-catenin protein - Homo sapiens, 906 aa. [JP06211898-A, 02-AUG-1994]	1..860 1..906	851/906 (93%) 855/906 (93%)	0.0
AAV07060	Renal cancer associated antigen precursor sequence - Homo sapiens, 906 aa. [WO9904265-A2, 28-JAN-1999]	8..859 9..905	694/899 (77%) 773/899 (85%)	0.0
AAU32945	Novel human secreted protein #3436 - Homo sapiens, 932 aa. [WO200179449-A2, 25-OCT-2001]	8..769 10..773	611/766 (79%) 683/766 (88%)	0.0
ABG10622	Novel human diagnostic protein #10613 - Homo sapiens, 932 aa. [WO200175067-A2, 11-OCT-2001]	8..769 10..773	610/766 (79%) 682/766 (88%)	0.0
ABG10622	Novel human diagnostic protein #10613 - Homo sapiens, 932 aa. [WO200175067-A2, 11-OCT-2001]	8..769 10..773	610/766 (79%) 682/766 (88%)	0.0

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

Table 1D. Public BLASTP Results for NOV1a				
Protein Accession Number	Protein/Organism/Length	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P30997	Alpha-2 catenin (Alpha N-catenin) (Neural alpha-catenin) - Gallus gallus (Chicken), 906 aa.	1..860 1..906	851/906 (93%) 855/906 (93%)	0.0
I49499	alpha N-catenin I - mouse, 905 aa.	1..860 1..905	850/905 (93%) 854/905 (93%)	0.0
A45011	alpha-catenin 2 - human, 945 aa.	1..769 1..770	768/770 (99%) 768/770 (99%)	0.0
P26232	Alpha-2 catenin (Alpha-catenin related protein) (Alpha N-catenin) - Homo sapiens (Human), 953 aa.	1..769 1..770	768/770 (99%) 768/770 (99%)	0.0
Q61301	Alpha-2 catenin (Alpha-catenin related protein) (Alpha N-catenin) - Mus musculus (Mouse), 953 aa.	1..769 1..770	759/770 (98%) 762/770 (98%)	0.0

PFam analysis predicts that the NOV1a protein contains the domains shown in the Table 1E.

Table 1E. Domain Analysis of NOV1a			
Pfam Domain	NOV1a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Vinculin	18..765	424/948 (45%) 736/948 (78%)	0
Vinculin	766..821	32/57 (56%) 56/57 (98%)	5.4e-30

5

Example 2.

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

Table 2A. NOV2 Sequence Analysis		
	SEQ ID NO: 3	2883 bp
NOV2a, CG100689-01	CGTGAATGGTGTAGTGAGTTCTAATGAACTTTATTTACAAGAGGAGACTGACCAGGT TTGGCCTGGGGGCCACAGTGTGTAGACCCCTGGAAAGATACATCCTGAGAAGAAAAA	

DNA Sequence	AGAATATATGCAGGAATGCTTAACTTTGTGGGTTTTCTCTCCTCTTGCCCTCACTGAC TCAGGATACACAAAGACCTATCAAGCTCACGCAAAGCAGAAATTCAGCCGCTTATGGT CCAGCAAGTCTGTCACTGAGATTACCTATACTTTGAGGAGGAAGTCAAGCAAGAAGA ATGTGACCATTGACCGCCTTTTGTCTCCCAAGGAAGCTGGGAAACAGCCACGTACA GTGATCATTCAAGGACCACAAGGAATTGGAAAAACGACACTCCTGATGAAGCTGATGA TGGCCTGGTCCGACAACAAGATCTTTCCGGGATAGGTTTCTGTACACGTTCTATTTCTG CTGCAGAGAACTGAGGGAGTTGCCGCCAACGAGTTTGGCTGACTTGATTTCAGAGAG TGGCCTGACCCCGCTGCTCCTATAACAGAGATCGTGTCTCAACCGGAGAGACTCTTGT TCGTCATCGACAGCTTCGAAGAGCTGCAGGGCGGCTTGAACGAACCCGATTCGGATCT GTGTGGTGAAGTGGAGAAACGGCCGGTGCAGGTGCTTCTGAGCAGTTTGCTGAGG AAGAAGATGCTCCCGGAGGCCTCCCTGCTCATCGCTATCAAACCCGTGTGCCCGAAGG AGCTCCGGGATCAGGTGACGATCTCAGAAATCTACCAGCCCCGGGGATTCAACGAGAG TGATAGGTTAGTGTATTTCTGCTGTTTCTTCAAAGACCCGAAAAGAGCCATGGAAGCC TTCAATCTTGTAAAGAGAAAGTGAACAGCTGTTTTCCATATGCCAAATCCCGCTCCTCT GCTGGATCCTGTGTACCAGTCTGAAGCAAGAGATGCAGAAAGGAAAAGACCTGGCCCT GACCTGCCAGAGCACTACCTCTGTGTACTCCTCTTTCGTCTTTAACCTGTTTACACCT GAGGGTGCCGAGGGCCCGACTCCGCAAACCCAGCACCAGCTGAAGGCCCTGTGCTCCC TGGCTGCAGAGGGTATGTGGACAGACACATTTGAGTTTTGTGAAGACGACCTCCGGAG AAATGGGGTGTGTGACGCTGACATCCCTGCGCTGCTGGGCACCAAGATACTTCTGAAG TACGGGGAGCGTGAGAGCTCCTACGTGTTTCTCCACGTGTGTATCCAGGAGTTCTGTG CCGCCTTGTCTATTTGCTCAAGAGCCACCTTGATCATCCTCACCAGCTGTGAGATG TGTACAGGAATTGCTAGTTGCCAATTTTGAAAAAGCAAGGAGAGCACATTGGATTTTT TTGGGGTGTTTTCTAACTGGCCTTTTAAATAAAAAGGAACAAGAAAACTGGATGCGT TTTTTGGCTTCCAACGTGTTCCCAAGAGATAAAGCAGCAAATTCACCAGTGCCTGAAGAG CTTAGGGGAGCGTGGCAATCCTCAGGGACAGGTGGATTCTTGGCGATATTTTACTGT CTCTTTGAAATGCAGGATCCTGCCTTTGTGAAGCAGGCAGTGAACCTCCTCCAAGAAG CTAACCTTTCATATTATTGACAACGTGGACTTGGTGGTTTCTGCCTACTGCTTAAATA CTGCTCCAGCTTGAGGAACTCTGTTTTTCCGTTCAAATGTCTTTAAGAAAGAGGAT GAACACAGCTCTACGTCGGATTACAGCCTCATCTGTTGGCATCACATCTGCTCTGTGC TCACCACCAGCGGGCACCTCAGAGAGCTCCAGGTGCAGGACAGCACCTCAGCGAGTC GACCTTTGTGACCTGGTGTAAACCAGCTGAGGCATCCCAGCTGTGCGCTTCAGAAGCTT GGAATAAATAACGTTTCTTTTCTGGCCAGAGTGTTCTGCTCTTTGAGGTGCTCTTTT ATCAGCCAGACTTGAAATACCTGAGCTTCACCTCAGCAAATCTCTCGTGATGACAT CAGGTCCCTCTGTGATGCCTTGAACCTACCAGCAGGCAACGTCAAAGAGCTAGCGCTG GTAAATTGTACCTCTCACCATTGATTGTGAAGTCCTTGCTGGCCTTCTAACCAACA ACAAGAAGCTGACGTATCTGAATGTATCCTGCAACCAGTTAGACACAGGCGTGCCCTT TTTGTGTGAAGCCCTGTGCAGCCAGACACGGTCTGGTATACCTGATGTTGGCTTTC TGCCACCTCAGCGAGCAGTGCTGCGAATACATCTCTGAAATGCTTCTGCGTAACAAGA GCGTGCGCTATCTAGACCTCAGTGCCAATGTCTGAAGGACGAAGGACTGAAAACCTCT CTGCGAGGCCTTGAAACATCCGGACTGCTGCCTGGATTCACTGTGTTTGGTAAATGT TTTATCACTGCTGCTGGCTGTGAAGACCTCGCCTCTGCTCTCATCAGCAATCAAACC TGAAGATTCTGCAAATTGGGTGCAATGAAATCGGAGATGTGGGTGTGCAGCTGTTGTG TCGGGCTCTGACGCATACGGATTGCCGCTTAGAGATTCTTGGGTGGAAGAATGTGGG TTAACGAGCACCTGCTGTAAGGATCTCGCGTCTGTTCTCACCTGCAGTAAGACCCTGC AGCAGCTCAACCTGACCTTGAACACCTTGGACCACACAGGGGTGGTTGTACTCTGTGA GGCCCTGAGACACCCAGAGTGTGCCCTGCAGGTGCTCGGGCTGAGAAAACTGATTTT GATGAGGAAACCCAGGCACTTCTGACGGCTGAGGAAGAGAGAAATCCTAACCTGACCA TCACAGACGACTGTGACACAATCACAAGGGTAGAGATCTGA		
	ORF Start: ATG at 124	ORF Stop: TGA at 2881	
	SEQ ID NO: 4	919 aa	MW at 103966.7kD
NOV2a, CG100689-01 Protein Sequence	MQECLTLWVFSPLALTDSGYTKTYQAHAKQKFSRLWSSKSVTEIHLVFEEVVKQEECD HLDRLFAPKEAGKQPRTVIIQGPQGIGKTTLLMKLMMAWS DNKIFRDRFLYTFYFCCR ELRELPP TSLADLISREWPDPAPITEIVSQPERLLFVIDSFEELOQGLNEPDS D LCG DLMEKRPVQVLLSSLRLKKMLPEASLLIAIKPVC PKELRDQVTISEIYQPRGFNESDR LVYFCCFFKDPKRAMEAFNLVRESEQLFSICQIPLLCWILCTSLKQEMQKGKDLALTC QSTTSVYSSFVFNLF TPEGAE GPTPQTQHQLKALCSLAAEGMWTDTFEFCEDDLRRNG VVDADIPALLG TKILLKYGERESSYVFLHVCIQEFCAALFYLLKSHLDHPHPAVRCVQ ELLVANFEKARRAHWIFLGCF LTGLLNKKEQEKLD A FFGFQLSQEI KQQIHQCLKSLG ERGNPQGQVDSLAI FYCLFEMQDPAFVKQAVNLLQEANFHI IDNVDLVVSAYCLKYCS		

SLRKLCFSVQNVFKKEDEHSSTSDYSLICWHHICSVLTTSGLHRELQVQDSTLSESTF VTWCNQLRHPSCRLQKLGINNVSFSGQSVLLFEVLIFYQPDLYLSFTLTSLRDDRIS LCDALNYPAGNVKELALVNCHLSPIDCEVLAGLLTNNKKLTYLNVSCNQLDTGVPLLC EALCSPDTVLVYMLAFCHLSEQCCEYISEMLLRNKSRYLDLSANVLKDEGLKTLCE ALKHPDCCLDSLCLVKCFITAAGCEDLASALISNQNLIKILQIGCNEIGDVGVLQLCRA LTHTDCRLEILGLEECGLTSTCKDLASVLTCSKTLQQLNLTNTLDHTGVVVLCEAL RHPECALQVLGLRKTDFTDEETQALLTAEERNPNLTITDDCDTITRVEI

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

Table 2B. Protein Sequence Properties NOV2a	
PSort analysis:	0.6000 probability located in nucleus; 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:	Cleavage site between residues 17 and 18

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

Table 2C. Geneseq Results for NOV2a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM50328	Human nucleotide binding site protein NBS-5 - Homo sapiens, 858 aa. [WO200183753-A2, 08-NOV-2001]	33..882 1..850	849/850 (99%) 850/850 (99%)	0.0
AAU07878	Polypeptide sequence for mammalian Spg65 - Mammalia, 748 aa. [WO200166752-A2, 13-SEP-2001]	165..907 5..744	375/743 (50%) 528/743 (70%)	0.0
AAE07514	Human PYRIN-1 protein - Homo sapiens, 1034 aa. [WO200161005-A2, 23-AUG-2001]	20..907 134..1028	320/926 (34%) 491/926 (52%)	e-146
AAG65895	Amino acid sequence of GSK gene Id 97078 - Homo sapiens, 1062 aa. [WO200172961-A2, 04-OCT-2001]	75..907 208..1043	301/849 (35%) 460/849 (53%)	e-137
AAE07513	Human nucleotide binding site 1 (NBS-1) protein - Homo sapiens, 1033 aa. [WO200161005-A2, 23-AUG-2001]	75..907 180..1014	299/849 (35%) 459/849 (53%)	e-134

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 2D.

Table 2D. Public BLASTP Results for NOV2a				
Protein Accession Number	Protein/Organism/Length	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96MN2	CDNA FLJ32126 FIS, CLONE PEBLM2000112, WEAKLY SIMILAR TO HOMO SAPIENS NUCLEOTIDE-BINDING SITE PROTEIN 1 MRNA - Homo sapiens (Human), 919 aa.	1..919 1..919	918/919 (99%) 918/919 (99%)	0.0
Q96MN2	NACHT-, LRR- and PYD-containing protein 4 (PAAD and NACHT-containing protein 2) (PYRIN-containing APAF1-like protein 4) (Ribonuclease inhibitor 2) - Homo sapiens (Human), 994 aa.	18..919 93..994	900/902 (99%) 901/902 (99%)	0.0
AAL88672	RIBONUCLEASE INHIBITOR 2 - Homo sapiens (Human), 916 aa.	18..919 15..916	894/902 (99%) 897/902 (99%)	0.0
CAD19386	SEQUENCE 7 FROM PATENT WO0183753 - Homo sapiens (Human), 858 aa (fragment).	33..882 1..850	849/850 (99%) 850/850 (99%)	0.0
Q99MW0	RIBONUCLEASE/ANGIOGENIN INHIBITOR 2 - Mus musculus (Mouse), 748 aa.	165..907 5..744	374/743 (50%) 528/743 (70%)	0.0

Pfam analysis predicts that the NOV2a protein contains the domains shown in the Table 2E.

Table 2E. Domain Analysis of NOV2a			
Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value
SRP54	71..93	11/23 (48%) 17/23 (74%)	0.18

Example 3.

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

Table 3A. NOV3 Sequence Analysis			
	SEQ ID NO: 5	2142 bp	
NOV3a, CG100760-01 DNA Sequence	TATTATTCAGCAAACAATCTCAATGTGTTCTGATGGGAGAGAGAGCATCTGGAAAAA CTATTGTTATAAATCTGGCTGTGTTGAGGTGGATCAAGGGTGAGATGTGGCAGAACAT GATCTCGTACGTCGTTTACCTCACTGCTCACGAAATAAACCAGATGACCAACAGCAGC TTGGCTGAGCTAATCGCCAAGGACTGGCCTGACGGCCAGGCTCCCATTGCAGACATCC TGTCTGATCCCAAGAACTCCTTTTTCATCCTCGAGGACTTGGACAACATAAGATTCTGA GTTAAATGTCAATGAAAGTGCTTTGTGTAGTAACAGCACCCAGAAAGTTCCCATTCCA GTTCTCCTGGTCAGTTTGCTGAAGAGAAAAATGGCTCCAGGCTGCTGGTTCCTCATCT CCTCAAGGCCACACGTGGGAATAATGTAAAAACGTTCTTGAAAGAGGTAGATTGCTG CACGACCTTGCAGCTGTCTGAATGGGAAGAGGGAGATATATTTTAACTCTTTCTTTAAA GACCGCCAGAGGGCGTCGGCAGCCCTCCAGCTTGTACATGAGGATGAAATACTCGTGG GTCTGTGCCGAGTCGCCATCTTATGCTGGATCACGTGTACTGTCCTGAAGCGGCAGAT GGACAAGGGGCGTGACTTCCAGCTCTGCTGCCAAACACCCACTGATCTACATGCCAC TTTCTTGCTGATGCGTTGACATCAGAGGCTGGACTTACTGCCAATCAGTATCACCTAG GTCTCCTAAAACGTCTGTGTTTGCTGGCTGCAGGAGGACTGTTTCTGAGCACCTTGAA TTTCAGTGGTGAAGACCTCAGATGTGTTGGGTTTACTGAGGCTGATGTCTCTGTGTTG CAGGCCGCGAATATTCTTTTGCCGAGCAACACTCATAAAGACCGTTACAAGTTCATAC ACTTGAACGTCCAGGAGTTTTGTACAGCCATTGCATTTCTGATGGCAGTACCCAACTA TCTGATCCCCTCAGGCAGCAGAGAGTATAAAGAGAAGAGAGAACAATACTCTGACTTT AATCAAGTGTTTACTTTTCAATTTTTGGTCTTCTAAATGCAAACAGGAGAAAGATTCTTG AGACATCCTTTGGATACCAGCTACCGATGGTAGACAGCTTCAAGTGGTACTCGGTGGG ATACATGAAACATTTGGACCGTGACCCGGAAGTTGACGCACCATATGCCTTTGTTT TACTGTCTCTATGAGAATCGGGAAGAAGAATTTGTGAAGACGATTGTGGATGCTCTCA TGGAGGTTACAGTTTACCTTCAATCAGACAAGGATATGATGGTCTCATTATACTGTCT GGATTACTGCTGTCACCTGAGGACACTTAAGTTGAGTGTTCAGCGCATCTTTCAAAC AAACTGGAGAAATGCAACTTGTGCGCAGCCAGCTGTCAGGACCTAGCCTTGTTTCTCA CCAGCATCCAACACGTAACCTCGATTGTGCCTGGGATTTAATCGGCTCCAAGATGATGG CATAAAGCTATTGTGTGCGGCCCTGACTCACCCCAAGTGTGCCTTAGAGAGACTGGAG CTCTGGTTTTTGCCAGCTGGCAGCACCCGCTTGCAAGCACTTGTGAGATGCTCTCCTGC AGAACAGGAGCCTGACACACCTGAATCTGAGCAAGAACAGCCTGAGAGACGAGGGAGT CAAGTTCCTGTGTGAGGCCTTGGGTGCCCCAGATGGTAACCTGCAGAGCCTGAGTTTG TCAGGTTGTTCTTTTACAAGAGAGGGCTGTGGAGAGCTGGCTAATGCCCTCAGCCATA ATCATAATGTGAAAATCTTGGATTTGGGAGAAAATGATCTTCAGGATGATGGAGTGAA GCTACTGTGTGAGGCTCTGAAACCACATCGTGCATTGCACACACTTGGGTTGGCGAAA TGCAATCTGACAACTGCTTGCTGCCAGCATCTCTTCTCTGTTCTCAGCAGCAGTAAGA GCCTGGTCAATCTGAACCTTCTAGGCAATGAATTGGATACTGATGGTGTCAAGATGCT ATGTAAGGCTTTGAAAAAGTCGACATGCAGGCTGCAGAAACTCGGGTAAACCTCACTG ACTTTTCTGCAGGGGAGAACATACAGGGACAAGGCTAGATTGACTAGGCTTCTA		
	ORF Start: ATG at 34	ORF Stop: TAA at 2077	
	SEQ ID NO: 6	681 aa	MW at 76724.1kD
NOV3a, CG100760-01 Protein Sequence	MGERASGKTIVINLAVLRWIKGEMWQNMISYVVHLTAHEINQMTNSSLAELIAKDWD GQAPIADILSDPKLLFILEDLNIRFELNVNESALCSNSTQKVPIPVLLVSLKLRKM APGCWFLISSRPTRGNNVKTFLKEVDCCTTLQLSNGKREIYFNSFFKDRQRASAALQL VHEDEILVGLCRVAILCWITCTVLKRQMDKGRDFQLCCQTPDLHAHFLADALTSEAG LTANQYHLGLLKLRLCLLAAGGLFLSTLNFSGEDLRCVGFTEADVSVLQAANILLPSNT HKDRYKFIHLNVQEFCTAIAFLMAVPNYLIPSGSREYKEKREQYSDFNQVFTFIFGLL NANRRKILETSFGYQLPMVDSFKWYSVGYMKHLDRDPEKLTHHMPLFYCLYENREEEF VKTIVDALMEVTVYLQSDKDMVSLYCLDYCCHLRTLKLSVQRIQNKLEKCNLSAAS		

CQDLALFLTSIQHVTRLCLGFNRLQDDGIKLLCAALTHPKCALERLELWFCQLAAPAC KHLSDALLQNRSLTHLNLSKNSLRDEGVKFLCEALGRPDGNLQSLSLSGCSFTREGCG ELANALSHNHNKILDLGENDLQDDGVKLLCEALKPHRALHTLGLAKCNLTACCQHL FSVLSSSKSLVNLNLLGNELDTDGVKMLCKALKKSTCRLQKLG

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

Table 3B. Protein Sequence Properties NOV3a	
PSort analysis:	0.8200 probability located in endoplasmic reticulum (membrane); 0.1900 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Cleavage site between residues 23 and 24

A search of the NOV3a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 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
AAM50330	Human nucleotide binding site protein NBS-3 - Homo sapiens, 875 aa. [WO200183753-A2, 08-NOV-2001]	1..681 116..859	539/745 (72%) 587/745 (78%)	0.0
AAM50326	Human nucleotide binding site protein NBS-3 - Homo sapiens, 631 aa. [WO200183753-A2, 08-NOV-2001]	1..510 116..631	469/517 (90%) 479/517 (91%)	0.0
AAM50328	Human nucleotide binding site protein NBS-5 - Homo sapiens, 858 aa. [WO200183753-A2, 08-NOV-2001]	2..681 48..792	247/750 (32%) 381/750 (49%)	e-105
AAE07514	Human PYRIN-1 protein - Homo sapiens, 1034 aa. [WO200161005-A2, 23-AUG-2001]	2..680 224..944	238/729 (32%) 362/729 (49%)	e-100
ABG03924	Novel human diagnostic protein #3915 - Homo sapiens, 952 aa. [WO200175067-A2, 11-OCT-2001]	2..680 178..908	228/741 (30%) 334/741 (44%)	7e-78

- 5 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
CAD19388	SEQUENCE 15 FROM PATENT WO0183753 - Homo sapiens (Human), 875 aa.	1..681 116..859	539/745 (72%) 587/745 (78%)	0.0
CAD19384	SEQUENCE 3 FROM PATENT WO0183753 - Homo sapiens (Human), 631 aa (fragment).	1..510 116..631	469/517 (90%) 479/517 (91%)	0.0
CAD19386	SEQUENCE 7 FROM PATENT WO0183753 - Homo sapiens (Human), 858 aa (fragment).	2..681 48..792	247/750 (32%) 381/750 (49%)	e-104
Q96MN2	CDNA FLJ32126 FIS, CLONE PEBLM2000112, WEAKLY SIMILAR TO HOMO SAPIENS NUCLEOTIDE-BINDING SITE PROTEIN 1 MRNA - Homo sapiens (Human), 919 aa.	2..681 80..824	247/750 (32%) 381/750 (49%)	e-104
Q96MN2	NACHT-, LRR- and PYD-containing protein 4 (PAAD and NACHT-containing protein 2) (PYRIN-containing APAF1-like protein 4) (Ribonuclease inhibitor 2) - Homo sapiens (Human), 994 aa.	2..681 155..899	247/750 (32%) 381/750 (49%)	e-104

Example 4.

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

Table 4A. NOV4 Sequence Analysis		
	SEQ ID NO: 7	782 bp
NOV4a, CG100851-02 DNA Sequence	TCTCAAGGGATAATCACTAAATTCTGCCGAAAGGACTGAGGAACGGTGCCTGGAAAAG GGCAAGAATATCACGGCATGGGCATGAGTAGCTTGAAACTGCTGAAGTATGTCCTGTT TTTCTTCAACTTGCTCTTTTGGATCTGTGGCTGCTGCATTTTGGGCTTTGGGATCTAC CTGCTGATCCACAACAACCTTCGGAGTGCTCTTCCATAACCTCCCCTCCCTCACGCTGG GCAATGTGTTTGTTCATCGTGGGCTCTATCAAGGAAAACAAGTGTCTGCTTATGTCGTT CTTCATCCTGCTGCTGATTATCCTCCTTGCTGAGGTGACCTTGGCCATCCTGCTCTTT GTATATGAACAGAAGCTGAATGAGTATGTGGCTAAGGGTCTGACCGACAGCATCCACC GTTACCACTCAGACAATAGCACCAAGGCAGCGTGGGACTCCATCCAGTCATTTCTGCA GTGTTGTGGTATAAATGGCACGAGTGATGGGACCAGTGGCCACCAGCATCTTGCCCC TCAGATCGAAAAGTGGAGGGTTGCTATGCGAAAGCAAGACTGTGGTTTCATTCCAATT TCCTGTATATCGGAATCATCACCATCTGTGTATGTGTGATTGAGGTGTTGGGGATGTC CTTTGCACTGACCCTGAAGTCCAGATTGACAAAACCAGCCAGACCATAGGGCTATGA	

	TCTGCAGTAGTTCTGTGGTGAAGAGACTTGTTCATCTCCTGGAAATGCAAAACCATT TATAGCATGAGCCCTACATGATCATCAG		
	ORF Start: ATG at 76	ORF Stop: TGA at 694	
	SEQ ID NO: 8	206 aa	MW at 22888.8kD
NOV4a, CG100851-02 Protein Sequence	MGMSLKLKLYVLFFFNLLFWICGCCILGFGIYLLIHNHFGVLFHNLPSLTGLGNVFVI VGSIKENKCLLMSFFILLIILLAEVTLAILLFVYEQKLNEYVAKGLTDSIHRYHSDN STKAAWDSIQSFLQCCGINGTSDGTSGPPASCPSPDRKVEGCGYAKARLWFHSNFLYIGI ITICVVCVIEVLGMSFALTLCQIDKTSQTIGL		

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

Table 4B. Protein Sequence Properties NOV4a	
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:	Cleavage site between residues 32 and 33

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

Table 4C. Geneseq Results for NOV4a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AA Y96141	Human haematopoietic CD53 - Homo sapiens, 219 aa. [US6111093-A, 29-AUG-2000]	1..206 1..219	205/219 (93%) 205/219 (93%)	e-115
AAB58136	Lung cancer associated polypeptide sequence SEQ ID 474 - Homo sapiens, 231 aa. [WO200055180-A2, 21-SEP-2000]	1..206 13..231	205/219 (93%) 205/219 (93%)	e-115
AAW89152	Human CD53 antigen - Homo sapiens, 219 aa. [US5849898-A, 15-DEC-1998]	1..206 1..219	205/219 (93%) 205/219 (93%)	e-115
AAW80455	Human CD53 antigen - Homo sapiens, 219 aa. [US5830731-A, 03-NOV-1998]	1..206 1..219	205/219 (93%) 205/219 (93%)	e-115
AAR91446	Human CD53 antigen - Homo sapiens, 219 aa. [US5506126-A, 09-APR-1996]	1..206 1..219	205/219 (93%) 205/219 (93%)	e-115

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 4D.

Table 4D. Public BLASTP Results for NOV4a				
Protein Accession Number	Protein/Organism/Length	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P19397	Leukocyte surface antigen CD53 (Cell surface glycoprotein CD53) - Homo sapiens (Human), 219 aa.	1..206 1..219	205/219 (93%) 205/219 (93%)	e-115
AAH21310	CD53 ANTIGEN - Mus musculus (Mouse), 219 aa.	1..206 1..219	168/219 (76%) 183/219 (82%)	6e-95
Q61451	Leukocyte surface antigen CD53 (Cell surface glycoprotein CD53) - Mus musculus (Mouse), 218 aa.	2..206 1..218	167/218 (76%) 182/218 (82%)	2e-94
A39574	leukocyte antigen OX-44 - rat, 219 aa.	1..206 1..219	164/219 (74%) 183/219 (82%)	7e-94
P24485	Leukocyte surface antigen CD53 (Cell surface glycoprotein CD53) (Leukocyte antigen MRC OX-44) - Rattus norvegicus (Rat), 218 aa.	2..206 1..218	163/218 (74%) 182/218 (82%)	3e-93

Pfam analysis predicts that the NOV4a protein contains the domains shown in the Table 4E.

Table 4E. Domain Analysis of NOV4a			
Pfam Domain	NOV4a Match Region	Identities/ Similarities for the Matched Region	Expect Value
transmembrane4	10..36	17/27 (63%) 27/27 (100%)	4.4e-08
transmembrane4	58..197	52/202 (26%) 120/202 (59%)	1.2e-44

5

Example 5.

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

Table 5A. NOV5 Sequence Analysis			
	SEQ ID NO: 9		1719 bp
NOV5a, CG101068-01 DNA Sequence	ATGCGGACGCCGGTGGTGATGACGCTGGGCATGGTGTGGCGCCCTGCGGGCTCCTGC TCAACCTGACCGGCACCCTGGCGCCCGGCTGGCGGCTGGTGAAGGGCTTCCTGAACCA GCCAGTGGACGTGGAGTTGTACCAGGGCCTGTGGGACATGTGTGCGGAGCAGAGCAGC CGCGAGCGCGAGTGCGGCCAGACGGACCAGTGGGGCTACTTCGAGGCCAGCCCGTGC TGGTGGCGCGGGCACTCATGGTCACCTCGCTGGCCGCCACGGTCCTGGGGCTTCTGCT GGCGTCGCTGGGCGTGCCTGCTGGCAGGACGAGCCCAACTTCGTGCTGGCAGGGCTC TCGGGCGTCGTGCTCTTCGTGCTGGCCTCCTCGGCCTCATCCCGGTGTCCTGGTACA ACCACTTCTTGGGGGACCGCGACGTGCTGCCCCGCCCGGCCAGCCCGGTACAGGTGCA GGTCAGCTACAGCCTGGTCCTGGGCTACCTGGGCAGCTGCCTCCTGCTGCTGGGCGGC TTCTCGCTGGCGCTCAGCTTCGCGCCCTGGTGCGACGAGCGTTGTGCGCCGCCGCCGA AGGGACCCTCCGCCGGGCCTCGCCGCAGCAGCGTCAGCACCATCCAAGTGGAGTGGCC CGAGCCCGACCTGGCGCCCCGCCATCAAGTACTACAGCGACGGCCAGCACCGACCGCCG CCTGCCCAGCACCGCAAGCCCAAGCCCAAGCCCAAGGTGCGCTTCCCCATGCCGCGGC CGCGGCCCAAGGCCTACACCAACTCGGTGGACGTCTTCGACGGGGAGGGGTGGGAGTC CCAGGACGCTCCCTCGTGCAGCACCCACCCCTGCGACAGCTCGCTGCCCTGCGACTCC GACCTCTAGACGCTTGTAGAGCCTGGGGGGCGCCGGGTGGCAAAGGACTCACCCCCGC ACAGGCCCCGCTTGGCTTCGAGTTGGAACCCGGACACTTGCCCCTCACTGGTGTGGATG GAAATCTGCCTTTCGTGGGACCAACAGGACTCCTTGACGATTAGTTCAGGTTGGGT TTGGTTTTCTTCTTAAAGAGTTTAGTTTTCTCTCCAGAGGGATCAGGGTCCTCTTAG GGAGTGACGGGCTTTTCATATATTTTGTCTGAAGAATATATGGAAAGGGTGGCATTG CGTCACGTGGACCAGGGACAGTGCTGAAATCAGCAGTGCTCAGAAACAATTTAACATG TTGAAACGACAATATTCTAAAATACTGATGAATCTTGCATCAATATAATTATTGGGTT TTTTTTCTTTTTCTGCTGTATAACTCCTTGCCATGCAAACTCTCAAGAGGCCAATAT ATTCCTGGCCATGTTTGAATGAGCCTCTTAAAATAAACTTAGAGCCATGCAAAATGCCA GCAGCTTAATGGATTTTCATGGAATGAAATACCGTGATTAATCATAGCTACATATCAT TGCATAAATGGGATTTATCTTTTTTCTCACTTATTTTTGCGGTGAAAGTCGAGGGCAT GCAAGAGTTTCTCTTCCAGAAGCCAAGAGGAGAACAAAGGTCCTAATGCTGTACTATT CCACCCTTTGGACGCCTCATCCAGGACGCAGAGGACTCTAGGTTTAACATTTTGTACA AAATGGAACCTGTTAATCATATTAAAGCACATATGTATATATCTTTTATTATAAATA AAATTTTAAACAATAGTTTCAGTATAGCCACAAAAA		
	ORF Start: ATG at 1		ORF Stop: TAG at 877
	SEQ ID NO: 10		292 aa MW at 31914.5kD
NOV5a, CG101068-01 Protein Sequence	MRTPVVMTLGMVLAPCGLLLNLTGTLAPGWRLVKGFLNQPDVELYQGLWDMCREQSS RERECGQTDQWGYFEAQPVLVARALMVTSLAATVLGLLLASLGVRWQDEPNFVLAGL SGVVL FVAGLLGLIPVSWYNHFLGDRDVLPA PASPVTVQVSYSLVLGYLGSCLLLLGG FSLALS FAPWCDERCRRRRKGPSAGPRRSSVSTIQVEWPEPDLAPA IKYYS DGQHRPP PAQHRKPKPKPKVGFPMRPRPKAYTNSVDVLDGEGWESQDAPSCSTHPCDSSLPCDS DL		

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

Table 5B. Protein Sequence Properties NOV5a	
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:	Cleavage site between residues 28 and 29

A search of the NOV5a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 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
AAB64401	Amino acid sequence of human intracellular signalling molecule INTRA33 - Homo sapiens, 217 aa. [WO200077040-A2, 21-DEC-2000]	9..206 9..209	74/206 (35%) 101/206 (48%)	1e-16
AAG75467	Human colon cancer antigen protein SEQ ID NO:6231 - Homo sapiens, 210 aa. [WO200122920-A2, 05-APR-2001]	6..187 7..192	59/188 (31%) 92/188 (48%)	2e-13
ABB50278	Claudin 4 ovarian tumour marker protein, SEQ ID NO:45 - Homo sapiens, 209 aa. [WO200175177-A2, 11-OCT-2001]	6..187 6..191	59/188 (31%) 92/188 (48%)	2e-13
AAB43133	Human ORFX ORF2897 polypeptide sequence SEQ ID NO:5794 - Homo sapiens, 209 aa. [WO200058473-A2, 05-OCT-2000]	6..187 6..191	59/188 (31%) 92/188 (48%)	2e-13
ABB50396	Human secreted protein encoded by gene 96 SEQ ID NO:344 - Homo sapiens, 202 aa. [WO200162891-A2, 30-AUG-2001]	9..187 1..183	59/185 (31%) 91/185 (48%)	2e-13

5 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
Q96B33	SIMILAR TO RIKEN CDNA 2310014B08 GENE - Homo sapiens (Human), 268 aa (fragment).	28..292 2..268	252/267 (94%) 254/267 (94%)	e-147
Q9D7D7	2310014B08RIK PROTEIN (RIKEN CDNA 2310014B08 GENE) - Mus musculus (Mouse), 296 aa.	1..292 1..296	230/296 (77%) 248/296 (83%)	e-135
O95484	Claudin-9 - Homo sapiens (Human), 217 aa.	9..206 9..209	74/206 (35%) 101/206 (48%)	4e-16
Q9Z0S7	Claudin-9 - Mus musculus (Mouse), 217 aa.	9..206 9..209	71/206 (34%) 99/206 (47%)	1e-14
Q98SR2	CLAUDIN-3 - Gallus gallus (Chicken), 214 aa.	10..206 9..207	64/202 (31%) 99/202 (48%)	1e-13

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
PMP22_Claudin	3..177	40/194 (21%) 108/194 (56%)	0.00018

5

Example 6.

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

Table 6A. NOV6 Sequence Analysis		
	SEQ ID NO: 11	2369 bp
NOV6a, CG101231-01 DNA Sequence	CGGCCGGAGCGCCGAGGCCCGGCCATGGCCACCACCAGCACCACGGGCTCCACCCTGC TGCAGCCCCCTCAGCAACGCCGTGCAGCTGCCCATCGACCAGGTCAACTTTGTAGTGTG CCAACTCTTTGCCTTGCTAGCAGCCATTTGGTTTCGAACCTTATCTACATTCAAGCAAA ACTAGCTCTTTTATAAGACATGTAGTTGCTACCCTTTTGGGCCTTTATCTTGCACTTT	

	TTTGCTTTGGATGGTATGCCTTACACTTTCTTGTACAAAGTGGAATTTCTACTGTAT CATGATCATCATAGGAGTGGAGAACATGCACAATTACTGCTTTGTGTTTGCTCTGGGA TACCTCACAGTGTGCCAAGTTACTCGAGTCTATATCTTTGACTATGGACAATATTCTG CTGATTTTTTCAGGCCCAATGATGATCATTACTCAGAAGATCACTAGTTTGGCTTGCGA AATTCATGATGGGATGTTTCGGAAGGATGAAGAACTGACTTCCTCACAGAGGGATTTA GCTGTAAGGCGCATGCCAAGCTTACTGGAGTATTTGAGTTACAACCTGTAACCTCATGG GGATCCTGGCAGGCCCACTTTGCTCTTACAAAGACTACATTACTTTCATTGAAGGCAG ATCATAACCATATCACACAATCTGGTGAAAATGGAAAAGAAGAGACACAGTATGAAAGA ACAGAGCCATCTCCAAATAGTGCGGTTGTTTCAGAAGCTCTTAGTTTGTGGGCTGTCCT TGTTATTTCACTTGACCATCTGTACAACATTACCTGTGGAGTACAACATTGATGAGCA TTTTCAAGCTACAGCTTCGTGGCCAACAAAGATTATCTATCTGTATATCTCTCTTTTG GCTGCCAGACCCAAATACTATTTTGCATGGACGCTAGCTGATGCCATTAATAATGCTG CAGGCTTTGGTTTCAGAGGGTATGACGAAAATGGAGCAGCTCGCTGGGACTTAATTTT CAATTTGAGAATTCAACAAATAGAGATGTCAACAAGTTTCAAGATGTTTCTTGATAAT TGGAATATTCAGACAGCTCTTTGGCTCAAAAGGGTGTGTTATGAACGAACCTCCTTCA GTCCAACATATCCAGACGTTTATTCTCTCTGCCATTTGGCACGGGGTATACCCAGGATA TTATCTAACGTTTCTAACAGGGGTGTTAATGACATTAGCAGCAAGAGCTGTAAGAAAT AACTTTAGACATTATTTTATTGAACCTTCCCAACTGAAATTATTTTATGATGTTATAA CATGGATAGTAACTCAAGTAGCAATAAGTTACACAGTTGTGCCATTTGTGCTTCTTTC TATAAAACCATCACTCACGTTTTACAGCTCCTGGTATTATTGCCTGCACATTCTTGGT ATCTTAGTATTATTGTTGTTGCCAGTAAAAAACTCAAAGAAGAAAGAATACACATG AAAACATTCAGCTCTCACAATCCAAAAGTTTGATGAAGGAGAAAATTCTTTGGGACA GAACAGTTTTTCTACAACAAACAATGTTTGCAATCAGAATCAAGAAATAGCCTCGAGA CATTCATCACTAAAGCAGTGATCGGGAAGGCTCTGAGGGCTGTTTTTTTTTTTGTATG TTAACAGAAACCAATCTTAGCACCTTTTCAAGGGGTTTGAGTTTGTGGAAGCAGT TAACTGGGGGGAAATGGACAGTTATAGATAAGGAATTTCTGTACACCAGATTGGAAA TGGAGTGAAACAAGCCCTCCCATGCCATGTCCCCGTGGGCCACGCCTTATGTAAGAAT ATTTCCATATTTTCAAGTGGGCACTCCCAACCTCAGCACTTGTCCGTAGGGTCACACGCG TGCCCTGTTGCTGAATGTATGTTGCGTATCCCAAGGCACTGAAGAGGTGGAATAATA TCGTGTCAATCTGGATGATAGAGAGAAATTAACCTTTTCCAAATGAATGTCTTGCCTTA AACCCTCTATTTCTTAAATATTGTTCTTAAATGGTATTTTCAAGTGAATATTGTGA GAACGCTACTGCAGTAGTTGATGTTGTGTGCTGTAAAGGATTTTAGGAGGAATTTGAA ACAGGATATTTAAGAGTGTGGATATTTTAAATGCAATAAACATCTCAGTATTTGAA GGGTTTTCTTAAAGTATGTCAAATGACTACAATCCATAGTGAACTGTAAACAGTAAT GGACGCCAAATTATAGGTAGCTGATTTTGTGCTGGAGAGTTTAATTACCTTGTGCAGTCA AAGAGCGCTTCCAGAAGGAATCTCTTAAACATAATGAGAGGTTTGGTAATGTGATAT TTTAAGCTTACTCTTTTTCTTAAAGAGAGAGGTGACGAAGGAAGGCAG		
	ORF Start: ATG at 25	ORF Stop: TGA at 1585	
	SEQ ID NO: 12	520 aa	MW at 59480.0kD
NOV6a, CG101231-01 Protein Sequence	MATTSTTGSTLLQPLSNAVQLPIDQVNFVVCQLFALLAAIWFRTYLHSSKTSSFIRHV VATLLGLYLALFCFGWYALHFLVQSGISYCIIMIIIGVENMHNYCFVFALGYLTVCQVT RVIYIFDYGQYSADFSGPMMIITQKITSLACEIHDGMFRKDEELTSSQRDLAVRRMPSL LEYLSYCNCFMGILAGPLCSYKDYITFIEGRSYHITQSGENGKEETQYERTEPSPNSA VVQKLLVCGLSLLFHLTICTTLPVEYNIDEHFQATASWPTKIIYLYISLLAARPKYYF AWTLADAINNAAGFGFRGYDENGARWDLISNLRIQQIEMSTSFKMFLDNWNIQTALW LKRVCYERTSFSPITQTFILSAIWHGVYPGYLTFLTGLVMTLAARAVRNNFRHYFIE PSQKLKLFYDVITWIVTQVAISYTVVPFVLLSIKPSLTFYSSWYYCLHILGILVLLLLP VKKTQRRKNTHENIQLSQSKKFDEGENSLGQNSFSTTNVNCNQNEIASRHSLSLKQ		
	SEQ ID NO: 13	2270 bp	
NOV6b, CG101231-02 DNA Sequence	CGGCCGGAGCGCCGAGGCCCGGCCATGGCCACCACCAGCACCACGGGCTCCACCCTGC TGCAGCCCCTCAGCAACGCCGTGCAGCTGCCCATCGACCAGGTCAACTTTGTAGTGTG CCAACCTCTTTGCCTTGCTAGCAGCCATTTGGTTTCGAACTTATCTACATTCAAGCAAA ACTAGCTCTTTTATAAGACATGTAGTTGCTACCCCTTTTGGGCCTTTATCTTGCACTTT TTTGCTTTTGGATGGTATGCCTTACACTTTCTTGTACAAAGTGGAATTTCTACTGTAT CATGATCATCATAGGAGTGGAGAACATGCAGCCAATGATGATCATTACTCAGAAGATC ACTAGTTTGGCTTGCGAAATTCATGATGGGATGTTTCGGAAGGATGAAGAACTGACTT CCTCACAGAGGGATTTAGCTGTAAGGCGCATGCCAAGCTTACTGGAGTATTTGAGTTA CAACTGTAACCTTCATGGGGATCCTGGCAGGCCCACTTTGCTCTTACAAAGACTACATT		

	ACTTTCATTGAAGGCAGATCATACCATATCACACAATCTGGTGAAAATGGAAAAGAAG AGACACAGTATGAAAGAACAGAGCCATCTCCAAATAGTGCGGTTGTTTCAGAAGCTCTT AGTTTGTGGGCTGTCCTTGTTATTTCACTTGACCATCTGTACAACATTACCTGTGGAG TACAACATTGATGAGCATTTTCAAGCTACAGCTTCGTGGCCAACAAAGATTATCTATC TGTATATCTCTCTTTTGGCTGCCAGACCCAAATACTATTTTGCATGGACGCTAGCTGA TGCCATTAATAATGCTGCAGGCTTTGGTTTCAGAGGGTATGACGAAAATGGAGCAGCT CGCTGGGACTTAATTTCCAATTTGAGAATTCAACAAATAGAGATGTCAACAAGTTTCA AGATGTTTCTTGATAATTGGAATATTCAGACAGCTCTTTGGCTCAAAAGGGTGTGTTA TGAACGAACCTCCTTCAGTCCAATATCCAGACGTTTCTCTCTGCCATTTGGCAC GGGGTATACCCAGGATATTATCTAACGTTTCTAACAGGGGTGTTAATGACATTAGCAG CAAGAGCTGTAAGAAATAACTTTAGACATTATTTTATTGAACCTTCCCAACTGAAATT ATTTTATGATGTTATAACATGGATAGTAACTCAAGTAGCAATAAGTTACACAGTTGTG CCATTTGTGCTTCTTTCTATAAAACCATCACTCACGTTTTTACAGCTCCTGGTATTATT GCCTGCACATTCTTGGTATCTTAGTATTATTGTTGTTGCCAGTAAAAAAACTCAAAG AAGAAAGAATACACATGAAAACATTCAGCTCTCACAATCCAAAAGTTTGATGAAGGA GAAAATTCTTTGGGACAGAACAGTTTTTCTACAACAAACAATGTTTGCAATCAGAATC AAGAAATAGCCTCGAGACATTCATCACTAAAGCAGTGATCGGGGAAGGCTCTGAGGGCT GTTTTTTTTTTTTTGATGTTAACAGAAACCAATCTTAGCACCTTTTCAAGGGGTTTGAG TTTGTTGGAAAAGCAGTTAACTGGGGGGAAATGGACAGTTATAGATAAGGAATTTCTCT GTACACCAGATTGGAAATGGAGTGAAACAAGCCCTCCCATGCCATGTCCCCGTGGGCC ACGCCTTATGTAAGAATATTTCCATATTTTCACTGGGCACTCCCAACCTCAGCACTTGT CCGTAGGGTCACACGCGTGCCCTGTTGCTGAATGTATGTTGCGTATCCCAAGGCACTG AAGAGGTGGAAAATAATCGTGTCAATCTGGATGATAGAGAGAAATTAACTTTTCCAA ATGAATGTCTTGCTTAAACCCTCTATTTCTTAAATATTGTTCTTAAATGGTATTTT CAAGTGTAATATTGTGAGAACGCTACTGCAGTAGTTGATGTTGTGTGCTGTAAAGGAT TTTAGGAGGAATTTGAAACAGGATATTTAAGAGTGTGGATATTTTAAATGCAATAA ACATCTCAGTATTTGAAGGGTTTTCTTAAAGTATGTCAAATGACTACAATCCATAGTG AAAGTGTAAACAGTAATGGACGCCAAATTATAGGTAGCTGATTTTGCTGGAGAGTTTA ATTACCTTGTGCAGTCAAAGAGCGCTTCCAGAAGGAATCTCTTAAACATAATGAGAG GTTTGGTAATGTGATATTTTAAGCTTACTCTTTTCTTAAAGAGAGAGGTGACGAAG GAAGGCAG		
	ORF Start: ATG at 25	ORF Stop: TGA at 1486	
	SEQ ID NO: 14	487 aa	MW at 55677.7kD
NOV6b, CG101231-02 Protein Sequence	MATTTSTTGSTLLQPLSNAVQLPIDQVNFVVCQLFALLAAIWFRTYLHSSKTSSFIRHV VATLLGLYLALFCFGWYALHFLVQSGISYCMIIIGVENMQPMMIITQKITSLACEIH DGMFRKDEELTSSQRDLAVRRMPSLLEYLSYNCFMGLILAGPLCSYKDYITFIEGRSY HITQSGENGKEETQYERTEPSPNSAVVQKLLVCGLSLLFHLTICTTLPVEYNIDEHFQ ATASWPTKIIYLYISLLAARPKYYFAWTLADAINNAAGFGFRGYDENGAAARWDLISNL RIQQIEMSTSFKMFLDNWNIQTALWLKRVCYERTSFSPTIQTFILSAIWHGVYPGYL TFLTGVLMTLAARAVRNNFRHYFIEPSQLKLFYDVITWIVTQVAISYTVVPFVLLSIK PSLTFYSSWYYCLHILGILVLLLLPVKKTQRRKNTHENIQLSQSKKFDEGENSLGQNS FSTTNVNCNQNEIASRHSSLKQ		

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

Table 6B. Comparison of NOV6a against NOV6b.		
Protein Sequence	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV6b	1..520	474/520 (91%)
	1..487	474/520 (91%)

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

Table 6C. Protein Sequence Properties NOV6a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3406 probability located in mitochondrial intermembrane space; 0.3384 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 44 and 45

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

Table 6D. Geneseq Results for NOV6a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG81345	Human AFP protein sequence SEQ ID NO:208 - Homo sapiens, 423 aa. [WO200129221-A2, 26-APR-2001]	98..520 1..423	419/423 (99%) 421/423 (99%)	0.0
AAB93797	Human protein sequence SEQ ID NO:13560 - Homo sapiens, 432 aa. [EP1074617-A2, 07-FEB-2001]	102..520 14..432	416/419 (99%) 419/419 (99%)	0.0
AAM93974	Human stomach cancer expressed polypeptide SEQ ID NO 17 - Homo sapiens, 432 aa. [WO200109317-A1, 08-FEB-2001]	102..520 14..432	416/419 (99%) 419/419 (99%)	0.0
ABG04835	Novel human diagnostic protein #4826 - Homo sapiens, 371 aa. [WO200175067-A2, 11-OCT-2001]	50..297 23..270	243/248 (97%) 246/248 (98%)	e-143
ABG04835	Novel human diagnostic protein #4826 - Homo sapiens, 371 aa. [WO200175067-A2, 11-OCT-2001]	50..297 23..270	243/248 (97%) 246/248 (98%)	e-143

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

Table 6E. Public BLASTP Results for NOV6a				
Protein Accession Number	Protein/Organism/Length	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAH25429	SIMILAR TO RIKEN CDNA 2810049G06 GENE - Mus musculus (Mouse), 519 aa.	1..520 1..519	451/520 (86%) 479/520 (91%)	0.0
CAC38595	SEQUENCE 207 FROM PATENT WO0129221 - Homo sapiens (Human), 423 aa.	98..520 1..423	419/423 (99%) 421/423 (99%)	0.0
AAH25020	RIKEN CDNA 2810049G06 GENE - Mus musculus (Mouse), 487 aa.	1..520 1..487	422/520 (81%) 449/520 (86%)	0.0
Q9CZ73	2810049G06RIK PROTEIN - Mus musculus (Mouse), 487 aa.	1..520 1..487	421/520 (80%) 448/520 (85%)	0.0
Q96KY4	SIMILAR TO RIKEN CDNA 2810049G06 GENE - Homo sapiens (Human), 350 aa.	171..520 1..350	348/350 (99%) 350/350 (99%)	0.0

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

Table 6F. Domain Analysis of NOV6a			
Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Adeno_Penton_B	204..222	8/20 (40%) 17/20 (85%)	0.54
MBOAT	148..442	108/334 (32%) 225/334 (67%)	4.1e-89

5

Example 7.

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

Table 7A. NOV7 Sequence Analysis		
	SEQ ID NO: 15	537 bp
NOV7a,	ATTAGCAACGGCTCATGATGAACTCAATCAAAGGGGGCTTGACCAGCATCTCAGGTCT	

CG101362-01 DNA Sequence	ACTCTATGTTTTCCAGTGCCCATCAGTGCCAAGGGTATTGATTAGCCTATCCGGGACA AAGAGAGAAGAAAGAGTGAGACACCACCCTAAAAGGGCTGCAGGTGGATACCGCCTC CCTCAAGCTGGAAAAAGATTAGAAAGATGGTGAAAACAGGAAGACCTTCCTCATCCCA CTATCAGGAAGATGAGGAAAGAGATCAGGAGGATCACAGGTGGAGAGGAGAAGAGGAC CATGCTCGATCCTCTCTGGTAATAGGCCTGAGATTCCCTCTCGTACTGGGTGATACAC ATCTGCTCCAGTGTTCCATCCTCCAGGCTTCGGGCGCTTCTTGCAGAGGCCAGGTC ACTCCATGTGGCCACAAAGAGAACCAGCATCCAGCAGCCAAGGTTGCCATAATGACT GCTCTGCCTCGGTCTGTGAGGAGAGGAGAAGCTCGCGGCGCCGCGGCTGTCAGCGACTG GCTCGGAGGACAGGC		
	ORF Start: ATG at 201	ORF Stop: TGA at 480	
	SEQ ID NO: 16	93 aa	MW at 10769.1kD
NOV7a, CG101362-01 Protein Sequence	MVKTGRPSSSHYQEDEERDQEDHRWRGEEDHARSSLVIGLRFPLVLGDTHLLPVFHPP GFGRFLQRPRSLHVATKRTSIQQPRFAIMTALPRS		

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

Table 7B. Protein Sequence Properties NOV7a	
PSort analysis:	0.6400 probability located in microbody (peroxisome); 0.4500 probability located in cytoplasm; 0.2288 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 7C. Geneseq Results for NOV7a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY19678	SEQ ID NO 396 from WO9922243 - Homo sapiens, 133 aa. [WO9922243-A1, 06-MAY-1999]	1..30 63..91	14/30 (46%) 19/30 (62%)	0.94
AAB92467	Human protein sequence SEQ ID NO:10527 - Homo sapiens, 563 aa. [EP1074617-A2, 07-FEB-2001]	2..90 318..398	24/90 (26%) 41/90 (44%)	1.2
AAU16292	Human novel secreted protein, Seq ID 1245 - Homo sapiens, 564 aa. [WO200155322-A2, 02-AUG-2001]	2..90 319..399	24/90 (26%) 41/90 (44%)	1.2
ABB50224	Human transcription factor TRFX-75 - Homo sapiens, 596 aa. [WO200172777-A2, 04-OCT-2001]	2..90 351..431	24/90 (26%) 41/90 (44%)	1.2
AAM33060	Peptide #7097 encoded by probe for measuring placental gene expression - Homo sapiens, 49 aa. [WO200157272-A2, 09-AUG-2001]	5..30 1..25	13/26 (50%) 18/26 (69%)	1.6

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
Q9D3A0	6330414C15RIK PROTEIN - Mus musculus (Mouse), 150 aa.	1..30 51..79	14/30 (46%) 19/30 (62%)	2.2
Q9Y269	Protein HSPC020 - Homo sapiens (Human), and, 121 aa.	1..30 51..79	14/30 (46%) 19/30 (62%)	2.2
Q9UKD0	DNA BINDING PROTEIN P96PIF (GLUCOCORTICOID MODULATORY ELEMENT BINDING PROTEIN 1) - Homo sapiens (Human), 563 aa.	2..90 318..398	24/90 (26%) 41/90 (44%)	2.9
Q9NWH1	HYPOTHETICAL 61.4 KDA PROTEIN - Homo sapiens (Human), 563 aa.	2..90 318..398	24/90 (26%) 41/90 (44%)	2.9
Q9Y692	GLUCOCORTICOID MODULATORY ELEMENT BINDING PROTEIN-1 - Homo sapiens (Human), 573 aa.	2..90 328..408	24/90 (26%) 40/90 (43%)	3.8

Example 8.

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

Table 8A. NOV8 Sequence Analysis		
	SEQ ID NO: 17	3653 bp
NOV8a, CG101458-01 DNA Sequence	CGGGATGCCCCGGCTTGCTGAATTGGATCACGGGGGCAGCCCTGCCCTCACC GCGTCT GATGTTACCTCCTGTGTGTCAGCGGTTATGCCCTGGGCCTAACTGCCTCCCTCACCTATG GCAACCTGGAAGCCCAGCCCTTCCAGGGTCTCTTCGTGTACCCCTGGATGAGTGCAC CACGGTGATCGGCTTTGAGGCAGTCATTGCCGACCGTGTCTGTGACAGTACAGATCAAG GACAAAGCCAAGCTGGAGAGCGGCCACTTCGATGCCTCCCATGTTTCGATCCCCAACAG TCACAGGTAAGGAGACCAGAAGGGCTGCCGCGGGACCTGGGAAGGTGACCTTGGACGA GGATTTGGAGCGGATCCTGTTTCGTGGCCAACCTGGGGACCATTTGCCCCATGGAGAAT GTCACCATCTTCATCAGCACCTCCTCGGAGCTCCCAACGCTGCCCAGCGGGGCTGTGA GGGTCCTTCTGCCTGCTGTCTGTGCCCAACCGTGCCCAAGTTCTGCACCAAGAGCAC TGGCACCTCCAACCAACAGGCCAGGGGAAAGACAGGCACTGCTTCGGTGCCTGGGCC CCGGGCTCCTGGAATAAGTTGTGCCTGGCGACTCTCCTGAACACCGAAGTGTCCAACC CCATGGAGTATGAGTTCAACTTCCAGCTGGAGATCCGTGGGCCATGTCTGCTCGCAGG TGTGGAGAGTCCCACTCATGAGATTCGTGCCGACGCCGCCCATCTGCCCGCTCGGCC AAGAGCATCATCATCACCTTGGCCAACAAGCACACCTTTGACCGGCCTGTGGAGATCC TCATCCACCCCAGCGAGCCCCATATGCCCATGTCTGTATAGAGAAAGGGGACATGAC	

	CCTGGGAGAGTTTGACCAGCACTTGAAGGGAAGAACAGATTTTCATTAAAGGGATGAAG AAGAAGAGCAGAGCAGAGCGGAAGACAGAAATCATTCGAAAACGCCTCCACAAAGACA TTCCCCACCACTCCGTCATCATGCTCAACTTCTGTCCCGACCTCCAGTCAGTCCAGCC GTGCCTGAGAAAGGCCCACGGGGAGTTCATCTTCCTCATTGACAGGAGCAGCAGCATG AGCGGGATCAGCATGCACCGAGTCAAGGATGCCATGTTGGTGGCCCTTAAGAGCCTCA TGCCAGCCTGCCTCTTCAATATCATTTGGGTTTGGATCCACATTTAAGAGCCTTTTTC TTCCAGCCAGACCTACAGTGAGGACAGCTTGGCCATGGCTTGTGATGACATCCAGAGA ATGAAGGCCGACATGGGTGGGACCAACATCCTTTCCCTCTCAAGTGGGTTCATCAGGC AGCCAGTGCACCGAGGCCACCCGCGGCTCCTCTTCGTGATCACAGATGGCGCTGTCAA CAACACAGGGAAGGTGCTGGAGCTGGTGCAGAAATCACGCCTTCTCCACCAGGTGCTAT AGCTTTGGAATTGGACCCAACGTCTGCCACAGACTGGTGAAAGGACTGGCATCTGTGT CCGAGGGCAGTGCTGAGCTCCTGATGGAGGGGGAGCGGCTGCAACCCAAGATGGTCAA ATCCTTGAAGAAGGCCATGGCCCCAGTCTGAGCGATGTGACTGTGGAGTGGATCTTC CCTGAGACCACTGAGGTCTTGGTCTCACCCGTCAGCGCCAGCTCCCTCTTCCCTGGAG AACGGCTGGTGGGGTATGGCATTGTATGTGATGCTTCTTTGCACATCTCCAATCCCAG ATCTGACAAGAGGGCGCCGGTACAGCATGCTGCACTCTCAGGAGTCTGGCAGCTCTGTC TTCTACCACTCTCAGGATGACGGACCCGGGCTGGAAGGTGGAGACTGTGCCAAGAACT CGGGGGCACCCCTTCATCCTAGGGCAGGCCAAAAATGCCCGGCTAGCCAGCGGAGACTC TACCACCAAGCACGGTCTGAACCTCTCTCAGCGACGGAGGGCATAACAGCACCAG ATCACCAATCACAAGCCCCCTCCCAAGAGCCACCATGGCAAGTGACCCCATGCCAGCTG CCAAGAGATACCCACTGCGGAAAGCCAGGCTGCAGGACCTCACCAACCAGACCAGCCT GGATGTCCAGCGGTGGCAGATTGATTTGCAGGTATTGCTGAACAGTGGTCAGGACCTG AACCAGGGCCCCAACTCCGTGGCCAGGGGGCCGAAGGCCCTCTCTGCTGCCCAAG GCTGCCAGCCCTTCTGCCCTGGGGCCAGGAGACCCAGGCCTGGAGCCCTGTGAGAGA GCGGACTTCTGACAGCCGAAGCCCTGGAGATCTGCCCGCAGAGCCGTCCCACCATCCC TCTGCCTTCGAGACAGAGACGTCTCGGACTGGGACCCCCCAGCCGAGTCCCAGGAGC GAGCCAGTCCCAGCAGGCCCGCCACCCCGGCCCGGTGCTGGGCAAGGCCCTGGTCAA AGGCCTGCACGACAGCCAACGCCTGCAGTGGGAGGTGAGCTTCGAGCTGGGGACCCCT GGACCGGAGCGGGGCGGCGCGCAGGATGCCGACCTATGGAGCGAGACCTTCCACCACC TGGCGGCCCCGCGCCATCATCCGCGACTTCGAGCAGCTGGCGGAGCGCGAGGGCGAGAT CGAGCAGGGTTCCAACCGCCGCTACCAAGTGAGCGCCTTGACACACCAGCAAGGCCTGC AACATCATTAGCAAATACACAGCCTTCGTGCCTGTGGACGTGAGCAAGAGCCGGTACC TGCCACCGTGGTGGAGTACCCCAACTCTGGTCTGATGCTTGGCTCTCGGGCCCTGGC CCAACAGTGGAGGGGCACCTCTTCTGGCTTTGGAAGGCCGAGACGATGCTTGGAGAA GATTTCGGCACCAGGAAATGGTAAATTTTCAGGTCTTAGACATGGAGGCAAGTCCCACTG CTCTCTTCAGCGAGGCCAGGTCCCCCGGCCGCGAGAAGCACGGTGCTTCTGAAGGTCC CCAGCGCAGCCTGGCTACAAATACTCTTTCTTCCATGAAGGCCCTCAGAGAATCTCTTT GGATCCAGGCTAAATCTCAACAAGTCCAGGCTACTGACGCGAGCAGCCAAGGGCTTCC TGAGCAAGCCACTGATCAAAGCTGTGGAGTCGACCTCCGGGAACCAGAGCTTCGACTA CATACCTCTGGTGTCTCTGCAGCTGGCCTCCGGAGCCTTCCTGCTCAACGAAGCCTTC TGTGAGGCCACGCACATCCCCATGGAGAAGCTCAAGTGGACGTCCCCCTTCACCTGCC ATCGAGTGTCCCTCACCACCCGCCCCGTCTGAGTCCAAGACCCCGAGTCCCCAGCTGTG CACCAGCTCCCCGCTTAGGCACCCGCTCCTGTGACAGCTTCTCCCTGGAGCCTCTGGCC AAGGGCAAGCTGGGCCTGGAGCCGAGGGCAGTGGTGGAGCACACTGGGAAGCTGTGGG CCACGGTGGTGGGGCTGGCATGGCTGGAGCACAGTTCGGCCTCCTACTTCACTGAGTG GGAGTTGGTGGCTGCCAAGGCCAACTCATGGCTGGAGCAGCAGGAAGTACCCGAGGGC CGCACGCAGGGCACACTCAAGGCCGCTGCCCGCCAGCTGTTTGTGCTTCTGCGGCACT GGGATGAGAATCTCGAGTTCAATATGCTCTGCTATAACCCGAATTATGTGTAGTTGA		
	ORF Start: ATG at 5	ORF Stop: TAG at 3647	
	SEQ ID NO: 18	1214 aa	MW at 133118.0kD
NOV8a, CG101458-01 Protein Sequence	MPGLLNWITGAALPLTASDVTSCVSGYALGLTASLTYNLEAQPFQGLFVYPLDECTT VIGFEAVIADRVVTVQIKDKAKLESGHFDASHVRSPTVTGKETRRAAAGPGKVTLD LERILFVANLGTIAPMENVTFISTSSSELPTLPSGAVRVLLPAVCAPTVPQFCTKSTG TSNQQAQGKDRHCFGAWAPGSWNKLCLATLLNTEVSNPMEYEFNFQLEIRGPCLLAGV ESPTHEIRADAAPSARSAKSIIITLANKHTFDRPVEILJHPSEPHMPHVLIIEKGDMTL GEFDQHLKGRTDIFIKMKKKSRAERKTEIRKRLHKDIPHHSVIMLNFCPDLQSVQPC LRKAHGEFIFLIDRSSMSGISMHRVKDAMLVALKSLMPACLFNIIGFGSTFKSLFPS SQTYSEDSLAMACDDIQRMKADMGGTNILSPLKWVIRQPVHRGHPRLLFVITDGAVNN TGKVLELVRNHAFSTRCYSFGIGPNVCHRLVKGLASVSEGSSELLMEGERLQPKMVKS		

	LKKAMAPVLSDTVVEWIFPETTEVLVSPVSASSLFPGERLVGYGIVCDASLHISNPRS DKRRRYSM LHSQESGSSVFYHSQDDGPGLEGDCAKNSGAPFILGQAKNARLASGDST TKHGLNLSQRRRAYSTNQITN HKPLPRATMASDPM PAAKRYPLRKARLQDLTNQTS LD VQRWQIDLQVLLNSGQDLNQGP KLRGPGARRPSLLPQGCQPFLPWGQETQAWSPVRER TSDSRSPGDLPAEP SHHPSAFETETSSDWDPPAESQERASPSRPATPAPVLGKALVKG LHDSQRLQWEVSFELGTPGPERGGAQDADLWSETFHHLAARAIIRD FEQLAEREGEIE QGSNRRYQVSALHTSKACNIISKYTAFVPVDVSKSRYLPTVVEYPNSGRMLGSRALAQ QWRGTSSGFGRPQTMLGEDSAPGNGKFQVLDMEASPTALFSEARS PGREKHGASEGPQ RSLATNTLSSMKASENLFGSRLNLNKSRL LTRAAKGFLSKPLIKAVESTSGNQSF DYI PLVSLQLASGAFL LNEAFCEATHIPMEKLKWTSPFTCHRVSLTTRPSESKTPSPQLCT SSPPRHPSCD SFSLEPLAKGKLGLEPRAVVEHTGKLWATVVGLAWLEHSSASYFTEWE LVAAKANSWLEQQEVPEGRTQGTLKAAARQLFVLLRHWDENLEFNMLCYNPNYV
--	---

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

Table 8B. Protein Sequence Properties NOV8a	
PSort analysis:	0.8700 probability located in nucleus; 0.8500 probability located in endoplasmic reticulum (membrane); 0.7900 probability located in plasma membrane; 0.3325 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 19 and 20

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

Table 8C. Geneseq Results for NOV8a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB82047	Human mast cell surface antigen - Homo sapiens, 786 aa. [JP2001025388-A, 30-JAN-2001]	13..565 15..500	168/565 (29%) 269/565 (46%)	2e-59
AAV82530	Human neurotransmitter associated protein sequence SEQ ID NO:6 - Homo sapiens, 210 aa. [WO200012685-A2, 09-MAR-2000]	1034..1211 16..207	82/194 (42%) 105/194 (53%)	1e-32
AAU33242	Novel human secreted protein #3733 - Homo sapiens, 1730 aa. [WO200179449-A2, 25-OCT-2001]	36..568 650..1096	120/537 (22%) 214/537 (39%)	4e-23
AAB51022	Human minor vault protein p193 - Homo sapiens, 1724 aa. [US6156879-A, 05-DEC-2000]	36..568 644..1090	120/537 (22%) 214/537 (39%)	6e-23
AAV54373	cDNA sequence encoding the human minor vault protein p193 - Homo sapiens, 1724 aa. [WO9962547-A1, 09-DEC-1999]	36..568 644..1090	120/537 (22%) 214/537 (39%)	6e-23

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

Table 8D. Public BLASTP Results for NOV8a				
Protein Accession Number	Protein/Organism/Length	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9CUE8	4931403E03RIK PROTEIN - Mus musculus (Mouse), 1209 aa (fragment).	1..1208 1..1209	883/1218 (72%) 1012/1218 (82%)	0.0
Q96M71	CDNA FLJ32784 FIS, CLONE TESTI2002245 - Homo sapiens (Human), 424 aa.	588..953 1..367	362/369 (98%) 362/369 (98%)	0.0
Q9BVH8	HYPOTHETICAL 106.2 KDA PROTEIN - Homo sapiens (Human), 1001 aa (fragment).	274..1211 32..998	311/1047 (29%) 467/1047 (43%)	e-106
O75668	DJ745E8.1 (BREAST CANCER SUPPRESSOR CANDIDATE 1 (BCSC-1) LIKE) - Homo sapiens (Human), 148 aa (fragment).	417..564 1..148	148/148 (100%) 148/148 (100%)	4e-80
Q9CTV9	5830475I06RIK PROTEIN - Mus musculus (Mouse), 565 aa (fragment).	13..565 15..500	165/567 (29%) 259/567 (45%)	5e-57

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

Table 8E. Domain Analysis of NOV8a			
Pfam Domain	NOV8a Match Region	Identities/ Similarities for the Matched Region	Expect Value
vwa	355..523	37/203 (18%) 107/203 (53%)	0.021

5 Example 9.

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

Table 9A. NOV9 Sequence Analysis		
	SEQ ID NO: 19	868 bp
NOV9a,	CGTTTTCTTCTACAATGTCTGAAGAAGTGACCTACGCGACACTCACATTTTCAGGATTC TGCTGGAGCAAGGAATAACCGAGATGGAAATAACCTAAGAAAAAGAGGTCATCCAGCT	

CG101475-01 DNA Sequence	CCATCTCCCATTGTTGGCGTCATGCTGCTCTGGGTCTGGTAACTCTTTGCCTGATGTTGCTGATTGGGCTGGTGACATTGGGGATGATGTGTTTGCAGATATCTAATGACATTAACTCAGATTCAGAGAAATTGAGTCAACTTCAGAAAACCATCCAACAGCAGCAGGATAACTTATCCCAGCAACTGGGCAACTCCAACAACCTTGTCATGGAGGAGGAATTTCTCAAGTCACAGATCTCCAGTGTACTGAAGAGGCAGGAACAAATGGCCATCAAACCTGTGCCAAGAGCTAATCATTCAATTTTTTCAGACCACAGATGTAATCCATGTCCTAAGATGTGGCAATGGTACCAAATAGTTGCTACTATTTTACAACAAATGAGGAGAAAACCTGGGCTAACAGTAGAAAGGACTGCATAGACAAGAACTCCACCCTAGTGAAGATAGACAGTTTGAAGAAAAGGATTTTTCTTATGTCACAGCCATTACTCATGTTTTTCGTTCTTTTGGCTGGGATTATCATGGGACTCCTCTGGCAGAAGTTGGTTCTGGGAAGATGGCTCTGTTCCCTCTCCATCCTTGA GTACTAAAGAACTTGACCAGATCAATGGATCCAAAGGATGTGCTTATTTTCAAAAAGGAAATATTTTATATTTCTCGCTGTAGTGCTGAAATTTTTTGGATTTGCGAGAAGACAGCTGCCCCAGTGAAGACTGAGGATTTGGATTAGTATGCTTCTTCCAAATTCTCCAAGAA		
	ORF Start: ATG at 15	ORF Stop: TAG at 840	
	SEQ ID NO: 20	275 aa	MW at 31470.4kD
NOV9a, CG101475-01 Protein Sequence	MSEEVTYATLTFQDSAGARNNRDGNLRLKRGH PAPSPIWRHAALGLVTLCLMLLIGLVTLGMMCLQISNDINS DSEKLSQLQKTIQQQQDNLSQQLGNSNNLSMEEEFLLKSQISSVLKRQEQMAIKLCQELIIHFS DHRCNPCPKMWQWYQNSCYFTTNEEKTWANSRKDCIDKNSTLVKIDSLEEKDFLMSQPLL MFSFFWLGLSWDSSGRSWFWEDGSVPSPSLSTKELDQINGSKGCAYFQKGNIIYISRCSAEI FWICEKTAAPVKTEDLD		
	SEQ ID NO: 21	819 bp	
NOV9b, CG101475-02 DNA Sequence	ACACTCACATTTTCAGGATTCTGCTGGAGCAAGGAATAACCGAGATGGAAATAACCTAAGAAAAAGAGGGCATCCAGCTCCATCTCCCATTTGGCGTCATGCTGCTCTGGGTCTGGTAACTCTTTGCCTGATGTTGCTGATTGGGCTGGTGACGTTGGGGATGATGTTTTTGCAGATATCTAATGACATTAACTCAGATTCAGAGAAATTGAGTCAACTTCAGAAAACCATCCAACAGCAGCAGGATAACTTATCCCAGCAACTGGGCAACTCCAACAACCTTGTCATGGA GGAGGAATTTCTCAAGTCACAGATCTCCAGTGTACTGAAGAGGCAGGAACAAATGGCCATCAAACCTGTGCCAAGAGCTAATCATTCACTTTCAGACCACAGATGTAATCCATGTCCTAAGATGTGGCAATGGTACC AAAATAGTTGCTACTATTTTACAACAAATGAGGAGAAACCTGGGCTAACAGTAGAAAGGACTGCATAGACAAGAACTCCACCCTAGTGAAGATAGACAGTTTGAAGAAAAGGATTTTCTTATGTCACAGCCATTACTCATGTTTTTCGTTCTTTTGGCTGGGATTATCATGGGACTCCTCTGGCAGAAGTTGGTTCTGGGAAGATGGCTCTGTTCCCTCTCCATCCTTATTTAGTACTAAAGAACTTGACCAGATCAATGGATCCAAAGGATGTGCTTATTTTCAAAAAGGAAATATTTATATTTCTCGCTGTAGTGCTGAAATTTTTTGGATTTGCGAGAAGACAGCTGCCCCAGTGAAGACTGAGGATTTGGATTAGAAGGGCGATTCC		
	ORF Start: at 1	ORF Stop: TAG at 805	
	SEQ ID NO: 22	268 aa	MW at 30704.5kD
NOV9b, CG101475-02 Protein Sequence	TLTFQDSAGARNNRDGNLRLKRGH PAPSPIWRHAALGLVTLCLMLLIGLVTLGMMFLQISNDINS DSEKLSQLQKTIQQQQDNLSQQLGNSNNLSMEEEFLLKSQISSVLKRQEQMAIKLCQELIIHTSDHRCNPCPKMWQWYQNSCYFTTNEEKTWANSRKDCIDKNSTLVKIDSLEEKDFLMSQPLL MFSFFWLGLSWDSSGRSWFWEDGSVPSPSLFSTKELDQINGSKGCAYFQKGNIIYISRCSAEI FWICEKTAAPVKTEDLD		

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

Table 9B. Comparison of NOV9a against NOV9b.		
Protein Sequence	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV9b	9..275	241/268 (89%)
	1..268	241/268 (89%)

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

Table 9C. Protein Sequence Properties NOV9a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 62 and 63

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

5

Table 9D. Geneseq Results for NOV9a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU29320	Human PRO polypeptide sequence #297 - Homo sapiens, 232 aa. [WO200168848-A2, 20-SEP-2001]	1..227 1..227	224/227 (98%) 225/227 (98%)	e-131
AAM79324	Human protein SEQ ID NO 2970 - Homo sapiens, 289 aa. [WO200157190-A2, 09-AUG-2001]	1..270 25..280	91/270 (33%) 147/270 (53%)	3e-37
ABB11776	Human macrophage Ag homologue, SEQ ID NO:2146 - Homo sapiens, 289 aa. [WO200157188-A2, 09-AUG-2001]	1..270 25..280	91/270 (33%) 147/270 (53%)	3e-37
AAM78340	Human protein SEQ ID NO 1002 - Homo sapiens, 265 aa. [WO200157190-A2, 09-AUG-2001]	1..270 1..256	88/270 (32%) 147/270 (53%)	1e-35
AAY02283	Secreted protein clone br342_11 polypeptide sequence - Homo sapiens, 265 aa. [WO9918127-A1, 15-APR-1999]	1..270 1..256	88/270 (32%) 147/270 (53%)	1e-35

In a BLAST search of public sequence databases, the NOV9a protein was found to have homology to the proteins shown in the BLASTP data in Table 9E.

Table 9E. Public BLASTP Results for NOV9a				
Protein Accession Number	Protein/Organism/Length	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9D403	4933425B16RIK PROTEIN - Mus musculus (Mouse), 275 aa.	1..275 1..275	197/276 (71%) 227/276 (81%)	e-113
AAL95693	C-TYPE LECTIN PROTEIN CLL-1 - Homo sapiens (Human), 265 aa.	1..270 1..256	88/270 (32%) 147/270 (53%)	2e-35
Q9NZH3	C-TYPE LECTIN-LIKE RECEPTOR-1 - Homo sapiens (Human), 280 aa.	28..274 36..269	83/249 (33%) 131/249 (52%)	5e-33
Q9XTA8	LECTIN-LIKE OXIDIZED LDL RECEPTOR - Oryctolagus cuniculus (Rabbit), 278 aa.	36..272 36..278	79/247 (31%) 124/247 (49%)	5e-27
P78380	LECTIN-LIKE OXIDIZED LDL RECEPTOR - Homo sapiens (Human), 273 aa.	36..266 32..268	79/245 (32%) 124/245 (50%)	3e-24

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

Table 9F. Domain Analysis of NOV9a			
Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
lectin_c	161..264	29/125 (23%) 61/125 (49%)	7.4e-06

5

Example 10.

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

Table 10A. NOV10 Sequence Analysis		
	SEQ ID NO: 23	516 bp
NOV10a, CG101772-01 DNA Sequence	CACTGCGCATGCTATTTGGGCGCCACCTCAGTGCACATGTTCACTGGGCGTCTTCTA CTCTACCCCTTCGCCCTCGTGGGGGTGTGAGGGTCGCGTTCCTGCTGTCTGGACTTTT TCTGTCCCACTGAGACGCAATGTATCGATAACAAACTTTTTATCTGCACACACACAC ACACACACACACACCCTGGTTCCAGGAGCCCGGTGATGAGGAGCCTCAGCAAGAGGA	

	ACCACCAACTGAAAGTCGGGATCCTGCACCTGGTCAGGAGAGAGAAGAAGATCAGGGT GCAGCTGAGACTCAATGCCTGACCTGGAAGCTGATCTCCAGGAGCTGTCTCAGTCAAA GACTGGGGGTGAATGTGGAAATGAAGATTCTGCCAAAATCAGAACAATTTAAAATGCC AGAAGGAGGTATGCTATCCATTATTATGTGCTTTCTGTTTTCCACAATATTATACTTT TGATAATAAAAGAGAACATTACTATCCCTTTAAAATCAGAGTTCAAATGCAG		
	ORF Start: ATG at 9	ORF Stop: TGA at 465	
	SEQ ID NO: 24	152 aa	MW at 17265.6kD
NOV10a, CG101772-01 Protein Sequence	MLFGRPPQCTCSLGVFYSTPSPSWGCEGRVPAVWTFVPLRRNVSITKLFICTHHTH THPWFAQEPGDEEPQQEPTESRDPAPGQEREEDQGAAETQCLTWKLISRSCLSQRLG VNVEMKILPKSEQFKMPEGMLSIIIMCFLFSTILYF		

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

Table 10B. Protein Sequence Properties NOV10a	
PSort analysis:	0.9190 probability located in plasma membrane; 0.2000 probability located in lysosome (membrane); 0.1021 probability located in microbody (peroxisome); 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 publication, yielded several 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
AAM39588	Human polypeptide SEQ ID NO 2733 - Homo sapiens, 111 aa. [WO200153312-A1, 26-JUL-2001]	65..136 28..105	51/78 (65%) 56/78 (71%)	4e-20
AAM41374	Human polypeptide SEQ ID NO 6305 - Homo sapiens, 106 aa. [WO200153312-A1, 26-JUL-2001]	65..135 29..105	52/77 (67%) 57/77 (73%)	2e-19
ABG05297	Novel human diagnostic protein #5288 - Homo sapiens, 112 aa. [WO200175067-A2, 11-OCT-2001]	65..136 29..106	48/78 (61%) 56/78 (71%)	8e-19
ABG05297	Novel human diagnostic protein #5288 - Homo sapiens, 112 aa. [WO200175067-A2, 11-OCT-2001]	65..136 29..106	48/78 (61%) 56/78 (71%)	8e-19
ABG27048	Novel human diagnostic protein #27039 - Homo sapiens, 249 aa. [WO200175067-A2, 11-OCT-2001]	64..135 70..147	45/78 (57%) 53/78 (67%)	5e-15

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
Q8WTP9	XAGE-3 PROTEIN - Homo sapiens (Human), 111 aa.	65..136 28..105	51/78 (65%) 56/78 (71%)	1e-19
Q8WYS9	HYPOTHETICAL 12.3-KDA PROTEIN - Homo sapiens (Human), 111 aa.	65..136 28..105	51/78 (65%) 56/78 (71%)	1e-19
Q9HD64	G antigen family D 2 protein (XAGE-1) - Homo sapiens (Human), 146 aa.	1..136 1..140	59/149 (39%) 76/149 (50%)	3e-18
Q8WWM1	XAGE-5 PROTEIN - Homo sapiens (Human), 108 aa.	65..136 25..102	45/78 (57%) 53/78 (67%)	9e-15
Q96GT9	SIMILAR TO G ANTIGEN 8 (XAGE-2 PROTEIN) - Homo sapiens (Human), 111 aa.	65..136 28..105	39/78 (50%) 53/78 (67%)	7e-13

Example 11.

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

Table 11A. NOV11 Sequence Analysis			
	SEQ ID NO: 25	709 bp	
NOV11a, CG102532-01 DNA Sequence	CGGCCGGT TTTGGTAGGCCCGGGCCGCCAGGCCTCCGCCTGAGCCCGCACCCGCC ATGGACA ACTACGCAGATCTTTCGGATACCGAGCTGACCACCTTGCTGCGCCGGTACA ACATCCCGCACGGGCTGTAGTAGGATCAACTCGTAGGCTTTACGAGAAGAAGATCTT CGAGTACGAGACCCAGAGGCGGCGGCTCTCGCCCCCAGCTCGTCCGCCGCCTCCTCT TATAGCTTCTCTGACTTGAATTCGACTAGAGGGGATGCAGATATGTATGATCTTCCCA AGAAAGAGGACGCTTTACTCTACCAGAGCAAGGGCTACAATGACGATCTTTTGTCTTC TTCTGAAGAGGAGTGCAAGGATAGGGAACGCCCCATGTACGGCCGGGACAGTGCCTAC CAGAGCATCACGCACTACCGCCCTGTTTCAGCCTCCAGGAGCTCCCTGGACCTGTCCT ATTATCCTACTTCCTCCTCCACCTCTTTTATGTCCTCCTCATCATCTTCCTCTTCATG GCTCACCCGCCGTGCCATCCGGCCTGAAAACCGTGCTCCTGGGGCTGGGCTGGGCCAG GATCGCCAGGTCCCGCTCTGGGGCCAGCTGCTGCTTTTCTGCTGCTTTTGTGATCGTCC TCTTCTTCATTTACCACTTCATGCAGGCTGAAGAAGGCAACCCCTTCTGACTGCAGCC AAGCTAATTCCGG		
	ORF Start: ATG at 59	ORF Stop: TGA at 686	
	SEQ ID NO: 26	209 aa	MW at 23844.1kD
NOV11a, CG102532-01	MDNYADLSDELTTLLRRYNIPHPVVGSTRRLYEKKIFEYETQRRRLSPSSSAASS YSFSDLNSTRGDADMYDLPKKEDALLYQSKGYNDLLSSSEEECKDRERPMYGRDSAY QSITHYRPVSASRSLDLSYYPTSSSTSFMSSSSSSSSWLTRRAIRPENRAPGAGLGQ		

Protein Sequence	DRQVPLWGQLLLFLVFVIVLFFIYHFMQAEEGNPF
------------------	-------------------------------------

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

Table 11B. Protein Sequence Properties NOV11a	
PSort analysis:	0.8500 probability located in endoplasmic reticulum (membrane); 0.6000 probability located in nucleus; 0.4400 probability located in plasma membrane; 0.2323 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 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
AA Y41294	Human emerin sequence (EMD_HU) - Homo sapiens, 254 aa. [WO9954468-A1, 28-OCT- 1999]	1..209 1..254	209/254 (82%) 209/254 (82%)	e-112
AAG02346	Human secreted protein, SEQ ID NO: 6427 - Homo sapiens, 51 aa. [EP1033401-A2, 06-SEP-2000]	1..51 1..51	51/51 (100%) 51/51 (100%)	2e-23
AA Y41297	Human thymopoietin gamma sequence - Homo sapiens, 345 aa. [WO9954468-A1, 28-OCT-1999]	6..209 114..333	60/231 (25%) 107/231 (45%)	7e-10
AAR93188	Thymopoietin-gamma - Homo sapiens, 345 aa. [WO9609526-A1, 28-MAR-1996]	6..209 114..333	60/231 (25%) 107/231 (45%)	7e-10
AAR76499	Human thymopoietin-gamma - Homo sapiens, 345 aa. [WO9517205-A1, 29-JUN-1995]	6..209 114..333	60/231 (25%) 107/231 (45%)	7e-10

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
P50402	Emerin - Homo sapiens (Human), 254 aa.	1..209 1..254	209/254 (82%) 209/254 (82%)	e-111
Q63190	Emerin - Rattus norvegicus (Rat), 260 aa.	1..209 1..256	162/256 (63%) 182/256 (70%)	1e-81
O08579	Emerin - Mus musculus (Mouse), 259 aa.	1..209 1..255	162/255 (63%) 182/255 (70%)	1e-81
Q61032	THYMOPOIETIN GAMMA - Mus musculus (Mouse), 342 aa.	6..209 112..331	66/231 (28%) 106/231 (45%)	2e-11
AAC25390	THYMOPOIETIN GAMMA - Homo sapiens (Human), 345 aa.	6..209 114..333	60/231 (25%) 107/231 (45%)	2e-09

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
LEM	1..44	22/47 (47%) 43/47 (91%)	4.4e-24

5 Example 12.

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

Table 12A. NOV12 Sequence Analysis		
	SEQ ID NO: 27	2812 bp
NOV12a, CG102575-01 DNA Sequence	CATTGAGTCGGCTTTTCTACTGCTTCGGCTAGGGTACCTTGTGACCATGTCTTCCAAG AAGAATAGAAAGCGGTTGAACCAAAGCGCGGAAATGGTTTCGTCCCTTGCCCTCTGCTG CTTCCTCTTGTGCGGAGGCACGGGCTCCTTCTGCTGGATCAGACTTCGCGGCAACCTC CGGGACTCTGACGGTGACCAACTTATTAGAAAAGGGTAAAATTCTTAAACATTCCAG AATTCCCTTATTCATCTTGGACTCAACACTATGAAGTCTGCAAATATATGTATAGGTC GACCAGTGTTGCTTACTAGTTTGAACGGAAAGCAAGAGGTATATACAGCCTGGCCTAT GGCAGGATTTCTTGGAGGCAAGGTCGGCCTGAGTGAAATGGCACAGAAAAATGTGGGT GTGAGGCCTGGTGATGCCATCCAGGTCCAGCCTCTTGTGGGTGCTGTGCTACAGGCTG AGGAAATGGATGTGGCACTGAGTGACAAAGATATGGAAATTAATGAAGAAGAACTGAC	

	TGGTTGTATCCTGAGAAAAGTAGATGGCAAGATTGTTTTACCAGGCAACTTTCTGTAT TGTACATTCTATGGACGACCGTACAAGCTGCAAGTATTGCGAGTGAAAGGGGCAGATG GCATGATATTGGGAGGGCCTCAGAGTGACTCTGACACTGATGCCCAAAGAATGGCCTT TGAACAGTCCAGCATGGAAACCAGTAGCCTGGAGTTATCCTTACAGCTAAGCCAGTTA GATCTGGAGGATACCCAGATCCCAACATCAAGAAGTACTCCTTATAAACCAATTGATG ACAGAATTACAAATAAAGCCAGTGATGTTTTGCTGGATGTTACACAGAGCCCTGGAGA TGGCAGTGGACTTATGCTAGAGGAAGTCACAGGTCTTAAATGTAATTTTGAATCTGCC AGAGAAGGAAATGAGCAACTTACTGAAGAAGAGAGACTGCTAAAGTTCAGCATAGGAG CAAAGTGCAATACTGATACTTTTTATTATTTCTTCAACAACAAGAGTCAATTTTAC AGAGATTGATAAAAATTCAAAGAGCAAGACAGTGATGTTAAAAGTAACTATGACCAT GATAGAGGATTAAGTAGCCAGCTGAAAGCAATTAGAGAAATAATTGAATTGCCCTCA AAATTCCTGCCCCTAGAGGATTGTTACTTTATGGTCCTCCATGTACTGGAAAAACAAT GATCGCCAGGGCTGTTGCTAATGAATTTGGAGCCTATGTTTCTGTAATTAATGGTCCT GAAATTATAAGCAAGTTCTATGGTGAGACTGAAGCAAAGTTACGTCAGATATTTGCTG AAGCCACTCTAAGACACCCATCAATTATTTTATTGATGAGCTGGATGCACCTTTGTCC GAAAAGAGAGGGGGCCAGAATGAAGTGGAAGGAGAGTTGTGGCTTCACTCTTAACA CTGATGGATGGCATTGGTTCAGAAGTAAGTGAAGGACAAGTGTTGGTCTTGGGGCCA CAAATCGCCCTCATGCCTTGGATGCTGCTCTCCGAAGACCTGGGCGATTTGATAAAGA GATTGAGATTGGAGTTCCCAATGCTCAGGACCGGCTAGATATTCTCCAGAACTGCTT CGAAGGGTACCCCATTTGCTCACTGAGGCTGAGCTGCTGCAGCTGGCAAATAGTGCTC ATGGATACGTTGGAGCAGACTTGAAAGTCTTGTGTAATGAAGCAGGTCTCTGTGCCTT GCGGAGAATCCTGAAAAACAGCCTAACCTCCCTGATGTCAAGGTGGCTGGACTGGTG AAGATTACTCTGAAGGATTCTTGCAGGCAATGAATGATATCAGACCCAGTGCCATGA GGGAAATAGCAATTGATGTCCCAATGTAAGTTATGATGATGTTGGTGGAGTTAGAAA GCAAATGGCCCAAATCAGAGAGCTTGTGAGCTTCCACTACGCCATCCTCAACTTTTC AAATCTATTGGTATTCCTGCCCCTAGAGGATTGTTACTTTATGGTCCTCCATGTACTG GAAAAACAATGATCGCCAGGGCTGTTGCTAATGAATTTGGAGCCTATGTTTCTGTAAT TAATGGTCCTGAAATTATAAGCAAGTATGTTGGTGAGAGTGAACGTGCTGTGCGACAA GTTTTTCAACGAGCCAAGAACTCAGCACCATCAATTATTTTTATTGATGAGCTGGATG CACTTTGTCCGAAAAGAGAGGGGGCCAGAATGAAGTGGAAGGAGAGTTGTGGCTTC ACTCTTAACACTGATGGATGGCATTGGTTCAGTAAGTATAGTGTGGTCTTGGGGCC ACAAATCGCCCTCATGCCTTGGATGCTGCTCTCCGAAGACCTGGGCGATTTGATAAAG AGATTGAGATTGGAGTTCCCAATGCTCAGGACCGGCTAGATATTCTCCAGAACTGCT TCGAAGGGTACCCCATTTGCTCACTGAGGCTGAGCTGCTGCAGCTGGCAAATAGTGCT CATGGATACGTTGGAGCAGACTTGAAAGTCTTGTGTAATGAAGCAGGTGAGTGTGGTT TGCTATGGGACATTCAAGCCAATCTCATGAAAAGACATTTCACTCAGGCCTTGAG CACTGTGACACCTAGAATTCCTGAGTCATTGAGACGTTTTTATGAAGATTATCAAGAG AAGAGTGGGCTGCATACACTCTGAGAAAATATATATATTCAAGATGCTGAAAATCCTT TCCAGAGAAAATTGTTTCTTTTAAATTTTGTAGAGTGTAAAAAAATTTTACTAG GCAAAATGTTTGAAGTATGTTTCAGTAGA		
	ORF Start: ATG at 47	ORF Stop: TGA at 2690	
	SEQ ID NO: 28	881 aa	MW at 96419.5kD
NOV12a, CG102575-01 Protein Sequence	MSSKKNRKRLNQSAENGSSLPASAASSCAEARAPSAGSDFAATSGTLTVTNLLEK GKIP KTFQNSLIHLGLNTMKSANICIGRPVLLTSLNGKQEVYTAWPMAGFPGGKVGLSEMAQ KNVGVRPGDAIQVQPLVGAVLQAEEMDVALSDKDMEINEEELTGCILRKLDGKIVLPG NFLYCTFYGRPYKLQVLRVKGADGMILGGPQSDSDTDAQRMAFEQSSMETSSLELSLQ LSQLDLEDTPQIPSTRSTPYKPIDDRITNKASDVLLDVTQSPGDGSGLMLEEVTGLKCN FESAREGNEQLTEEERLLKFSIGAKCNTDTFYFISSTTRVNFTEIDKNSKEQSDSVKS NYDHDRGLSSQLKAIREIIELPLKIPAPRGLLLYGPPCTGKTMIAARAVANEF GAYVSV INGPEIISKFYGETEAKLRQIFAEATLRHPSIIFIDELDALCPKREGAQNEVEKRVVA SLLTLMDGIGSEVSEGQVLVLGATNRPHALDAALRRPGRFDKEIEIGVPNAQDRLDIL QKLLRRVPHLLTEAELLQLANSAHGYVGADLVLCNEAGLCALRRILKKQPNLPDVKV AGLVKITLKDFLQAMNDIRPSAMREIAIDVPNVSYDDVGGVRKQMAQIRELVELPLRH PQLFKSIGIPAPRGLLLYGPPCTGKTMIAARAVANEF GAYVSVINGPEIISKYVGESER AVRQVFQRAKNSAPSIIFIDELDALCPKREGAQNEVEKRVVASLLTLMDGIGSVSIVL VLGATNRPHALDAALRRPGRFDKEIEIGVPNAQDRLDILQKLLRRVPHLLTEAELLQL ANSAHGYVGADLVLCNEAGECGLLWDIQANLIMKRHFTQALSTVTPRIPESLRRFYE DYQEKSGLHTL		

	SEQ ID NO: 29	2789 bp
NOV12b, CG102575-02 DNA Sequence	CAGAGTTCGCCCTTCATTGAGTCGGCTTTTCTACTGCTTCGGCTAGGGTACCTTGTGA CCATGTCTTCCAAGAAGAATAGAAAGCGGTTGAACCAAAGCGCGGAAAATGGTTCGTC CTTGCCCTCTGCTGCTTCTCTTGTGTGGAGGCACGGGCTCCTTCTGCTGGATCAGAC TTCGCGGCAACCTCCGGGACTCTGACGGTGACCAACTTATTAGAAAAGGTAGATGACA AAATTCCTAAAACATTCCAGAATTCCTTATTTCATCTTGGACTCAACACTATGAAGTC TGCAAATATATGTATAGGTGACCAAGTGTGCTTACTAGTTTGAACGGAAAGCAAGAG GTGTATACAGCCTGGCCTATGGCAGGATTTCTTGGAGGCAAGGTTCGGCCTGAGTGAAA TGGCACAGAAAAATGTGGGTGTGAGGCCTGGTGATGCCATCCAGGTCCAGCCTCTTGT GGGTGCTGTGCTACAGGCTGAGGAAATGGATGTGGCACTGAGTGACAAAGATATGGAA ATTAATGAAGAAGAAGTACTGGTTGTATCCTGAGAAAAGTACTAGATGGCAAGATTGTTT TACCAGGCAACTTTCTGTATTGTACATTCTATGGACGACCGTACAAGCTGCAAGTATT GCGAGTGAAAGGGGCAGATGGCATGATATTGGGAGGGCCTCAGAGTGACTCTGACACT GATGCCCCAAGAATGGCCTTTGAACAGTCCAGCATGGAAACCAGTAGCCTGGAGTTAT CCTTACAGCTAAGCCAGTTAGATCTGGAGGATACCAGATCCCAACATCAAGAAGTAC TCCTTATAAACCAATTGATGACAGAATTACAAATAAAGCCAGTGATGTTTTGCTGGAT GTTACACAGAGCCCTGGAGATGGCAGTGGACTTATGCTAGAGGAAGTCACAGGTCTTA AATGTAATTTTGAATCTGCCAGAGAAGGAAATGAGCAACTTACTGAAGAAGAGAGACT GCTAAAGTTCAGCATAGGAGCAAAGTGCAATACTGATACTTTTTTATTTTATTCTTCA ACAACAAGAGTCAATTTTACAGAGATTGATAAAAATTCAAAGAGCAAGACAACCAAT TTAAAGTAACTTATGACATGATAGGAGGATTAAGTAGCCAGCTGAAAGCAATTAGAGA AATAATTGAATTGCCCTCAAACAGCCTGAGCTTTTCAAGAGTTATGGAATTCCTGCC CCTAGAGGAGTGTTACTTTATGGTCTCCAGGTACTGGAAAAACAATGATCGCCAGGG CTGTTGCTAATGAAGTTGGAGCCTATGTTTCTGTAATTAATGGTCCTGAAATTATAAG CAAATTCATGCTGAGACTGAAGCAAAGTTACGTCAGATATTTGCTGAAGCCACTCTA CGACACCCATCAATTATTTTATTTGATGAGCTGGATGCACTTTGTCCGAAAAGAGAGG GGGCCCAGAATGAAGTGGAAGAAAGAGTTGTGGCTTCACTCTTAACACTGATGGATGG CATTGGTTCAGAAGTAAGTGAAGGACAAGTGTGGTTCTTGGGGCCACAAATCGCCCT CATGCCTTGATGCTGCTCTCCGAAGACCTGGGCGATTGATAAAGAGATTGAGATTG GAGTTCCCAATGCTCAGGACCGGCTAGATATTTCTCCAGAACTGCTTCCGAAGGGTACC CCATTTGCTCACTGAGGCTGAGCTGCTGCAGCTGGCAAATAGTGCTCATGGATACGTT GGAGCAGACTTGAAAGTCTTGTGTAATGAAGCAGGTCTCTGTGCCTTGCGGAGAATCC TGAAAAACAGCCTAACCTCCCTGATGTCAAGGTGGCTGGACTGGTGAAGATTACTCT GAAGGATTTCTTGCAGGCAATGAATGATATCAGACCCAGTGCCATGAGGGAAATAGCA ATTGATGTCCCAAATGTATCCTGGTCAGATATAGGAGGACTGGAAAGTATCAAAGTGA AGTTGGAACAGGCTGTGGAATGGCCCTTAAACATCCAGAGTCTTTCATTGGAATGGG TATTCAGCCACCTAAAGGAGTTCTTCTCTATGGGCCACCTGGGTGCTCTAAACAATG ATAGCAAAGGCTTTGGCCAATGAGAGTGGACTGAATTTTCTAGCTATAAAGGGGCTG AATTAATGAATAAATATGTTGGTGAATCTGAAAGAGCAGTTAGAGAGACCTTCCGAAA AGCAAGAGCAGTGGCGCCTTCCATTATTTTCTTTGATGAACTGGATGCCTTAGCAGTT GAAAGGGGCAGTTCTTTAGGTGCTGGGAATGTAGCCGATCGTGTTTTGGCTCAGCTCT TAACAGAAATGGATGGGATTGAACAGCTAAAGGATGTGACCATTTTGGCAGCTACTAA CCGTCCAGATAGGATAGACAAGGCTTTGATGCGGCCTGGAAGAATTGATAGAATCATC TATGTGCCTTTACCGGATGCAGCAACAAGAAGGGAAATATTTAAGCTGCAGTTTCACT CCATGCCTGTGAGTAATGAAGTTGACCTGGATGAACTCATCCTTCAAACCGACGCATA CTCAGGAGCAGAGATTGTAGCTGTCTGCAGAGAGGCAGCTCTTCTGGCTCTGGAAGAA GACATTCAAGCCAATCTCATCATGAAAGACATTTCACTCAGGCCTTGAGCACTGTGA CACCTAGAATTCCTGAGTCATTGAGACGTTTTTATGAAGATTATCAAGAGAAGAGTGG GCTGCATACACTCTGAGAAAATATATATATTCAAGATGCTGAAAATCCTTTCCAGAGA AAATT	
	ORF Start: ATG at 61	ORF Stop: TGA at 2740
	SEQ ID NO: 30	893 aa MW at 97931.2kD
NOV12b, CG102575-02 Protein Sequence	MSSKKNRKRLNQSAENGSSLPSAASSCVEARAPSAGSDFAATSGTLTVTNLLEKVDK IPKTFQNSLIHLGLNTMKSANICIGRPVLLTSLNGKQEVYTAWPMAGFPGGKVGLSEM AQKNVGVPRGDAIQVQPLVGAVLQAEEMDVALSDKDMEINEEELTGCLRLKLDGKIVL PGNFLYCTFYGRPYKLQVLRVKGADGMILGGPQSDSDTDAQRMAFEQSSMETSSLELS LQLSQLDLEDTPQIPSRSTPYKPIDDRITNKASDVLLDVTQSPGDGSGLMLEEVGTGLK CNFESAREGNEQLTEEERLLKFSIGAKCNTDTFYFISSTTRVNFTEIDKNSKEQDNQF	

	KVTYDMIGGLSSQLKAIREIIELPLKQPELFKSYGIPAPRGVLLYGPPGTGKTMIAARA VANEVGAYVSVINGPEIISKFYGETEAKLRQIFAEATLRHPSIIFIDELDALCPKREG AQNEVEKRVVASLLTMDGIGSEVSEGQVLVLGATNRPHALDAALRRPGRFDKEIEIG VPNAQDRLDILQKLLRRVPHLLTEAELLQLANSAHGYPVADLKVLCNEAGLCALRRIL KKQPNLPDVKVAGLVKITLKDFLQAMNDIRPSAMREIAIDVPNVSWSDIGGLESIKLK LEQAVEWPLKHPESFIRMGIQPPKGVLLYGPPGCSKTMIAKALANESGLNFLAIKGPE LMNKYVGESERAVRETFRKARAVAPSIIFDELDAVERGSSLGAGNVADRVLAQLL TEMDGIEQLKDVITLAATNRPDRIKALMRPGRIDRIIYVPLPDAATRREIFKLQFHS MPVSNEVDLDELILQTDAYSAGAEIVAVCREAALLALEEDIQANLIMKRHFTQALSTVT PRIPESLRRFYEDYQEKSLHTL
--	---

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

Table 12B. Comparison of NOV12a against NOV12b.		
Protein Sequence	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV12b	1..881	724/895 (80%)
	1..893	764/895 (84%)

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

Table 12C. Protein Sequence Properties NOV12a	
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:	No Known Signal Sequence Predicted

- 5 A search of the NOV12a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 12D.

Table 12D. Geneseq Results for NOV12a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU17209	Novel signal transduction pathway protein, Seq ID 774 - Homo sapiens, 574 aa. [WO200154733-A1, 02-AUG-2001]	261..823 2..574	442/575 (76%) 481/575 (82%)	0.0
AAB59399	Protein tyrosine phosphatase related sequence - Unidentified, 806 aa. [WO200075339-A1, 14-DEC-2000]	337..848 190..711	229/527 (43%) 340/527 (64%)	e-120
AAE09327	Human intracellular regulatory molecule, VCP - Homo sapiens, 806 aa. [US6274312-B1, 14-AUG-2001]	337..848 190..711	229/527 (43%) 340/527 (64%)	e-120
AAB05879	Human transitional endoplasmic reticulum ATPase protein sequence - Homo sapiens, 806 aa. [WO200034470-A1, 15-JUN-2000]	337..848 190..711	229/527 (43%) 340/527 (64%)	e-120
ABB59038	Drosophila melanogaster polypeptide SEQ ID NO 3906 - Drosophila melanogaster, 801 aa. [WO200171042-A2, 27-SEP-2001]	322..844 170..704	228/540 (42%) 342/540 (63%)	e-117

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
AAM00262	SPERMATOGENESIS ASSOCIATED FACTOR - Homo sapiens (Human), 893 aa.	1..881 1..893	745/895 (83%) 785/895 (87%)	0.0
Q9Z2K7	SPAF - Mus musculus (Mouse), 892 aa.	1..881 1..892	640/895 (71%) 721/895 (80%)	0.0
Q9CXZ7	2510048F20RIK PROTEIN - Mus musculus (Mouse), 893 aa.	1..881 1..893	640/896 (71%) 721/896 (80%)	0.0
Q8ZYN4	AAA FAMILY ATPASE, POSSIBLE CELL DIVISION CONTROL PROTEIN CDC48 - Pyrobaculum aerophilum, 731 aa.	356..876 184..714	265/537 (49%) 358/537 (66%)	e-136
Q58556	Cell division cycle protein 48 homolog MJ1156 - Methanococcus jannaschii, 903 aa.	309..855 127..697	271/578 (46%) 378/578 (64%)	e-136

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
AAA	378..566	95/217 (44%) 165/217 (76%)	3.3e-75
AAA	652..837	98/217 (45%) 165/217 (76%)	2e-77

5

Example 13.

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

Table 13A. NOV13 Sequence Analysis		
	SEQ ID NO: 31	420 bp
NOV13a,	TGCAGAAGGTGACCCTGGGCCTGCTTGTGTTCTTGGCAGGCTTTCCTGTCCTGGACGC	

CG102615-01 DNA Sequence	CAATGACCTAGAAGATAAAAAACAGTCCTTTCTACTATGACTGGCACAGCCTCCAGGTT GGCGGGCTCATCTGCGCTGGGGTTCTGTGCGCCATGGGCATCATCATCGTCATGAGTG CAAAATGCAAATGCAAGTTTGGCCAGAAGTCCGGTCACCATCCAGGGGAGACTCCACC TCTCATCACCCCAGGCTCAGCCCCAAAGCTGATGAGGACAGACCAGCTGAAATTGGGTG GAGGACCGTTCTCTGTCCCCAGGTCCTGTCTCTGCACAGAACTTGA ACTCCAGGATG GAATTCTTCCTCCTCTGCTGGGACTCCTTTGCATGGCAGGGCCTCATCTCACCTCTCG CAAGAGGGTCTCTT		
	ORF Start: at 3	ORF Stop: TGA at 261	
	SEQ ID NO: 32	86 aa	MW at 9131.6kD
NOV13a, CG102615-01 Protein Sequence	QKVTLLGLLVFLAGFPVLDANDLEDKNSPFYYDWHSLQVGGLICAGVLCAMGIIIVMSA KCKCKFGQKSGHHPGETPPLITPGSAQS		
	SEQ ID NO: 33	462 bp	
NOV13b, CG102615-04 DNA Sequence	TCAGCCTGGTGAACCACACAGAGGCTGGGGCGAGGAGGATACCATCTGTCA GTCTTGG CTGGATGACATCATGGGAAGGGGGTATAGTGGGGCCTTGCAGGCCAGAGGTGGCTTGG AGGAGCCCCTGGAAAGAGGCTTAAGAGGCCAGCGCTCTGACATGCAGAAGGTGACCCT GGGCCTGCTTGTGTTCTTGGCAGGCTTTCCTGTCTTGGACGCCAATGACCTAGAAGAT AAAAACAGTCCTTTCTACTATGACTGGCACAGCCTCCAGGTTGGCGGGCTCATCTGCG CTGGGGTTCTGTGCGCCATGGGCATCATCATCGTCATGAGTGCAAAATGCAAATGCAA GTTTGGCCAGAAGTCCGGTCACCATCCAGGGGAGACTCCACCTCTCATCACCCCAGGC TCAGCCCAAAGCTGATGAGGACAGACCAGCTGAAATTGGGTGGAGGACCGTTCTCT		
	ORF Start: ATG at 71	ORF Stop: TGA at 419	
	SEQ ID NO: 34	116 aa	MW at 12362.2kD
NOV13b, CG102615-04 Protein Sequence	MGRGYSGALQARGGLEEPLERGLRGQRSDMQKVTLLGLLVFLAGFPVLDANDLEDKNSP FYYDWHSLQVGGLICAGVLCAMGIIIVMSAKCKCKFGQKSGHHPGETPPLITPGSAQS		

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

Table 13B. Comparison of NOV13a against NOV13b.		
Protein Sequence	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV13b	1..86 31..116	86/86 (100%) 86/86 (100%)

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

Table 13C. Protein Sequence Properties NOV13a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.2000 probability located in lysosome (membrane); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 20 and 21

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

Table 13D. Geneseq Results for NOV13a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM23962	Human EST encoded protein SEQ ID NO: 1487 - Homo sapiens, 87 aa. [WO200154477-A2, 02-AUG-2001]	1..86 2..87	86/86 (100%) 86/86 (100%)	7e-47
AAW92959	Human MAT-8 protein - Homo sapiens, 87 aa. [WO9905276-A1, 04-FEB-1999]	1..86 2..87	86/86 (100%) 86/86 (100%)	7e-47
AAV48304	Human prostate cancer-associated protein 1 - Homo sapiens, 87 aa. [DE19811194-A1, 16-SEP-1999]	1..86 2..87	86/86 (100%) 86/86 (100%)	7e-47
AAR90990	Human Mat-8 polypeptide - Homo sapiens, 87 aa. [WO9605322-A1, 22-FEB-1996]	1..86 2..87	86/86 (100%) 86/86 (100%)	7e-47
AAB53415	Human colon cancer antigen protein sequence SEQ ID NO:955 - Homo sapiens, 150 aa. [WO200055351-A1, 21-SEP-2000]	1..86 39..150	86/112 (76%) 86/112 (76%)	2e-42

5 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
Q14802	FXYP domain-containing ion transport regulator 3 precursor (Chloride conductance inducer protein Mat-8) (Mammary tumor 8 kDa protein) (Phospholemman-like) - Homo sapiens (Human), 87 aa.	1..86 2..87	86/86 (100%) 86/86 (100%)	2e-46
Q61835	FXYP domain-containing ion transport regulator 3 precursor (Chloride conductance inducer protein Mat-8) (Mammary tumor 8 kDa protein) (Phospholemman-like) - Mus musculus (Mouse), 88 aa.	1..86 2..87	63/86 (73%) 72/86 (83%)	2e-33
O97797	FXYP domain-containing ion transport regulator 3 precursor (Chloride conductance inducer protein Mat-8) (Mammary tumor 8 kDa protein) - Sus scrofa (Pig), 88 aa.	2..84 3..85	60/83 (72%) 68/83 (81%)	8e-32
Q9D2W0	FXYP domain-containing ion transport regulator 4 precursor (Channel inducing factor) (CHIF) - Mus musculus (Mouse), 88 aa.	1..86 2..87	45/86 (52%) 59/86 (68%)	4e-21
Q63113	FXYP domain-containing ion transport regulator 4 precursor (Channel inducing factor) (CHIF) (Corticosteroid-induced protein) - Rattus norvegicus (Rat), 87 aa.	3..86 4..87	44/84 (52%) 55/84 (65%)	7e-20

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
ATP1G1_PLM_MAT8	19..74	27/57 (47%) 55/57 (96%)	2.7e-35

Example 14.

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

Table 14A. NOV14 Sequence Analysis			
	SEQ ID NO: 35		1638 bp
NOV14a, CG102646-01 DNA Sequence	TTATCTAATATGTTTGTTTTAGCTACATCTTTATCAAGCCAAGTGAATCCTGACTGGC GAACATATATCATGCTTGCAGTATATTTTCTAATTTTACTTGTTATTGGATATTATGG TTATAAGCAAGCAACCGGAGATTTAAGTGAATATATGCTTGGCGAAAGAAATATTGGT CCATATGTCACTGCCTTATCTGCCGGAGCTTCAGATATGAGCGGTTGGATGATTATGG GATTACCTGGAGAAGTTTATACTACAGGTTTATCAGCAGCATGGTTAGCTATTGGGTT AACTATCGGAGCTTATGTTAACTACATACTTGTAGCACCAAGACTTCGTGTGTACACT GAAAAAGCCAATGACTCAATTACATTGCCTAATTACTTTACACATCGTCTTAATGATA ATTCCAATATTATTAATAATTATCTCTGGTGGTATCATTGTTGTATTTTTTACACTCTA TACTCATTTCAGGTATGGTATCAGGTGGTAAATTATTTGATAGTGCTTTTGGTTTAGAC TATCATATTGGACTTATTTTAATCTCTGTCAATTGTAATTTTATATACTTTTTTTGGTG GCTATTTAGCAGTGTCGTTAACTGACTTTTTTCCAAGGGGTTGTCATGTTAATTGCGAT GGTTATGGTACCTATTGTAGCCATGATGCAGCTCGGAGGTATGGATGCTTTTTTCACAA GCAGCAACATTAAACCTACTAATTTAGATTTATTTAAAGGAACAACATTATATAGGCA TCATTTTCATTCTTTGCTTGGGGATTAGGCTATTTTGGCCAGCCTCATATCATTGTACG ATTTATGTCTATCAAATCCGTACGACAATTAAAAACGTCTAGAAGATTTGGTATTAGT TGGATGGCTATTAGTTTAATCGGTGCAGTATGTGTTGGATTAATTGGCATTTCGTTTG TACAAGATAAAGGTGTTGAATTAAGATCCAGAAACACTATTTATTTTAATGGGACA AATTTTATTCCATCCTCTTGTAGGTGGGTTTCTACTTGCAGCCATTTTGGCAGCAATT ATGAGTACGATTTCTTCCCAATTACTTGTGACTTCAAGTTCACCTACAGAAGATTTTT ACAAGTTAATTCGTGGTGAAGAAGCAGCAAAGCAACATAAGAAAGAATTTTTATTAGT GGGTCGATTATCTGTTGTAGTCGTTGCGATTATCTCCATCCTCATTGCATGGACGCCA AATGACACTATCTTAAATCTTGTTGGTAACGCTTGGGCTGGATTCGGTGCAGCATTTG GTCCACTGGTATTATTATCTCTCTATTTCGAAAGGTTTAAGTCGTACTGGAGCTATTT TGGAATGTTATCAGGAGCAATTGTCGTCATTCTTTGGATTGTGTTTGTAAACCATTA GGAGCATATAATGATTTCTTTAATTTATATGAAATTATTCCTGGTTTCTTAACAAGTC TTATTGTGACATATGTAGTGAGTCTTGTAACATAAAAGCCAGATCTCAATGTTCAAAA AGATTTAGAAGACGTCAAACGTATTGTAAAAGGACAATAAATTAATAATATTCAACGA TGCTTAATGTCAATATTATTTCAATTAGTGCATTACTCTTATAATATGAAACACAAAT AAATTTTTTATACAT		
	ORF Start: ATG at 10		ORF Stop: TAA at 1546
	SEQ ID NO: 36		512 aa MW at 55813.4kD
NOV14a, CG102646-01	MFVLATSLSSQVNPDWRTYIMLAVYFLILLVIGYYGYKQATGDLSEYMLGERNIGPYV TALSAGASDMSGWMIMGLPGEVYTTGLSAAWLAIGLTIGAYVNYILVAPRLRVYTEKA NDSITLPNYFTHRLNDNSNI IKIISGGIIVVFFTLYTHSGMVSGGKLFDSAFGLDYHI		

Protein Sequence	GLILISVIVILYTFEGGYLAVSLTDFQGVVMLIAMVMVPIVAMMQLGGMDAFSQAAT LKPTNLDLFGKTTIIGIISFFAWGLGYFGQPHIIVRFMSIKSVRQLKTSRRFGISWMA ISLIGAVCVGLIGISFVQDKGVELKDPETLFILMGQILFHPLVGGFLAAILAAIMST ISSQLLVTSSSLTDFYKLIRGEEAAKQHKKEFLLVGRLSVVVVAIISILIAWTPNDT ILNLVGNAWAGFGAAGPLVLLSLYSKGLSRTGAISGMLSGAIVVILWIVFVKPLGAY NDFFNLYEIIIPGFLTSLIVTYVVSIVTKKPDNLNVQKDLEDVKRIVKGQ
------------------	--

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

Table 14B. Protein Sequence Properties NOV14a	
PSort analysis:	0.8200 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:	Cleavage site between residues 37 and 38

5 A search of the NOV14a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 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
AAB76757	Corynebacterium glutamicum MCT protein SEQ ID NO:496 - Corynebacterium glutamicum, 524 aa. [WO200100805-A2, 04-JAN-2001]	20..506 9..499	233/502 (46%) 339/502 (67%)	e-127
AAG93195	C glutamicum protein fragment SEQ ID NO: 6949 - Corynebacterium glutamicum, 524 aa. [EP1108790-A2, 20-JUN-2001]	20..506 9..499	233/502 (46%) 339/502 (67%)	e-127
AAW20806	H. pylori transporter protein, 09ap20802orf27 - Helicobacter pylori, 446 aa. [WO9640893-A1, 19-DEC-1996]	64..506 5..445	208/450 (46%) 306/450 (67%)	e-112
AAG82596	S. epidermidis open reading frame protein sequence SEQ ID NO:2286 - Staphylococcus epidermidis, 408 aa. [WO200134809-A2, 17-MAY-2001]	266..510 163..407	171/245 (69%) 208/245 (84%)	1e-94
AAB96626	Putative P. abyssi permease #22 - Pyrococcus abyssi, 537 aa. [FR2792651-A1, 27-OCT-2000]	24..508 11..507	174/503 (34%) 275/503 (54%)	4e-83

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
Q99SY5	HIGH AFFINITY PROLINE PERMEASE - Staphylococcus aureus (strain Mu50 / ATCC 700699), and, 512 aa.	1..510 1..510	378/510 (74%) 443/510 (86%)	0.0
O30986	HIGH AFFINITY PROLINE PERMEASE - Staphylococcus aureus, 497 aa.	1..494 1..493	366/494 (74%) 431/494 (87%)	0.0
Q53584	PROLINE PERMEASE HOMOLOG - Staphylococcus aureus, 497 aa.	1..494 1..493	366/494 (74%) 430/494 (86%)	0.0
O06493	Osmoregulated proline transporter (Sodium/proline symporter) - Bacillus subtilis, 492 aa.	20..494 7..473	268/478 (56%) 371/478 (77%)	e-158
P94392	HOMOLOGUE OF PROLINE PERMEASE OF E. COLI - Bacillus subtilis, 449 aa.	54..504 1..442	243/452 (53%) 336/452 (73%)	e-142

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
SSF	47..447	134/449 (30%) 318/449 (71%)	5.7e-121

5

Example 15.

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

Table 15A. NOV15 Sequence Analysis		
	SEQ ID NO: 37	1146 bp
NOV15a,	CTAGCTCGACAGCTTCCCGGCGGCTGCGCGATGGACAGCCCCGAGGTGACCTTCACTC TCGCCTATCTGGTGTTCGCCGTGTGCTTCGTGTTACGCCCAACGAGTTCCACGCGGC	

CG102878-01 DNA Sequence	GGGGCTCACGGTGCAGAACCTGCTGTCTGGGGCTGGCTGGGCAGCGAGGACGCCGCCTTC GTGCCCTTCCACTTGCGCCGCACGGCCGCCACGCTGTTGTGCCACTCGCTGCTGCCGC TCGGCTACTATGTGGGCATGTGCCTTGC GGCTTCAGAAAAGCGGCTCCACGCCCTCAG CCAGGCCCCCTGAGGCCTGGCGGCTCTTCCTGCTGCTGGCCGTGACCCTCCCCTCCATC GCCTGCATCCTGATCTACTACTGGTCCCGTGACCGGTGGGCCTGCCACCCACTGGCGC GCACCCTGGCCCTCTACGCCCTCCCACAGTCTGGCTGGCAGGCTGTTGCCTCCTCTGT CAACACTGAGTTCCGGCGGATTGACAAGTTTGCCACCGGTGCACCAGGTGCCCGTGTG ATTGTGACAGACACGTGGGTGATGAAGGTAACCACCTACCGAGTGCACGTGGCCCAGC AGCAGGACGTGCACCTGACTGTGACGGAGTCTCGGCAGCATGAGCTCTCGCCAGACTC GAACTTGCCCGTGCAGCTCCTCACCATCCGTGTGGCCAGCACCAACCCTGCTGTGCAG GCCTTTGACATCAGGCTGAACTCCACTGAGTACGGGGAGCTCTGCGAGAAGCTCCGGG CACCCATCCGCAGGGCAGCCCATGTGGTCATCCACCAGAGCCTGGGCGACCTGTTCTT GGAGACATTTGCCTCCCTGGTAGAGGTCAACCCGGCCTACTCAGTGCCAGCAGCCAG GAGCTGGAGGCCTGCATAGGCTGCATGCAGACACGTGCCAGCGTGAAGCTGGTGAAGA CCTGCCAGGAGGCAGCCACAGGCGAGTGCCAGCAGTGTTACTGCCGCCCCATGTGGTG CCTCACCTGCATGGGCAAGTGGTTCGCCAGCCGCCAGGACCCCTGCGCCCTGACACC TGGCTGGCCAGCCGCGTGCCCTGCCCCACCTGCCGCGCACGCTTCTGCATCCTGGATG TGTGCACCGTGCCTGAGTGGGCTGGGGCCTTGAGGTGACTCTG		
	ORF Start: ATG at 31	ORF Stop: TGA at 1117	
	SEQ ID NO: 38	362 aa	MW at 40433.3kD
NOV15a, CG102878-01 Protein Sequence	MDSPEVTFTLAYLVFAVCFVFTPNEFHAAGLTVQNLLSGWLGSEDAAFVPFHLRRTAA TLLCHSLLPLGYVGMCLAASEKRLHALSQAPEAWRLFLLLA VTLPSIACILIYYWSR DRWACHPLARTLALYALPQSGWQAVASSVNTEFRRIDKFATGAPGARVIVTDTWVMKV TTYRVHVAQQQDVHLTVTESRQHELSPDSNLPVQLLTIRVASTNPAVQAFDIRLNSTE YGELCEKLRAPIRRAAHVVIHQSLGDLFLETFA SLVEVNPAYSVPSSELEACIGCMQ TRASVKLVKTCQEAATGECQQCYCRPMWCLTCMGKWFASRQDPLRPDTWLASRVPCPT CRARFCILDVCTVR		
	SEQ ID NO: 39	1115 bp	
NOV15b, CG102878-02 DNA Sequence	TTCGCCCTTGGCTGCGCGATGGACAGCCCCGAGGTGACCTTCACTCTCGCCTATCTGG TGTTTCGCCGTGTGCTTTCGTGTTACGCCCAACGAGTTCACGCGGGCGGGGCTCACGGT GCAGAACCTGCTGTCTGGGCTGGCTGGGCAGCGAGGACGCCGCTTCGTGCCCTTCCAC TTGCGCCGCACGGCCGCCACGCTGTTGTGCCACTCGCTGCTGCCGCTCGGCTACTACG TGGGCATGTGCCTTGC GGCTTCAGAAAAGCGGCTCCACGCCCTCAGCCAGGCCCTGA GGCCTGGCGGCTCTTCCTGCTGCTGGCCGTGACCCTCCCCTCCATTGCCTGCATCCTG ATCTACTACTGGTCCCGTGACCGGTGGGCCTGCCACCCACTGGCGCGCACCTGGCCC TCTACGCCCTCCCACAGTCTGGCTGGCAGGCTGTTGCCTCCTCTGTCAACACTGAGTT CCGGCGGATTGACAAGTTTGCCACCGGTGCACCAGGTGCCCGTGTGATTGTGACAGAC ACGTGGGTGATGAAGGTAACCACCTACCGAGTGCACGTGGCCCAGCAGCAGGACGTGC ACCTGACTGTGACGGAGTCTCGGCAGCATGAGCTCTCGCCAGACTCGAACTTGCCCGT GCAGCTCCTCACCATCCGTGTGGCCAGCACCAACCCTGCTGTGTCAGGCCTTTGACATC TGGCTGAACTCCACTGAGTACGGGGAGCTCTGCGAGAAGCTCCGGGCACCCATCCGCA GGGCAGCCCATGTGGTTCATCCACCAGAGCCTGGGCGACCTGTTCTTGAGACATTTGC CTCCCTGGTAGAGGTCAACCCGGCCTACTCAGTGCCCAGCAGCCAGGAGCTGGAGGCC TGCATAGGCTGCATGCAGACACGTGCCAGCGTGAAGCTGGTGAAGACCTGCCAGGAGG CAGCCACAGGCGAGTGCCAGCAGTGTTACTGCCGCCCCATGTGGTGCCTCACCTGCAT GGGCAAGTGGTTCGCCAGCCGCCAGGACCCCTGCGCCCTGACACCTGGCTGGCCAGC CGCGTGCCCTGCCCCACCTGCCGCGCACGCTTCTGCATCCTGGATGTGTGCACCGTGC GCTGATGTGGCGG		
	ORF Start: ATG at 19	ORF Stop: TGA at 1105	
	SEQ ID NO: 40	362 aa	MW at 40463.4kD
NOV15b, CG102878-02 Protein Sequence	MDSPEVTFTLAYLVFAVCFVFTPNEFHAAGLTVQNLLSGWLGSEDAAFVPFHLRRTAA TLLCHSLLPLGYVGMCLAASEKRLHALSQAPEAWRLFLLLA VTLPSIACILIYYWSR DRWACHPLARTLALYALPQSGWQAVASSVNTEFRRIDKFATGAPGARVIVTDTWVMKV TTYRVHVAQQQDVHLTVTESRQHELSPDSNLPVQLLTIRVASTNPAVQAFDIWLNSTE YGELCEKLRAPIRRAAHVVIHQSLGDLFLETFA SLVEVNPAYSVPSSELEACIGCMQ TRASVKLVKTCQEAATGECQQCYCRPMWCLTCMGKWFASRQDPLRPDTWLASRVPCPT		

	CRARFCILDVCTVR
--	----------------

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

Table 15B. Comparison of NOV15a against NOV15b.		
Protein Sequence	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV15b	1..362	361/362 (99%)
	1..362	361/362 (99%)

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

Table 15C. Protein Sequence Properties NOV15a	
PSort analysis:	0.6760 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:	Cleavage site between residues 29 and 30

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

Table 15D. Geneseq Results for NOV15a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG81377	Human AFP protein sequence SEQ ID NO:272 - Homo sapiens, 362 aa. [WO200129221-A2, 26-APR-2001]	1..362 1..362	360/362 (99%) 360/362 (99%)	0.0
ABB69639	Drosophila melanogaster polypeptide SEQ ID NO 35709 - Drosophila melanogaster, 409 aa. [WO200171042-A2, 27-SEP-2001]	1..358 1..383	122/389 (31%) 200/389 (51%)	7e-60
AAG23427	Arabidopsis thaliana protein fragment SEQ ID NO: 26729 - Arabidopsis thaliana, 284 aa. [EP1033405-A2, 06-SEP-2000]	337..362 77..102	13/26 (50%) 16/26 (61%)	2.8
AAG23426	Arabidopsis thaliana protein fragment SEQ ID NO: 26728 - Arabidopsis thaliana, 413 aa. [EP1033405-A2, 06-SEP-2000]	337..362 206..231	13/26 (50%) 16/26 (61%)	2.8
ABG11786	Novel human diagnostic protein #11777 - Homo sapiens, 198 aa. [WO200175067-A2, 11-OCT-2001]	285..354 54..141	23/89 (25%) 37/89 (40%)	3.6

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.

Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC38627	SEQUENCE 271 FROM PATENT WO0129221 - Homo sapiens (Human), 362 aa.	1..362 1..362	361/362 (99%) 361/362 (99%)	0.0
Q9DCF3	0610039G24RIK PROTEIN - Mus musculus (Mouse), 362 aa.	1..362 1..362	323/362 (89%) 341/362 (93%)	0.0
Q96GP5	SIMILAR TO RIKEN CDNA 0610039G24 GENE - Homo sapiens (Human), 232 aa.	1..226 1..226	226/226 (100%) 226/226 (100%)	e-129
Q9VN16	CG14646 PROTEIN - Drosophila melanogaster (Fruit fly), 409 aa.	1..358 1..383	122/389 (31%) 200/389 (51%)	2e-59
Q95TM4	LD39811P - Drosophila melanogaster (Fruit fly), 393 aa.	20..358 4..367	116/370 (31%) 190/370 (51%)	1e-55

5

Table 16A. NOV16 Sequence Analysis		
	SEQ ID NO: 41	2765 bp
NOV16a, CG103459-01 DNA Sequence	CTGGCGGCGTCGCATGGAGGGCTCTGGGGGCGGTGCGGGCGAGCGGGCGCCGCTGCTG GGCGCGCGGCGGGCGGCGGCGGCCGCGGCGGCGGCTGGGGCGTTTCGCGGGCCGGCGCG CGGCGTGCGGGGCCGTGCTGCTGACGGAGCTGCTGGAGCGCGCCGCTTTCTACGGCAT CACGTCCAACCTGGTGCTATTCCTGAACGGGGCGCCGTTCTGCTGGGAGGGCGCGCAG GCCAGCGAGGCGCTGCTGCTCTTCATGGGCCTCACCTACCTGGGCTCGCCGTTTCGGAG GCTGGCTGGCCGACGCGCGGCTGGGCCGGGCGCGGCCATCCTGCTGAGCCTGGCGCT CTACCTGCTGGGCATGCTGGCCTTCCCGCTGCTGGCCGCGCCCGCCACGCGAGCCGCG CTCTGCGGTTCCGCGCGCCTGCTCAACTGCACGGCGCCTGGTCCCGACGCCGCCGCC GCTGCTGCTCACC GGCCACCTTCGCGGGGCTGGTGCTGGTGGGCCTGGGCGTGCCAC CGTCAAGGCCAACATCACGCCCTTCGGCGCCGACCAGGTTAAAGATCGAGGTCCGGAA GCCACTAGGAGATTTTTTAATTGGTTTTATTGGAGCATTAACTGGGAGCGATCCTGT CGTTAGGTGGCATTGCCCTATATT CAGCAGAACGTCAGCTTTGTCACTGGTTATGCGAT CCCCACTGTCTGCGTCGGCCTTGCTTTTGTGGCCTTCCTCTGTGGCCAGAGCGTTTTTC ATCACCAAGCCTCCTGATGGCAGTGCCTTCACCGATATGTTCAAGATACTGACGTATT CCTGCTGTTCC CAGAAGCGAAGTGGAGAGCGCCAGAGTAATGGTGAAGGCATTGGAGT CTTTCAGCAATCTTCTAAACAAAGTCTGTTTGATT CATGTAAGATGTCTCATGGTGGG CCATTTACAGAAGAGAAAGTGG AAGATGTGAAAGCTCTGGTCAAGATTGTCCCTGTTT TCTTGGCTTTGATACCTTACTGGACAGTGATTTTCAAATGCAGACAACATATGTTTT ACAGAGTCTTCATTTGAGGATTCCAGAAATTTCAAATATTACAACCACTCCTCACACG CTCCCTGCAGCCTGGCTGACCATGTTTGATGCTGTGCTCATCCTCCTGCTCATCCCTC TGAAGGACAAACTGGTCGATCCCATTTTTGAGAAGACATGGCCTGCTCCCATCCTCCCT	

	GAAGAGGATCGCCGTGGGCATGTTCTTTGTCATGTGCTCAGCCTTTGCTGCAGGAATT TTGGAGAGTAAAAGGCTGAACCTTGTTAAAGAGAAAACCATTAATCAGACCATCGGCA ACGTCGTCTACCATGCTGCCGATCTGTCGCTGTGGTGGCAGGTGCCGCAGTACTTGCT GATTGGGATCAGCGAGATCTTTGCAAGTATCGCAGGCCTGGAATTTGCATACTCAGCT GCCCCAAGTCCATGCAGAGTGCCATAATGGGCTTGTTCTTTTCTTCTCTGGCGTCG GGTCGTTTCGTGGGTTCTGGACTGCTGGCACTGGTGTCTATCAAAGCCATCGGATGGAT GAGCAGTCACACAGACTTTGGTAATATTAACGGCTGCTATTTGAACTATTACTTTTTTT CTTCTGGCTGCTATTCAAGGAGCTACCCTCCTGCTTTTCCTCATTATTTCTGTGAAAT ATGACCATCATCGAGACCATCAGCGATCAAGAGCCAATGGCGTGCCACCAGCAGGAG GGCCTGACCTTCCTGAGGCCATGTGCGGTTTCTGAGGCTGACATGTCAGTAACTGACT GGGGTGCACTGAGAACAGGCAAGACTTTAAATTCCCATAAAATGTCTGACTTCACTGA AACTTGCAATGTTGCCTGGATTGATTTCTTCTTTCCCTCTATCCAAAGGAGCTTGGTAA GTGCCTTACTGCAGCGTGTCTCCTGGCACGCTGGGCCCTCCGGGAGGAGAGCTGCAGA TTTCGAGTATGTCGCTTGTCATTCAAGGTCTCTGTGAATCCTCTAGCTGGGTTCCCTT TTTTACAGAACTCACAAATGGAGATTGCAAAGTCTTGGGGAACCTCCACGTGTTAGTT GGCATCCCAGTTTCTTAAACAAATAGTATCACCTGCTTCCCATAGCCATATCTCACTG TAAAAAAAAAATAATAAACTGTTACTTATATTTAAGAAAGTGAGGATTTTTTTTTTT TTAAAGATAAAAGCATGGTCAGATGCTGCAAGGATTTTACATAAATGCCATATTTATG GTTTCCTTCCTGAGAACAATCTTGCTCTTGCCATGTTCTTTGATTTAGGCTGGTAGTA AACACATTTTCATCTGCTGCTTCAAAAAGTACTTACTTTTTTAAACCATCAACATTACTT TTCTTTCTTAAGGCAAGGCATGCATAAGAGTCATTTGAGACCATGTGTCCCATCTCAA GCCACAGAGCAACTCACGGGGTACTTCACACCTTACCTAGTCAGAGTGCTTATATATA GCTTTATTTTGGTACGATTGAGACTAAAGACTGATCATGGTTGTATGTAAGGAAAACA TTCTTTTGAACAGAAATAGTGTAATTAATAAATAATTGAAAGTGTTAAATGTGAACCTG AGCTGTTTGACCAGTCACATTTTTGTATTGTTACTGTACGTGTATCTGGGGCTTCTCC GTTTGTTAATACTTTTTCTGTATTTGTTGCTGTATTTTTGGCATAACTTTATTATAAA AAGCATCTCAAATGCGAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 14	ORF Stop: TGA at 1745	
	SEQ ID NO: 42	577 aa	MW at 62004.6kD
NOV16a, CG103459-01 Protein Sequence	MEGSGGGAGERAPLLGARRAAAAAAAGAFAGRAACGAVLLTELLERAIFYGITSNL VLFLNGAPFCWEGAQASEALLLFMGLTYLGSPPGGWLADARLGRARAILLSLALYLLG MLAFPLLAAPATRAALCGSARLLNCTAPGPDAAARCCSPATFAGLVLVGLGVATVKAN ITPFGADQVKDRGPEATRRFFNWFWYSINLGAILSLGGIAYIQQNVSFVTGYAIPVC VGLAFVAFLCGQSVFITKPPDGSAFTDMFKILTYSQKRSGERQSNNEGIGVFQQS SKQSLFDSCKMSHGGPFTEEKVEDVKALVKIVPVFLALIPYWTVYFQMOTTYVLQSLH LRIPEISNITTTTPHTLPAAWLTMFDAVLILLIPLKDKLVDPILRRHGLLPSSLKRIA VGMFFVMCSAFAAGILESKRLNLVKEKTINQITIGNVVYHAADLSLWWQVPQYLLIGIS EIFASIAGLEFAYSAPKSMQSAIMGLFFFFSGVGSFVGSGLLALVSIKAIGWMSST DFGNINGCYLNYFFLLAAIQGATLLLFLIISVKYDHHRDHQSRANGVPTSRA		

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

Table 16B. Protein Sequence Properties NOV16a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 16C. Geneseq Results for NOV16a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU12071	Human PHT1 variant protein from Caco-2 cells - Homo sapiens, 577 aa. [WO200192468-A2, 06-DEC-2001]	1..577 1..577	577/577 (100%) 577/577 (100%)	0.0
AAU12068	Human PHT1 protein isolated from Caco-2 cells - Homo sapiens, 577 aa. [WO200192468-A2, 06-DEC-2001]	1..577 1..577	577/577 (100%) 577/577 (100%)	0.0
AAU12070	Human PHT1 variant protein from BeWo cells - Homo sapiens, 577 aa. [WO200192468-A2, 06-DEC-2001]	1..577 1..577	575/577 (99%) 576/577 (99%)	0.0
AAE16771	Human transporter and ion channel-8 (TRICH-8) protein - Homo sapiens, 576 aa. [WO200192304-A2, 06-DEC-2001]	1..577 1..576	576/577 (99%) 576/577 (99%)	0.0
AAB82821	Human proton/oligonucleotide transporter hPHT1 polypeptide - Homo sapiens, 556 aa. [WO200160854-A1, 23-AUG-2001]	22..577 1..556	555/556 (99%) 555/556 (99%)	0.0

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 16D.

Table 16D. Public BLASTP Results for NOV16a				
Protein Accession Number	Protein/Organism/Length	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O09014	PEPTIDE/HISTIDINE TRANSPORTER - Rattus norvegicus (Rat), 572 aa.	9..576 3..571	500/578 (86%) 531/578 (91%)	0.0
Q91W98	SIMILAR TO PEPTIDE TRANSPORTER 3 - Mus musculus (Mouse), 574 aa.	9..576 3..573	496/578 (85%) 531/578 (91%)	0.0
AAH28394	SIMILAR TO PEPTIDE TRANSPORTER 3 - Homo sapiens (Human), 461 aa.	117..577 1..461	460/461 (99%) 460/461 (99%)	0.0
Q9P2X9	PEPTIDE TRANSPORTER 3 - Homo sapiens (Human), 581 aa.	9..558 14..564	289/570 (50%) 379/570 (65%)	e-152
Q9WU80	CAMP INDUCIBLE 1 PROTEIN - Mus musculus (Mouse), 578 aa.	8..567 6..570	279/577 (48%) 366/577 (63%)	e-144

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

Table 16E. Domain Analysis of NOV16a			
Pfam Domain	NOV16a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PTR2	103..496	109/448 (24%) 310/448 (69%)	6.7e-103

5 Example 17.

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

Table 17A. NOV17 Sequence Analysis		
	SEQ ID NO: 43	1393 bp
NOV17a, CG104210-01 DNA Sequence	CCCATGGAGGCTCCGGGACCCCGCGCCTTGCGGACTGCGCTCTGTGGCGGCTGTTGCT GCCTCCTCCTATGTGCCAGCTGGCTGTGGCTGGTAAAGGAGCTCGAGGCTTTGGGAG GGGAGCCCTGATCCGCCTGAATATCTGGCCGGCGGTCCAAGGGGCTGCAAACAGCTG GAGGTCTGTGAGCACTGCGTGGAGGGAGACAGAGCGCGCAATCTCTCCAGCTGCATGT	

	GGGAGCAGTGCCGGCCAGAGGAGCCAGGTCACTGTGTGGCCCAATCTGAGGTGGTCAA GGAAGGTTGCTCCATCTACAACCGCTCAGAGGCATGTCCAGCTGCTCACCACCACCCC ACCTATGAACCGAAGACAGTCACAACAGGTAGCCCCCAGTCCCTGAGGCCACAGCC CTGGATTTGACGGGGCCAGCTTTATCGGAGGTGTCGTGCTGGTGTGAGCCTACAGGC GGTGGCTTTCTTTGTGCTGCACTTCCTCAAGGCCAAGGACAGCACCTACCAGACGCTG TGAGTACCTGGCCAGCAGCAAGTACCTGAGTCCCAGCTCACCTCCTGGTTCTGCCCC ACCGTTCCCCTTCAGTACCCAGGGTGCTGTCTTCTCCACTGGCAAGCCCTCAGGACGG TGACAGCGTGCTCCATGTGAGCCACACCCCTTTTGTCTCCTCCAGTTGGGGTGTTTCC TTTGTGATGATGTTGGCTGGGACCAGGACTCAGCCTGGGCCAGTCTAGGAGCCCAGCTG AGCCCTCCTGTGTCTTTTCCCTTCATGCTGCCAGCAGGGAAGAGAACCAGTAGGTGCC AGCCCAGCAACCTGTGGCCCGCGTTTCTGTGGCTGTGGGCAGGAGCTGGGCCTTGTTGT CTAGTTGGGTTTTGCTCTGAGAAGGGGAGCTGTGCTGAGGCCCTCTGTGTGCCGTGTG TGCTGTGGGGCGGGTCGCCACAGCCTGTGTTAAAGTGTGTTGCTCTTCTCTGCTGCCT CCTCTCGAGGCAGGGGGTCTTGGCTGGCTGAGGCAGTGTACCTTCTGAGTGTCTCT CTTTGGCCTCTGCAGAATCTGACCCCTTTGGGCCTGGACTCCATCCTGAGGGGAAAGG AGGATGCAGAGGGTGGCCTCTGGGCACCCTTGTGGGTAAAGCGGGGGGCGGGGCGGGA AAACTCTGGCCGCCAGTTTTTTGGCTCCTGCGGGCACCAAGCAGGCTCAGTGTCTGAT GCTTGACATCTCCTCCTGTCCTGGGCCTGGAACCTGCAGCTGAGAAAATCCCTCAACC ACCTCGTCTCCTCCATCGCCCCCTGCTGGGCCCCCAGCCTGACAGTGGGTTGTATGCC TGCTCTTTCCACCAACTGGCCTGGGCACTGCCCCCAAATAAAGGAACTCTGCACTGC A		
	ORF Start: ATG at 4	ORF Stop: TGA at 523	
	SEQ ID NO: 44	173 aa	MW at 18421.0kD
NOV17a, CG104210-01 Protein Sequence	MEAPGPRLRTALCGGCCLLLLCAQLAVAGKGARGFGRGALIRLNIWPAVQGACKQLE VCEHCVEGDRARNLSSCMWEQCRPEEPGHCVAQSEVVKEGCSIYNRSEACPAHHHPT YEPKTVTTGSPVPPEAHSPGFDGASFIGGVVLVLSLQAVAFFVLHFLKAKDSTYQTL		
	SEQ ID NO: 45	561 bp	
NOV17b, CG104210-02 DNA Sequence	CCCATGGAGGCTCCGGGACCCCGCGCCTTGCGGACTGCGCTCTGTGGCGGCTGTTGCT GCCTCCTCCTATGTGCCAGCTGGCTGTGGCTGGTAAAGGAGCTCGAGGCTTTGGGAG GGGAGCCCTGATCCGCCTGAATATCTGGCCGGCGGTCCAAGGGGCCTGCAAACAGCTG GAGGTCTGTGAGCACTGCGTGGAGGGAGACAGAGCGCGCAATCTCTCCAGCTGCGTGT GGGAGCAGTGCCGGCCAGAGGAGCCAGGACACTGTGTGGCCCAATCTGAGGTGGTCAA GGAAGGTTGCTCCATCTACAACCGCTCAGAGGCATGTCCAGCTGCTCACCACCACCCC ACCTATGAACCGAAGACAGTCACAACAGGGAGCCCCCAGTCCCTGAGGCCACAGCC CTGGATTTGACGGGGCCAGCTTTATCGGAGGTGTCGTGCTGGTGTGAGCCTACAGGC GGTGGCTTTCTTTGTGCTGCACTTCCTCAAGGCCAAGGACAGCACCTACCAGACGCTG TGAGTACCTGGCCAGCAGCAAGTACCTGAGTCCCAGCTC		
	ORF Start: ATG at 4	ORF Stop: TGA at 523	
	SEQ ID NO: 46	173 aa	MW at 18389.0kD
NOV17b, CG104210-02 Protein Sequence	MEAPGPRLRTALCGGCCLLLLCAQLAVAGKGARGFGRGALIRLNIWPAVQGACKQLE VCEHCVEGDRARNLSSCVWEQCRPEEPGHCVAQSEVVKEGCSIYNRSEACPAHHHPT YEPKTVTTGSPVPPEAHSPGFDGASFIGGVVLVLSLQAVAFFVLHFLKAKDSTYQTL		
	SEQ ID NO: 47	349 bp	
NOV17c, 272249075 DNA Sequence	CACCGGATCCGGTAAAGGAGCTCGAGGCTTTGGGAGGGGAGCCCTGATCCGCCTGAAT ATCTGGCCGGCGGTCCAAGGGGCCTGCAAACAGCTGGAGGTCTGTGAGCACTGCGTGG AGGGAGACAGAGCGCGCAATCTCTCCAGCTGCATGTGGGAGCAGTGCCGGCCAGAGGA GCCAGGACACTGTGTGGCCCAATCTGAGGTGGTCAAGGAAGGTTGCTCCATCTACAAC CGCTCAGAGGCATGTCCAGCTGCTCACCACCACCCACCTATGAACCGAAGACAGTCA CAACAGGGAGCCCCCAGTCCCTGAGGCCACAGCCCTGGATTTGACGGGGTTCGACGG C		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 48	116 aa	MW at 12383.7kD

WO 02/099062

PCT/US02/17559

NOV17c, 272249075 Protein Sequence	TGSGKGARGFGRGALIRLNIWPAVQGACKQLEVCEHCVEGDRARNLSSCMWEQCRPEE PGHCVAQSEVVKEGCSIYNRSEACPAHHHTYEPKTVTTGSPPVPEAHSPGFDGVDG
--	--

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 17B.

Table 17B. Comparison of NOV17a against NOV17b and NOV17c.		
Protein Sequence	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV17b	1..173	139/173 (80%)
	1..173	140/173 (80%)
NOV17c	41..139	99/99 (100%)
	15..113	99/99 (100%)

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

Table 17C. Protein Sequence Properties NOV17a	
PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 30 and 31

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

Table 17D. Geneseq Results for NOV17a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE03827	Human gene 10 encoded secreted protein HBINS58, SEQ ID NO: 73 - Homo sapiens, 173 aa. [WO200136440-A1, 25-MAY-2001]	1..173 1..173	173/173 (100%) 173/173 (100%)	e-103
AAE03852	Human gene 10 encoded secreted protein HBINS58, SEQ ID NO:98 - Homo sapiens, 210 aa. [WO200136440-A1, 25-MAY-2001]	1..160 1..160	159/160 (99%) 159/160 (99%)	5e-94
AAB58415	Lung cancer associated polypeptide sequence SEQ ID 753 - Homo sapiens, 214 aa. [WO200055180-A2, 21-SEP-2000]	73..173 95..214	41/124 (33%) 56/124 (45%)	8e-10
AAG03771	Human secreted protein, SEQ ID NO: 7852 - Homo sapiens, 197 aa. [EP1033401-A2, 06-SEP-2000]	73..173 78..197	38/124 (30%) 52/124 (41%)	1e-07
ABB65987	Drosophila melanogaster polypeptide SEQ ID NO 24753 - Drosophila melanogaster, 183 aa. [WO200171042-A2, 27-SEP-2001]	116..173 127..183	29/60 (48%) 35/60 (58%)	9e-05

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

Table 17E. Public BLASTP Results for NOV17a				
Protein Accession Number	Protein/Organism/Length	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9D6W7	2310047N01RIK PROTEIN - Mus musculus (Mouse), 172 aa.	1..173 1..171	140/173 (80%) 150/173 (85%)	1e-82
Q9BPV0	CD164 ISOFORM DELTA 4 - Homo sapiens (Human), 184 aa.	73..173 78..184	41/111 (36%) 56/111 (49%)	6e-11
Q9CVT7	CD164 ANTIGEN - Mus musculus (Mouse), 161 aa (fragment).	25..173 5..161	51/173 (29%) 67/173 (38%)	2e-10
Q9QX82	ENDOLYN PRECURSOR - Rattus norvegicus (Rat), 195 aa.	54..173 57..195	41/140 (29%) 59/140 (41%)	4e-10
Q9Z317	MGC-24V - Mus musculus (Mouse), 197 aa.	54..173 58..197	44/144 (30%) 58/144 (39%)	7e-10

Example 18.

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

Table 18A. NOV18 Sequence Analysis			
	SEQ ID NO: 49	788 bp	
NOV18a, CG104251-01 DNA Sequence	CTTTTGCCTTTATGCAACCAACATGGAGATTTTGTACCATGTCCTGTTCTTAGTGCTT GAATGTCCTAACCTGAAGCTGAAGAAGCCGCCCTGGCTGCACATGCTGTGCGCCATGA CTGTATGCTCTGGTGGTGGTGTCTTCCTCATTACCGGAGGAATCATTTATGATGTTAT TGTTGAACCTCCAAGTGTGGCTCTATGACTGATGAACATGGGCATCAGAGGCCAGTA GCTTTCTTTGCCTATAGAGTAAATGGACAATATATTATGGAAGGACTTGCATCCAGCT TCCTGTTTACAATGGGAGGTTTAGGTTTCATAATCCTGGACCAATTGAATGCACCAAA TATCCCAAACTCAATAGATTTCTTCTTCTATTTCATTGGATTTGTCTGTGTTCTATTG AGTATTTTTCATGGCTAGAGTATTCATGAGAATGAACTGCCGAGCTATCTGATGGGTT AGAGTGCCTTTGAGAAGAAATCAGTGGATACTGGATTTTTTCTTGTCAATGAAGTTTT AAAGGCTGTACCAATCCTCTAATATGAAATGTGGAAAAGAATGAAGAGCAGCAGTAAA AGAAATATCTAGTGAAAAACAGGAAGCGTATTGAAGCTTGGACTAGAATTTCTTCTT GGTATTAAAGAGACAAGTTTATCACAGAATTTTTTTTCTGCTGGCCTATTGCTATAC CAATGATGTTGAGTGGCATTTTCTTTTCTAGTTTTTTCATTAAAATATATTCATATCTA CAACTATAATATCAAATAAAGTGATTATTTTTTA		
	ORF Start: ATG at 23	ORF Stop: TAG at 464	
	SEQ ID NO: 50	147 aa	MW at 16447.7kD
NOV18a, CG104251-01 Protein Sequence	MEILYHVLFLVLECPNLKLLKPPWLHMLSAMTVCSGGVFLITGGIIYDVIVEPPSVG SMTDEHGHQRPVAFFAYRVNGQYIMEGLASSFLFTMGGLGFIILDQLNAPNIPKLNRF LLLFIGFVCVLLSIFMARVFMRMKLPSYLMG		

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

Table 18B. Protein Sequence Properties NOV18a	
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:	Cleavage site between residues 42 and 43

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

Table 18C. Geneseq Results for NOV18a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY53631	A bone marrow secreted protein designated BMS155 - Homo sapiens, 149 aa. [WO9933979-A2, 08-JUL-1999]	1..147 1..149	133/149 (89%) 135/149 (90%)	1e-69
AAY53042	Human secreted protein clone pu282_10 protein sequence SEQ ID NO:90 - Homo sapiens, 149 aa. [WO9957132-A1, 11-NOV-1999]	1..147 1..149	133/149 (89%) 135/149 (90%)	1e-69
AAB12143	Hydrophobic domain protein isolated from WERI-RB cells - Homo sapiens, 149 aa. [WO200029448-A2, 25-MAY-2000]	1..147 1..149	133/149 (89%) 135/149 (90%)	1e-69
AAY59670	Secreted protein 108-005-5-0-F6-FL - Homo sapiens, 149 aa. [WO9940189-A2, 12-AUG-1999]	1..147 1..149	133/149 (89%) 135/149 (90%)	1e-69
AAY60146	Human endometrium tumour EST encoded protein 206 - Homo sapiens, 171 aa. [DE19817948-A1, 21-OCT-1999]	1..147 23..171	133/149 (89%) 135/149 (90%)	1e-69

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 18D.

Table 18D. Public BLASTP Results for NOV18a				
Protein Accession Number	Protein/Organism/Length	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NRP0	DC2 (HYDROPHOBIC PROTEIN HSF-28) (HYPOTHETICAL 16.8 KDA PROTEIN) - Homo sapiens (Human), 149 aa.	1..147 1..149	133/149 (89%) 135/149 (90%)	3e-69
Q9P075	HSPC307 - Homo sapiens (Human), 167 aa (fragment).	1..147 19..167	133/149 (89%) 135/149 (90%)	3e-69
Q9CPZ2	2310008M10RIK PROTEIN (RIKEN CDNA 2310008M10 GENE) - Mus musculus (Mouse), 149 aa.	1..147 1..149	132/149 (88%) 135/149 (90%)	6e-69
Q9P1R4	HDCMD45P - Homo sapiens (Human), 160 aa (fragment).	1..147 12..160	132/149 (88%) 134/149 (89%)	2e-68
AAH24224	SIMILAR TO DC2 PROTEIN - Homo sapiens (Human), 119 aa.	40..147 12..119	96/108 (88%) 100/108 (91%)	7e-50

Example 19.

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

Table 19A. NOV19 Sequence Analysis		
	SEQ ID NO: 51	3761 bp
NOV19a, CG104934-01 DNA Sequence	GGGCGCGCCGAGCCGGGCGCGGGGGCGCTGAACGGCGGAGCGGGAGCGGCCGGAGGAG CCATGGACTGCAGCCTCGTGCGGACGCTCGTGACAGATACTGTGCAGGAGAAGAGAA TTGGGTGGACAGCAGGACCATCTACGTGGGACACAGGGAGCCACCTCCGGGCGCAGAG GCCTACATCCACAGAGATAACCCAGACAACAGGATCGTCTCGTCCAAGTACACATTTT GGAAC TTTATACCCAAGAATTTATTTGAACAATTCAGAAGAGTAGCCAAC TTTTATTT CCTTATCATATTTCTGGTGCAGTTGATTATTGATACACCCACAAGTCCAGTGACAAGC GGACTTCCACTCTTCTTTGTCTATTACTGTGACGGCTATCAAACAGGGTTATGAAGACT GGCTTCGACATAAAGCAGACAATGCCATGAACCAGTGTCTGTTTCATTTTCATTCAGCA CGGCAAGCTCGTTTCGAAACAAAGTCGAAAGCTGCGAGTTGGGGACATTGTTCATGGTT AAGGAGGACGAGACCTTTCCCTGCGACTTGATCTTCCTTTCCAGCAACCGGGGAGATG GGACGTGCCACGTCACCACCGCCAGCTTGGATGGAGAATCCAGCCATAAAACGCATTA CGCGGTCCAGGACACCAAAGGCTTCCACACAGAGGAGGATATCGGCGGACTTCACGCC ACCATCGAGTGTGAGCAGCCCCAGCCGACCTCTACAAGTTCGTGGGTGCGATCAACG TTTACAGTGACCTGAATGACCCCGTGGTGAGGCCCTTAGGATCGGAAAACCTGCTGCT TAGAGGAGCTACACTGAAGAACACTGAGAAAATCTTTGGTGTGGCTATTTACACGGGA ATGGAAACCAAGATGGCATTAAATTATCAATCAAATCTCAGAAGCGATCTGCCGTGG AAAAATCGATGAATGCGTTCCTCATTTGTGTATCTCTGCATTCTGATCAGCAAAGCCCT GATAAACACTGTGCTGAAATACATGTGGCAGAGTGAGCCCTTTCGGGATGAGCCGTGG	

	TATAATCAGAAAACGGAGTCGGAAAGGCAGAGGAATCTGTTCCCTCAAGGCATTCACGG ACTTCCTGGCCTTCATGGTCCTCTTTAACTACATCATCCCTGTGTCCATGTACGTCAC GGTCGAGATGCAGAAGTTCCTCGGCTCTTACTTCATCACCTGGGACGAAGACATGTTT GACGAGGAGACTGGCGAGGGGCTCTGGTGAACACGTCGGACCTCAATGAAGAGCTGG GACAGGTGGAGTACATCTTCACAGACAAGACCGGCACCCTCACGGAAAAACAACATGGA GTTCAAGGAGTGCTGCATCGAAGGCCATGTCTACGTGCCCCACGTCTCTGCAACGGG CAGGTCCTCCCGAGATCGTCAGGAATCGACATGATTGACTCGTCCCCCAGCGTCAACG GGAGGGAGCGCGAGGAGCTGTTTTTCCGGGCCCTCTGTCTCTGCCACACCGTCCAGGT GAAAGACGATGACAGCGTAGACGGCCCCAGGAAATCGCCGGACGGGGGGAAATCCTGT GTGTACATCTCATCCTCGCCCCGACGAGGTGGCGCTGGTTCGAAGGTGTCCAGAGACTTG GCTTTACCTACCTAAGGCTGAAGGACAATTACATGGAGATATTAAACAGGGAGAACCA CATCGAAAGGTTTGAATTGCTGGAAATTTTGAGTTTTGACTCAGTCAGAAGGAGAATG AGTGTAATTGTAAAATCTGCTACAGGAGAAATTTATCTGTTTTGCAAAGGAGCAGATT CTTCGATATTCCCCCGAGTGATAGAAGGCAAAGTTGACCAGATCCGAGCCAGAGTGGA GCGTAACGCAGTGGAGGGGCTCCGAACCTTGTGTGTTGCTTATAAAAGGCTGATCCAA GAAGAATATGAAGGCATTTGTAAGCTGCTGCAGGCTGCCAAAGTGGCCCTTCAAGATC GAGAGAAAAAGTTAGCAGAAGCCTATGAGCAAATAGAGAAAGATCTTACTCTGCTTGG TGCTACAGCTGTTGAGGACCGGCTGCAGGAGAAAGCTGCAGACACCATCGAGGCCCTG CAGAAGGCCGGGATCAAAGTCTGGGTCTCACGGGAGACAAGATGGAGACGGCCGCGG CCACGTGCTACGCCTGCAAGCTCTTCGCAGGAACACGCAGCTGCTGGAGCTGACCAC CAAGAGGATCGAGGAGCAGAGCCTGCACGACGTCTGTTCGAGCTGAGCAAGACGGTC CTGCGCCACAGCGGGAGCCTGACCAGAGACAACCTGTCCGGACTTTCAGCAGATATGC AGGACTACGGTTTAAATTATCGACGGAGCTGCACTGTCTCTGATAATGAAGCCTCGAGA AGACGGGAGTTCCGGCAACTACAGGGAGCTCTTCTGGAATCTGCCGGAGCTGCAGC GCGGTGCTCTGCTGCCGCATGGCGCCCTTGCAGAAGGCTCAGATTGTTAAATTAATCA AATTTTCAAAGAGCACCCAATCACGTTAGCAATTGGCGATGGTGCAAATGATGTCAG CATGATTCTGGAAGCGCACGTGGGCATAGGTGTCATCGGCAAGGAAGGCCCGCCAGGCT GCCAGGAACAGCGACTATGCAATCCCAAAGTTTAAAGCATTGGAAGAAGATGCTGCTTG TTCACGGGCATTTTTATTACATTAGGATCTCTGAGCTCGTGCAGTACTTCTTCTATAA GAACGTCTGCTTCATCTTCCCTCAGTTTTTATACCAGTTCTTCTGTGGGTTTTTCACAA CAGACTTTGTACGACACCGCGTATCTGACCCTCTACAACATCAGCTTCACCTCCCTCC CCATCCTCCTGTACAGCCTCATGGAGCAGCATGTTGGCATTGACGTGCTCAAGAGAGA CCCGACCCTGTACAGGGACGTCGCCAAGAATGCCCTGCTGCGCTGGCGCGTGTTCATC TACTGGACGCTCCTGGGACTGTTTGACGCACTGGTGTCTTCTTTGGTGCTTATTTTCG TGTTTGAAAATACAACCTGTGACAAGCAACGGGCAGATATTTGGAAACTGGACGTTTGG AACGCTGGTATTCACCGTGATGGTGTTCACAGTTACACTAAAGCTTGCATTGGACACA CACTACTGGACTTGGATCAACCATTTTGTATCTGGGGGTCGCTGCTGTTCTACGTTG TCTTTTCGCTTCTCTGGGGAGGAGTGATCTGGCCGTTCTCAACTACCAGAGGATGTA CTACGTGTTTCATCCAGATGCTGTCCAGCGGGCCCGCTGGCTGGCCATCGTGCTGCTG GTGACCATCAGCCTCCTTCCCGACGTCCTCAAGAAAGTCCTGTGCCGGCAGCTGTGGC CAACAGCAACAGAGAGAGTCCAGAATGGGTGCGCACAGCCTCGGGACCGCGACTCAGA ATTCACCCCTCTTGCCCTCTCTGCAGAGCCAGGCTACCAGAGCACCTGTCCCTCGGCC GCCTGGTACAGCTCCCACTCTCAGCAGGTGACACTCGCGGCCTGGAAGGAGAAGGTGT CCACGGAGCCCCCACCATCCTCGGCGGTTCCCATCACCCTGCAGTTCCATCCCAAG TCACAGCTGCCCTAGGTCCCGTGTGGGAATGCTCGTGTGATGGATGGTCTTAAGCCTG TGGAGACTGTGCACGTGCCTCTTCTGCCCCCAGCAGGCAAGGAGGGGGGTACAGG CCTTGCCCTCGAGCATGGCACCTGGCCGCTGGACCCAGCACTGTGGT		
	ORF Start: ATG at 61	ORF Stop: TGA at 3634	
	SEQ ID NO: 52	1191 aa	MW at 135846.0kD
NOV19a, CG104934-01 Protein Sequence	MDCSLVRTLVHRYCAGEENWVDSRTIYVGHREPPPGAEAYIPQRYPDNRIVSSKYTFW NFIPKNLFEQFRRVANFYFLIIIFLVQLIIDTPTSPVTSGLPLFFVITVTAIKQGYEDW LRHKADNAMNQCPVHFIQHGKLVKQSRKLRVGDIVMVKEDETFPCDLIFLSSNRGDG TCHVTTASLDGESSHKTHYAVQDTKGFHTEEDIGGLHATIECEQPQPDLYKFVGRINV YSDLNDPVVRPLGSENLLLRGATLKNTEKIFGVAIYTMETKMALNYQSKSQKRSAGE KSMNAFLIVYLCILISKALINTVLKYMWQSEPFRDEPWYNQKTESERQRNLFKAFTD FLAFMVLFNYYIIPVSMYVTVMQKFLGSYFITWDEDMFDEETGEGPLVNTSDLNEELG QVEYIFTDKTGTLTENNMEFKECCIEGHVYVPHVICNGQVLPESGIDMIDSSPSVNG REREELFFRALCLCHTVQVKDDDSVDGPRKSPDGGKSCVYISSSPDEVALVEGVQRLG FTYLRLKDNMEILNRENHIERFELLEILSFDSVRRRMSVIVKSATGEIYLFCKGADS		

	SIFPRVIEGKVDQIRARVERNAVEGLRTLCAVAYKRLIQEEYEGICKLLQAAKVALQDR EKKLAEAYEQIEKDLTLLGATAVEDRLQEKAAADTIEALQKAGIKVWVLTGDKMETAAA TCYACKLFRNTQLLELTTKRIEEQSLHDVLFELSKTVLRHSGSLTRDNLSGLSADMQ DYGLIIDGAALSLIMKPREDDGSSGNRELFLFELCRSCSAVLCCRMALPLQKAQIVKLIK FSKEHPITLAIGDGANDVSMILEAHVGIGVIGKEGRQAARNSDYAIPKFKHLKKMLLV HGHFYIIRISELVQYFFYKNCVFIFFPQFLYQFFCGFSQQTLYDTAYLTLYNISFTSLP ILLYSLMEQHVIGIDVLKRDPTLYRDVAKNALLRWRVFIYWTLGLFDALVFFFGAYFV FENTTVTSNGQIFGNWTFGLTVFTVMVFTVTLKLALDTHYWTWINHFVIWGSLLFYVV FSLWGGVIWPFLLNYQRMYYVFIQMLSSGPAWLAIVLLVTISLLPDVLKKVLCRQLWP TATERVQNGCAQPRDRDSEFTPLASLQSPGYQSTCPSAAWYSSHSQQVTLAAWKEKVS TEPPPILGSGHHHCSSIPSHSCPRSRVGMLV
--	---

Further analysis of the NOV19a protein yielded the following properties shown in Table 19B.

Table 19B. Protein Sequence Properties NOV19a	
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:	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 publication, yielded several homologous proteins shown in Table 19C.

5

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
AAO14200	Human transporter and ion channel TRICH-17 - Homo sapiens, 1192 aa. [WO200204520-A2, 17-JAN-2002]	1..1191 1..1192	1190/1192 (99%) 1191/1192 (99%)	0.0
AAB42368	Human ORFX ORF2132 polypeptide sequence SEQ ID NO:4264 - Homo sapiens, 797 aa. [WO200058473-A2, 05-OCT-2000]	338..1109 1..772	770/772 (99%) 772/772 (99%)	0.0
AAG67546	Amino acid sequence of a human transporter protein - Homo sapiens, 1177 aa. [WO200164878-A2, 07-SEP-2001]	22..1109 18..1106	657/1119 (58%) 833/1119 (73%)	0.0
AAO14203	Human transporter and ion channel TRICH-20 - Homo sapiens, 1096 aa. [WO200204520-A2, 17-JAN-2002]	22..1109 18..1040	583/1119 (52%) 755/1119 (67%)	0.0
AAM39290	Human polypeptide SEQ ID NO 2435 - Homo sapiens, 815 aa. [WO200153312-A1, 26-JUL-2001]	370..1109 1..744	424/771 (54%) 544/771 (69%)	0.0

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
P98197	Potential phospholipid-transporting ATPase IH (EC 3.6.3.1) - Mus musculus (Mouse), 1187 aa.	1..1189 1..1185	1074/1195 (89%) 1117/1195 (92%)	0.0
P98196	Potential phospholipid-transporting ATPase IS (EC 3.6.3.1) - Homo sapiens (Human), 797 aa (fragment).	338..1109 1..772	772/772 (100%) 772/772 (100%)	0.0
Q8WX24	BB206I21.1 (ATPASE, CLASS VI, TYPE 11C) - Homo sapiens (Human), 962 aa (fragment).	14..997 1..962	633/992 (63%) 770/992 (76%)	0.0
Q9N0Z4	RING-FINGER BINDING PROTEIN - Oryctolagus cuniculus (Rabbit), 1107 aa (fragment).	22..1109 10..1036	574/1123 (51%) 752/1123 (66%)	0.0
Q9Y2G3	Potential phospholipid-transporting ATPase IR (EC 3.6.3.1) - Homo sapiens (Human), 672 aa (fragment).	486..1109 1..601	358/625 (57%) 462/625 (73%)	0.0

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
Hydrolase	408..846	46/448 (10%) 258/448 (58%)	0.0058

5

Example 20.

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

Table 20A. NOV20 Sequence Analysis			
	SEQ ID NO: 53	2588 bp	
NOV20a, CG105463-01 DNA Sequence	AGTTC CGAACAGAAGGCTGTGTATTCTCTGCCGCTTATTGTGGCCTCGACAGGCCATG GTTACTTTGGCCACTGCCAGAGCAGCCTTGGCACTATGGAGGAGCCTAGGGCTACCCC TCAGCTGTACTTGGGGCTGGTCCTGCAGTTGCTACCCAGGGTTATGGCAGCACTGCCT GAAGGTGTGAGACCAAATTCGAATCCTTATGGTTTTTCCATGGGAATTGGTGATATGTG CAGCTGTCCTTGGATTTGTTGCTGTTCCCTTTTTTTTTGTGGAGAAGTTTTAGATCGGT TAGGAGTCGGCTTTATGTGGGAAGAGAGAAAGAGCTTGCTATAGCGCTTTCTGGACTA ATTGAAGAAAAATGTAGACTACTTGAAAAATTTAGCCTTGTTCAAAAAGAGTATGAAG GCTATGAAGTAGAGTCATCTTTAGAGGATGCCAGCTTTGAGAAGGAGGCAACAGAAGC ACAAAGTCTGGAGGCAAACTGTGAAAAGCTGAACAGGTCCAATTCTGAACTGGAGCAT GAAATACTCTGTCTAGAAAAGGGGATAAAAAGAAGAGAAATCTAAACATTCTGAACAAG ATGAGGTGATGGCAGATATTTCCAAAAAGATACAGTCTCTAGAAGATGAGTCAAAATC CCTCAAATCACTACTAACTGAAGCCAAAATGACCTTCAAGGGATTTCAAATGAATGAA GAAAACTGGAGATAGGAATACAAGATGCTTCGAGTGAAAATTGTCAACTTCAGGAAA GCCAGAAACAGCTTTTGGCAAGAAGCTGAAGTATGGAAAGAACAAGTGAGTGAACCTTAA TAAACAGAAAATAACATTTGAAGACTCCAAAGTACACGCAGAACAAAGTTCTAAATGAT AAAGAAAATCACATCGAGACTCTGACTGAACGCTTGCTAAAGATCAAAGATCAGGCTG CTGTGCTGGAAGAAGACATAACGGATGATGGTAACTTGGAAATTAGAAATGAACAGTGA ATTGAAAGATGGTGCTTACTTAGATAATCCTCCAAAAGGAGCTTTGAAGAACTGATT CATGCTGCTAAGTTAAATGCTTCTTTAACAACCTTAGAAGGAGAAAGAAACCAATTTA TATTCAGTTATCTGAAGTTGATAAAACCAAGGAAGAGCTTAGAGAGCATATTAATAAT CTTCAGACGGAACAAGCATCTTTGCAGTCGGAAAACACACATTTTGAAAGTGAGAATC AGAACTTCAACAGAAAGTTAATGACTGAGTTATATCAAGAAAATGAAATGAACTCT ACAGGAAATTAATAGTAGAGGAAAATAACCGGTTAGAGAAAGAGAACTTTCTAAAGT AGACGAAATGATCAGCCATGCCACTGAAGAGCTGGAGACCTGCAGAAAGCGAGCCAAA GATCTTGAAGAAGAAGCTTGAGAGAACTATTCTTTTTTATCAAGGGAAGATTATATACC ATGAGAAAAAAGCACATGATAATTGTTTGGCAGCATGGACTGCTGAAAGAAACCTCAA TGATTTAAGGAAAGAAAATGCTCACAAAAGACAAAAATTAGCTGAAACAGAGTTTAAA ATTAACTTTTAGAAAAAGATCCTTATGCACCTTGATGTTCCAAATACAGCATTTGGCA GAGAGCATTCCTCATATGGTCCCTCACCATTGGGTTCGGCCTTCATCTGAAACGAGAGC TTTTCTCTATCTTCCGACTTTGTTGGAGGGTCCACTGAGACTCTCACCTTTGCTTCCA GGGGGAGGAGGAAGAGACCCAAGAGGGCCAGGGAATCCTCTGGACCACCAGATTACCA AGGAAAGAGGAGAATCAAGCTGTGATAGGTTTACTGATCCTCACAAGGCTCCTTCTGA CACTGGGCCCCCTGTCACCTCCGTGGGAACAGGACCGTAGGATGATGTTTCTCCACCA GGACAATCATATCCTGATTGAGCTCTTCTCCACAAAGGCAAGACAGATTTTATTCTA ATTCTGCTAGACGCTCTGGACTAGCAGAACTCAGAAGTTTTAATATACCTTCTTTGGA TAAAATGGATGGGTCAATGCCTTCAGAAATGGAATCCAGTGGAATGATACCAAAGAT AATCTTGGTAATTTAAATGTGGCTGATTCATCTCTCCCTGCTGGAAATGAAGTGAGTG GCCCTGGCTTTGTTCTCCACCTCTTGCTTCAATCAGAGGTCCATTGTTTCCAGTGGA TACGAGGGGCCCCGTTTCATGAGAAGAGGACCTCCTTTCCCTCCACCTCCTCCAGGAACC ATGTTTGGAGCTTCTCCAGATTATTTTCCACCAAGGGATGTCCCAGGTCCACCACGTG CTCCATTTGCAATGAGAAATGTCTGTCCACCGAGGGGTTTTCTCCTTACCTTCCCCC AAGACCTGGATTTTGCCCCCACCCTCCACCTCCACCTCCAGTGAAGTTCCCTTTAGGGTTGAGT CTGCCTTCAAATGAGCCTGCTGCTGAAGATCCAGAACCACGGCAAGAAACCTGATAAT ATTTTTGCTGTCTTCAAAAGTCATTTTGACTATTCTCATTTTCAGTTGAAGTAACTGC TGTTACTTCAGTGATTACACTTTTGCTCAAATTGAA		
	ORF Start: ATG at 94	ORF Stop: TGA at 2488	
	SEQ ID NO: 54	798 aa	MW at 90383.6kD
NOV20a, CG105463-01 Protein Sequence	MEEPRATPQLYLGLVLQLLPRVMAALPEGVRPNSNPYGFPELVICA AVLGFVAVPFF LWRSFRSVRSRLVVGREKELAIALSGLIEEKRLLEKFSLVQKEYEGYEVESSLEDAS FEKEATEAQSLANCEKLNRSNSELEHEILCLEKGIKEEKSXHSEQDEVMA DISKKIQ SLEDESKSLKSLLTEAKMTFKGFQMNEEKLEIGIQDASSENQQLQESQKQLLQEA EVW KEQVSELNKQKITFEDSKVHAEQVLNDKENHIETLTERLLKIKDQAAVLEEDITDDGN LELEMNSELKDGAYLDNPPKGALKKLIHAAKLNASLTTLEGERNQFIFSYLKLKIPRK		

	SLESILKIFRRNKHLC SRKTHILKVRIRNFNRKLMTELYQENEMKLYRKLIVEENNRL EKEKLSKVDEMISHATEELETCTRKRAKDLEEELEERTILFYQGKIIYHEKKAHDNCLAA WTAERNLNDLRKENAHKRQKLAETEFKIKLLEKDPYALDVPNTAFGREHSSYGPSPLG RPSSETRAFLYLPTLLEGLRLSPLLPGGGGRDPRGPGNPLDHQITKERGESSCDRFT DPHKAPSDTGPLSPWEQDRRMFPPPGQSYPDALPPQRQDRFYSNSARRSGLAELR SFNIPSLDKMDGSMPSSESSGNDTKDNLGNLNVADSSLPAGNEVSGPGFVPPPLASI RGPLFPVDTRGPFMRRGPPFPPPPPGTMFGASPDYFPPRDVPGPPRAPFAMRNVCPPR GFPPYLPPRPGFCPHPHPHSEFPLGLSLPSNEPAAEDPEPRQET		
	SEQ ID NO: 55	2483 bp	
NOV20b, CG105463-02 DNA Sequence	AGCTGGAATTCGCCCTTCTCGACAGGCCATGGTTACTTTGGCCACTGCCAGAGCAGCC TTGGCACTATGGAGGAGCCTAGGGCTACCCCTCAGCTGTACTTGGGGCTGGTCCTGCA GTTGCTACCCAGGGTTATGGCAGCACTGCCTGAAGGTGTGAGACCAAATTCGAATCCT TATGGTTTTCCATGGGAATTGGTGATATGTGCAGCTGTCCTTGGATTTGTTGCTGTTT CCTTTTTTTTGTGGAGAAGTTTTAGATCGGTTAGGAGTCGGCTTTATGTGGGAAGAGA GAAAGAGCTTGCTATAGCGCTTTCTGGACTAATTGAAGAAAAATGTAGACTACTTGAA AAATTTAGCCTTGTTCAAAAAGAGTATGAAGGCTATGAAGTAGAGTCATCTTTAGAGG ATGCCAGCTTTGAGAAGGAGGCAACAGAAGCACAAAGTCTGGAGGCAAACTGTGAAAA GCTGAACAGGTCCAATTCTGAACCTGGAGCATGAAATACTCTGTCTAGAAAAGGGGATA AAAGAAGAGAAATCTAAACATTCTGAACAAGATGAGGTGATGGCAGATATTTCCAAAA AGATACAGTCTCTAGAAGATGAGTCAAAATCCCTCAAATCACTACTAAGCTAA AATGACCTTCAAGGGATTTCAAATGAATGAAGAAAACTGGAGATAGGAATACAAGAT GCTTCGAGTGAAAATTGTCAACTTCAGGAAAGCCAGAAACAGCTTTTGCAAGAAGCTG AAGTATGGAAAGAACAAGTGAGTGAACCTTAATAAACAGAAAATAACATTTGAAGACTC CAAAGTACACGCAGAACAAAGTTCTAAATGATAAAGAAAATCACATCGAGACTCTGACT GAACGCTTGCTAAAGATCAAAGATCAGGCTGCTGTGCTGGAAGAAGACATAACGGATG ATGGTAACTTGGAATTAGAAATGAACAGTGAATTGAAAGATGGTGCTTACTTAGATAA TCCTCCAAAAGGAGCTTTGAAGAACTGATTCATGCTGCTAAGTTAAATGCTTCTTTA ACAACCTTAGAAGGAGAAAGAAACCAATTTATATTCACTTATCTGAAGTTGATAAAAC CAAGGAAGAGCTTAGAGAGCATATTAATAATCTTCAGACGGAACAAGCATCTTTGCAG TCGGAAAACACACATTTTGAAAGTGAGAATCAGAACTTCAACAGAAAGTTAATGACT GAGTTATATCAAGAAAATGAAATGAACTCTACAGGAAATTAATAGTAGAGGAAAATA ACCGGTTAGAGAAAGAGAACTTTCTAAAGTAGACGAAATGATCAGCCATGCCACTGA AGAGCTGGAGACCTGCAGAAAGCGAGCCAAAGATCTTGAAGAAGAACTTGAGAGAACT ATTCTTTTTTATCAAGGGAAGATTATATACCATGAGAAAAAGCACATGATAATTGTT TGGCAGCATGGACTGCTGAAAGAAACCTCAATGATTTAAGGAAAGAAAATGCTCACAA AAGACAAAATTAGCTGAAACAGAGTTTAAATTAACCTTTTAGAAAAAGATCCTTAT GCACTTGATGTTCCAAATACAGCATTGTCAGAGAGCATTCTCATATGGTCCCTCAC CATTGGGTGCGCCTTCATCTGAAACGAGAGCTTTTCTCTATCTTCCGACTTTGTTGGA GGGTCCACTGAGACTCTCACCTTTGCTTCCAGGGGGAGGAGGAAGAGGCCCAAGAGGC CCAGGGAATCCTCTGGACCACCAGATTACCAAGGAAAGAGGAGAATCAAGCTGTGATA GGTTTACTGATCCTCACAAGGCTCCTTCTGACACTGGGCCCCTGTCACCTCCGTGGGA ACAGGACCGTAGGATGATGTTTCTCCACCAGGACAATCATATCCTGATTCAGCTCTT CCTCCACAAAGGCAAGACAGATTTTATTCTAATTCTGCTAGACGCTCTGGACTAGCAG AACTCAGAAGTTTTAATATACCTTCTTTGGATAAAATGGATGGGTCAATGCCTTCAGA AATGGAATCCAGTGGAATGATACCAAGATAATCTTGGTAATTTAAATGTGGCTGAT TCATCTCTCCCTGCTGGAAATGAAGTGAGTGGCCCTGGCTTTGTTCTCCACCTCTTG CTCCAATCAGAGGTCCGTTGTTTCCAGTGATACGAGGGGGCCGTTTCATGAGAAGAGG ACCTCCTTTCCCTCCACCTCCTCCAGGAACCATGTTTGGAGCTTCTCCAGATTATTTT CCACCAAGGGATGTCCAGGTCTACCACGTGCTCCATTTGCAATGAGAAATGTCTGTC CACCGAGGGGTTTTCTCCTTACCTTCCCCCAAGACCTGGATTTTGCCCCCACCCTCA CCCCACATTCTGAAGATAGAGTGAGTTCCCTTTAGGGTTGAGTGCCTTCAAATGAGC CTGCTGCTGAAGATCCAGAACCACGGCAAGAAACCTGATAATATTTT		
	ORF Start: ATG at 67	ORF Stop: TGA at 2401	
	SEQ ID NO: 56	778 aa	MW at 88255.5kD
NOV20b, CG105463-02 Protein Sequence	MEEPRA TPQLYLGLVLQLLPRVMAALPEGVRPNSNPYGF PWELVICAAVLGFVAVPFF LWRSFRSVRSRLYVGREKELAIALSGLIEEKCRLLLEKFSLVQKEYEGYEVESSLEDAS FEKEATEAQSLANCEKLNRSNSELEHEILCLEKGIKEEKSKHSEQDEVMDISKKIQ SLEDESKSLKSLLTEAKMTFKGFQMNEEKLEIGIQDASSENQQLQESQKQLLQEAQEVW		

KEQVSELNKQKITFEDSKVHAEQVLNDKENHIETLTERLLKIKDQAAVLEEDITDDGN LELEMNSELKDGAYLDNPPKGALKKLIHAAKLNASLTTEGERNQFIFSULKLIKPRK SLESILKIFRRNKHLC SRKTHILKVRIRNFNRKLMTELYQENEMKLYRKLI VEENNRL EKEKLSKVDEMISHATEELETCTRKRAKDLEEEELERTILFYQGKI IYHEKKAHDNCLAA WTAERNLNDLRKENAHKRQKLAETEFKIKLLEKDPYALDVPNTAFGREHSSYGPSPLG RPSSETRAFLYLPTLLEGPLRLSPLLPGGGGRGPRGPGNPLDHQITKERGESSCDRFT DPHKAPSDTGPLSPPWEQDRMMFPFPGQSYFDSALPPQRQDRFYNSARRSGLAELR SFNIPSLDKMDGSMPSMESSGNDTKDNLGNLNVADSSLPAGNEVSGPGFVPPPLAPI RGPLFPVDTRGPFMRRGPFPPPPPGTMFGASPDYFPPRDVPGLPRAPFAMRNVCPPR GFPPYLPFPPRGFCPPHPHILKIE
--

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 20B.

Table 20B. Comparison of NOV20a against NOV20b.		
Protein Sequence	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV20b	1..750	662/750 (88%)
	1..750	662/750 (88%)

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

Table 20C. Protein Sequence Properties NOV20a	
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:	Cleavage site between residues 25 and 26

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

Table 20D. Geneseq Results for NOV20a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY77574	Human cytoskeletal protein (HCYT) (clone 3768043) - Homo sapiens, 806 aa. [WO200006730-A2, 10-FEB-2000]	1..798 1..806	638/812 (78%) 683/812 (83%)	0.0
ABG05280	Novel human diagnostic protein #5271 - Homo sapiens, 881 aa. [WO200175067-A2, 11-OCT-2001]	1..797 59..867	639/814 (78%) 684/814 (83%)	0.0
ABG05280	Novel human diagnostic protein #5271 - Homo sapiens, 881 aa. [WO200175067-A2, 11-OCT-2001]	1..797 59..867	639/814 (78%) 684/814 (83%)	0.0
ABG20258	Novel human diagnostic protein #20249 - Homo sapiens, 881 aa. [WO200175067-A2, 11-OCT-2001]	1..797 59..867	634/814 (77%) 681/814 (82%)	0.0
ABG20258	Novel human diagnostic protein #20249 - Homo sapiens, 881 aa. [WO200175067-A2, 11-OCT-2001]	1..797 59..867	634/814 (77%) 681/814 (82%)	0.0

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

Table 20E. Public BLASTP Results for NOV20a				
Protein Accession Number	Protein/Organism/Length	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O15320	Meningioma-expressed antigen 6/11 (MEA6) (MEA11) - Homo sapiens (Human), 804 aa.	1..798 1..804	653/810 (80%) 696/810 (85%)	0.0
Q96SG9	BA500G10.2 (NOVEL PROTEIN SIMILAR TO MENINGIOMA EXPRESSED ANTIGEN 6 (MEA6) AND 11 (MEA11)) - Homo sapiens (Human), 825 aa (fragment).	1..798 15..816	616/805 (76%) 670/805 (82%)	0.0
Q96RT6	CTAGE-2 - Homo sapiens (Human), 754 aa.	30..775 1..746	605/749 (80%) 642/749 (84%)	0.0
O95046	WUGSC:H_DJ0988G15.3 PROTEIN (DJ1005H11.2) (WUGSC:H_DJ0988G15.3 PROTEIN) - Homo sapiens (Human), 777 aa.	1..770 1..775	570/775 (73%) 633/775 (81%)	0.0
AAH26864	SIMILAR TO MENINGIOMA EXPRESSED ANTIGEN 6 (COILED-COIL PROLINE-RICH) - Mus musculus (Mouse), 779 aa.	30..796 1..778	520/783 (66%) 600/783 (76%)	0.0

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

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

5

Example 21.

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

Table 21A. NOV21 Sequence Analysis			
	SEQ ID NO: 57	1269 bp	
NOV21a, CG105491-01 DNA Sequence	<u>AGCGGGGGGCGCTGCCTCGAGCCTCATGGCTGCCCCCTGCTTCCGTCATGGGCCCCACTC</u> GGGCCCTCTGCCCTGGGCCTTCTGCTGCTGCTCCTGGTGGTGGCCCCCTCCCCGGGTCCG CAGCATTGGTCCACAGACAGCCAGAGAACCAGGGAATCTCCCTAACTGGCAGCGTGGC CTGTGGTTCGGCCCAGCATGGAGGGGAAAATCCTGGGCGGCGTCCCTGCGCCCGAGAGG AAGTGGCCCGTGGCAGGTCAGCGTGCACCTACGCAGGCCTCCACGTCTGCGGCGGCTCCA TCCTCAATGAGTACTGGGTGCTGTGAGCTGCGCACTGCTTTTCACAGGGACAAGAATAT CAAAATCTATGACATGTACGTAGGCCTCGTAAACCTCAGGGTGGCCGGCAACCACACC CAGTGGTATGAGGTGAACAGGGTGATCCTGCACCCACATATGAGATGTACCACCCCA TCGGAGGTGACGTGGCCCTGGTGCAGCTGAAGACCCGCATTGTGTTTTCTGAGTCCGT GCTCCCGGTTTGCCTTGCAACTCCAGAAGTGAACCTTACCAGTGCCAATTGCTGGGCT ACGGGATGGGGACTAGTCTCAAAACAAGGTGAGACCTCAGACGAGCTGCAGGAGGTGC AGCTCCCGCTGATCCTGGAGCCCTGGTGCCACCTGCTCTACGGACACATGTCTTACAT CATGCCCCGACATGCTGTGTGCTGGGGACATCCTGAATGCTAAGACCGTGTGTGAGGGC GACTCCGGGGGGCCCACTTGTCTGTGAATTCAACCGCAGCTGGTTGCAGATTGGAATTG TGAGCTGGGGCCGAGGCTGCTCCAACCTCTGTACCCTGGAGTGTATGCCAGTGTTC CTATTTCTCAAAATGGATATGTGATAACATAGAAATCACGCCCACTCCTGCTCAGCCA GCCCTGCTCTCTCTCCAGCTCTGGGGCCCACTCTCAGCGTCTTAATGGCCATGCTGG CTGGCTGGTCAGTGTGTGAGGTCAGGATACCCACTCTAGGATTCTCATGGCTGCACA CCCTGCCCCAGCCCAGCTGCCTCCAGACCCCTAAGCATCTCCTGTCCTGGCCTCTCTG AAGCAGACAAGGGCCACCTATCCCGGGGGTGGATGCTGAGTCCAGGAGGTGATGAGCA AGTGTACAAAAGAAAAAGGGAAGGGGGAGAGGGGCTGGTCAGGGAGAACCAGCTTG GGCAGAGTGCACCTGAGATTTGATAAGATCATTAAATATTTACAAAGCAA		
	ORF Start: ATG at 26	ORF Stop: TGA at 1004	
	SEQ ID NO: 58	326 aa	MW at 35323.8kD
NOV21a, CG105491-01 Protein Sequence	MAAPASVMGPLGPSALGLLLLLLVVAPPRVAALVHRQPENQGISLTGSVACGRPSMEG KILGGVPAPERKWPWQVSVHYAGLVCGGSILNEYWVLSAAHCFHRDKNIKIYDMYVG LVNLRVAGNHTQWYEVNRVILHPTYEMYHPIGGDVALVQLKTRIVFSESVLPVCLATE EVNLTSANCWATGWGLVSKQGETSDELQEVQLPLILEPWCHLLYGHMSYIMPDMLCAG DILNAKTVCEGDSGGPLVCEFNRSWLQIGIVSWGRGCSNPLYPGVYASVSYFSKWICD NIEITPTPAQPAPALSPALGPTLSVLMAMLAGWSVL		

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

Table 21B. Protein Sequence Properties NOV21a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane); 0.1007 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 33 and 34

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

Table 21C. Geneseq Results for NOV21a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU82747	Amino acid sequence of novel human protease #46 - Homo sapiens, 326 aa. [WO200200860-A2, 03-JAN-2002]	1..326 1..326	326/326 (100%) 326/326 (100%)	0.0
AAB73945	Human protease T - Homo sapiens, 290 aa. [WO200116293-A2, 08-MAR-2001]	13..288 4..272	115/286 (40%) 152/286 (52%)	6e-53
AAE03821	Human gene 4 encoded secreted protein HWHIH10, SEQ ID NO: 67 - Homo sapiens, 290 aa. [WO200136440-A1, 25-MAY-2001]	13..288 4..272	115/286 (40%) 152/286 (52%)	6e-53
AAU12282	Human PRO4327 polypeptide sequence - Homo sapiens, 290 aa. [WO200140466-A2, 07-JUN-2001]	13..288 4..272	115/286 (40%) 152/286 (52%)	6e-53
AAY73388	HTRM clone 3376404 protein sequence - Homo sapiens, 290 aa. [WO9957144-A2, 11-NOV-1999]	13..288 4..272	115/286 (40%) 152/286 (52%)	6e-53

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 21D. .

Table 21D. Public BLASTP Results for NOV21a				
Protein Accession Number	Protein/Organism/Length	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BQR3	Marapsin precursor (EC 3.4.21.-) - Homo sapiens (Human), 290 aa.	13..288 4..272	115/286 (40%) 152/286 (52%)	1e-52
Q9PVX7	EPIDERMIS SPECIFIC SERINE PROTEASE - Xenopus laevis (African clawed frog), 389 aa.	50..314 16..287	98/275 (35%) 159/275 (57%)	5e-52
AAH24903	RIKEN CDNA 2010001P08 GENE - Mus musculus (Mouse), 331 aa.	51..326 45..329	114/288 (39%) 158/288 (54%)	2e-51
Q91XC4	SIMILAR TO DISTAL INTESTINAL SERINE PROTEASE - Mus musculus (Mouse), 310 aa.	51..288 28..272	106/247 (42%) 138/247 (54%)	6e-51
Q9QYZ9	DISTAL INTESTINAL SERINE PROTEASE - Mus musculus (Mouse), 310 aa.	51..288 28..272	105/247 (42%) 137/247 (54%)	7e-50

Pfam analysis predicts that the NOV21a protein contains the domains shown in the Table 21E.

Table 21E. Domain Analysis of NOV21a			
Pfam Domain	NOV21a Match Region	Identities/ Similarities for the Matched Region	Expect Value
trypsin	60..288	87/265 (33%) 172/265 (65%)	5.3e-72

5

Example 22.

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

Table 22A. NOV22 Sequence Analysis		
	SEQ ID NO: 59	4131 bp
NOV22a, CG105954-01	GTGCCGAGGATGGCCAGGCAGCCACCGCCGCCCTGGATCCATGCAGCCTTCCTCCTCT GCCTCCTCAGTCTTGCGGAGCCATCGAAATTCCTATGGTTCCAAGCATTTCAGAATGA GCTGACGCAGCCGCCAACCATCACCAAGCAGTCAGCGAAGGATCACATCGTGGACCCC	

DNA Sequence	CGTGATAACATCCTGATTGAGTGTGAAGCAAAAGGGGAACCCTGCCCCAGCTTCCACT GGACACGAAACAGCAGATTCTTCAACATCGCCAAGGACCCCCGGGTGTCCATGAGGAG GAGGTCTGGGACCCTGGTGATTGACTTCCGCAGTGGCGGGCGGCCGGAGGAATATGAG GGGGAATATCAGTGCTTCGCCCCGAACAAATTTGGCACGGCCCTGTCCAATAGGATCC GCCTGCAGGTGTCTAAATCTCCTCTGTGGCCCAAGGAAAACCTAGACCCTGTCGTGGT CCAAGAGGGCGCTCCTTTGACGCTCCAGTGCAACCCCCCGCCTGGACTTCCATCCCCG GTCATCTTCTGGATGAGCAGCGCCATGGAGCCCATCACCCAAGACAAACGTGTCTCTC AGGGCCATAACGGAGACCTATACTTCTCCAACGTGATGCTGCAGGACATGCAGACCGA CTACAGTTGTAACGCCCCGCTTCCACTTACCCACACCATCCAGCAGAAGAACCCTTTC ACCCTCAAGGTCTCACCAGTAAGCCTTATAATGACTCGTCCTTAAGAAACCACCCTG ACATGTACAGTGCCCGAGGAGTTGCAGAAAGAACACCAAGCTTCATGTATCCCCAGGG CACCGCGAGCAGCCAGATGGTGCTTCGTGGCATGGACCTCCTGCTGGAATGCATCGCC TCCGGGGTCCCAACACCAGACATCGCATGGTACAAGAAAGGTGGGGACCTCCCATCTG ATAAGGCCAAGTTTGAGAACTTTAATAAGGCCCTGCGTATCACAAATGTCTCTGAGGA AGACTCCGGGGAGTATTTCTGCCTGGCCTCCAACAAGATGGGCAGCATCCGGCACACG ATCTCGGTGAGAGTAAAGGCTGCTCCCTACTGGCTGGACGAACCCAAGAACCTTATTC TGGCTCCTGGCGAGGATGGGAGACTGGTGTGTGCGAGCCAATGGAAACCCCAAACCCAC TGTCCAGTGGATGGTGAATGGGGAACCTTTGCAAGCGGCACCACCTAACCCAAACCGT GAGGTGGCCGGAGACACCATCATCTTCCGGGACACCCAGATCAGCAGCAGGGCTGTGT ACCAGTGCAACACCTCCAACGAGCATGGCTACCTGCTGGCCAACGCCTTTGTCAGTGT GCTGGATGTGCCGCTCGGATGCTGTGCCCCCGGAACCAGCTCATTGAGTGATTCTT TACAACCGGACGCGGCTGGACTGCCCTTTCTTTGGGTCTCCCATCCCCACACTGCGAT GGTTTAAGAATGGGCAAGGAAGCAACCTGGATGGTGGCAACTACCATGTTTATGAGAA CGGCAGTCTGGAAATTAAGATGATCCGCAAAGAGGACCAGGGCATCTACACCTGTGTC GCCACCAACATCCTGGGCAAAGCTGAAAACCAAGTCCGCCTGGAGGTAAAAGACCCCA CCAGGATCTACCGGATGCCCGAGGACCAGGTGGCCAGAAGGGGCACCACGGTGCAACT GGAGTGTGCGGTGAAGCACGACCCCTCCCTGAAACTCACCGTCTACTGGCTGAAGGAT GACGAGCCGCTCTATATTGGAAACAGGATGAAGAAGGAAGACGACTCCCTGACCATCT TTGGGGTGGCAGAGCGGGACCAGGGCAGTTACACGTGTGTGCGCCAGCACCGAGCTAGA CCAAGACCTGGCCAAGGCCTACCTCACCGTGTAGGACGGCCAGACCGGCCCGGGAC CTGGAGCTGACCGACCTGGCCGAGAGGAGCGTGCGGCTGACCTGGATCCCCGGGGATG CTAACAACAGCCCCATCACAGACTACGTGCTCCAGTTTGAAGAAGACCAGTTCCAACC TGGGGTCTGGCATGACCATTTCCAAGTACCCCGGCAGCGTTAACTCAGCCGTCTTCCGG CTGTCCCCGTATGTCAACTACCAGTTCCGTGTGTCATTGCCATCAACGAGGTGGGAGCA GCCACCCAGCCTCCCATCCGAGCGCTACCGAACCAGTGGAGCACCCCCCGAGTCCAA TCCTGGTGACGTGAAGGGAGAGGGGACCAGAAAGAACAACATGGAGATCACGTGGACG CCCATGAATGCCACCTCGGCCTTTGGCCCCAACCTGCGCTACATTGTCAAGTGGAGGC GGAGAGAGACTCGAGAGGCCTGGAACAACGTACAGTGTGGGGCTCTCGCTACGTGGT GGGGCAGACCCAGTCTACGTGCCCTATGAGATCCGAGTCCAGGCTGAAAATGACTTC GGGAAGGGCCCTGAGCCAGAGTCCGTGTCATCGGTTACTCCGGAGAAGATTATCCCAGGG CTGCGCCCACTGAAGTTAAAGTCCGAGTCATGAACAGCACAGCCATCAGCCTTCAGTG GAACCGCGTCTACTCCGACACGGTCCAGGGCCAGCTCAGAGAGTACCGAGCCTACTAC TGGAGGGAGAGCAGCTTGCTGAAGAACCTGTGGGTGTCTCAGAAGAGACAGCAAGCCA GCTTCCCTGGTGACCGCCTCCGTGGCGTGGTGTCCCGCCTCTTCCCCTACAGTAACTA CAAGCTGGAGATGGTTGTGGTCAATGGGAGAGGTGATGGGCCTCGCAGTGAGACCAAG GAGTTCACCACCCCGGAAGGAGTACCCAGTGCCCCCTAGGCGTTTCCGAGTCCGGCAGC CCAACCTGGAGACAATCAACCTGGAATGGGATCATCCTGAGCATCCAAATGGGATCAT GATTGGATACACTCTCAAATATGTGGCCTGTACGTTCTCCCCAGTTAACGGGACCAAA GTAGGAAAGCAGATAGTGGAAAACCTTCTCTCCAATCAGACCAAGTTACGGTGCAAA GAACGGACCCCGTGTACGCTACCGCTTTACCCTCAGCGCCAGGACGCAGGTGGGCTC TGGGGAAGCCGTCACAGAGGAGTCACCAGCACCCCGAATGAAGGTAGGTGCATGGCA GCAGCCCCTGGGGTAAAACCCCGACTACCGTGGGTGCGACGGGCGCTGTGAGCAGTA CCGATGCTACTGCCATTGCTGCCACCACCGAAGCCACAACAGTCCCCATCATCCCAAC TGTCGCACCTACCACCATGGCCACCACCACCGTCGCCACAACCTACTACAACCACT GCTGCCGCCACCACCACCGAGAGTCCTCCCACCACCACCTCCGGGACTAAGATAC ACGAATCCGGTACTGCGCATCGCCCATGCTCCCCAGCCCCCTGATGAGCAGTCCATATG GAACGTACCGGTGCTCCCCAACAGTAAATGGGCCAACATCACCTGGAAGCACAATTTTC GGGCCCCGGAAGTGAATTTGTGGTTGAGTACATCGACAGTAACCATAACGAAAAAACTG TCCCAGTTAAGGCCCAGGCTCAGCCTATACAGCTGACAGACCTCTATCCCGGGATGAC ATACACGTTGCGGGTTTATTCCCAGGACAACGAGGGCATCAGCAATCATTCACGGGTT TGCTTCCGGCCCTCCCCGCCAGCTTACACCAACAACCAAGCGGACATCGCCACCCAGG
--------------	---

	GCTGGTTCATTGGGCTTATGTGCGCCATCGCCCTCCTGGTGCTGATCCTGCTCATCGTCTGTTTCATCAAGAGGAGTCGCGGCGGCAAGTACCCAGTACGAGAAAAGAAGGATGTTCCCCTTGGCCCTGAAGACCCCAAGGAAGAGGATGGCTCATTGACTATAGGTCTTTGGCCAGTGATGAGGACAACAAGCCCCTGCAGGGCAGTCAGACATCTCTGGACGGCACCATCAAGCAGCAGGAGAGTGACGACAGCCTGGTGGACTATGGCGAGGGTGGCGAGGGTCAGTTCAATGAAGACGGCTCCTTCATCGGCCAGTACACGGTCAAAAAGGACAAGGAGGAAACAGAGGGCAACGAAAGCTCAGAGGCCACGTCACCTGTCAATGCTATCTACTCTCTGGCCTAACGGAGCCCA		
	ORF Start: ATG at 10	ORF Stop: TAA at 4120	
	SEQ ID NO: 60	1370 aa	MW at 152752.5kD
NOV22a, CG105954-01 Protein Sequence	MARQPPPPWIHAFLCLLSLGGAIIEIPMVPSIQNELTQPPTITKQSAKDHIVDPRDNILIECEAKGNPAPSFHWTRNSRFFNIAKDPRVSMRRSGTLVIDFRSGGRPEEYEGEYQCFARNKFGTALSNRIRLQVSKSPLWPKENLDPVVVQEGAPLTLQCNPPPGLPSPVIFWMSSAMEPITQDKRVSQGHNGDLYFSNVMLQDMQTDYSCNARFHFTHTIQQKNPFTLKVLTSKPYNDSSLRNHPDMYSARGVAERTPSFMYPQGTASSQMVLRGMDLLECIASGVPTPDI AWYKKGDLPSDKAKFENFNKALRITNVSEEDSGEYFCLASNKMGSI RHTISVRVKAAPYWLDEPKNLILAPGEDGRLVCRANGNPKPTVQWMVNGEPLQAAPPNPNREVAGDTIIFRDTQISSRAVYQCNTSNEHGYLLANAFVSVDVPPRMLSPRNQLIRVILYNRTRLDPCFFGSPIPTLRWFKNGQGSNLDGGNYHVYENGSL EIKMIRKEDQGIYTCVATNILGKAENQVRLEV KDPTRIYRMPEDQVARRGTTVQLECRVKHDP SLKLT VYWLKDD EPLYIGNRMKKEDDSLTI FGVAERDQGSYTCVASTELDQDLAKAYLTVLGRPD RPRDLELTDLAERSVRLTWIPGDANNSPITDYVVQFEEDQFQPGVWHDH SKYPG SVNSAVLRLSPYVNYQFRVIAINEVGSSHPSLPSEYRTSGAPPESNPGDVKGEGTRKNMEITWTPMNATS AFGPNLRYIVKWRRETREAWNNTVWGSRYVVGQTPVYVPYEIRVQAENDFGKGPEPESVIGYSGEDYPRAAPTEVKVRVMNSTAISLQWNRVYS DTVQGGQLREYRAYYWRESSLLKNLWVSQKRQQASFPGDRLRGVVSRLFPYSNYKLEMVVVNGRGDGRSETKEFTPEGVPSAPRRFRVRQPNLETINLEWDHPEHPNGIMIGYTLKYVACTFSPVNGTKVGKQIVENFSPNQTKFTVQRTDPVSRYRFTLSARTQVGSGEAVTEES PAPPNEGR CMAAAPGVKPPTTVGATGAVSSTDATAIAATTEATTVP I IPTVAPT TMATTTT VATT TTTTAAATTTTESPPTTTS GTKIHESGTAHRPCSPAPDEQSIWNVTVL P NSKWANITWKHNFGPGTDFVVEYIDSNHTKKTVPVKAQAQPIQLTDLYPGMTYTLRVYSRDNEGISNHSRVCFRPSPPAYTNNQADIATQGWFIGLMCAIALLV LILLIVCFIKRSRGGKYPVREKKDVPLGPEDPKEEDGSFDYRSLASDEDNKPLQGSQTSLDGTIKQQESDDSLVDYGE GEGEGQFNE DGSFIGQYTVKKDK EETEGNESSEATSPVNATYSLA		

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

Table 22B. Protein Sequence Properties NOV22a	
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:	Cleavage site between residues 25 and 26

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

Table 22C. Geneseq Results for NOV22a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM78714	Human protein SEQ ID NO 1376 - Homo sapiens, 937 aa. [WO200157190-A2, 09-AUG- 2001]	92..1050 1..937	928/959 (96%) 933/959 (96%)	0.0
AAM78715	Human protein SEQ ID NO 1377 - Homo sapiens, 952 aa. [WO200157190-A2, 09-AUG- 2001]	92..1050 1..952	928/974 (95%) 933/974 (95%)	0.0
AAW59994	Human neural cell adhesion molecule splice variant NrCAMvar - Homo sapiens, 1304 aa. [WO9836062-A1, 20-AUG-1998]	15..1072 20..1082	517/1075 (48%) 708/1075 (65%)	0.0
AAU10650	Chicken Nr-CAM protein sequence - Gallus sp, 1268 aa. [US6313265- B1, 06-NOV-2001]	12..1054 12..1042	508/1055 (48%) 696/1055 (65%)	0.0
AAB90717	Human CO722_1 protein sequence SEQ ID 130 - Homo sapiens, 1192 aa. [WO200119988-A1, 22-MAR- 2001]	15..1060 20..1047	509/1058 (48%) 701/1058 (66%)	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
O42414	NEUROFASCIN PRECURSOR - Gallus gallus (Chicken), 1369 aa.	13..1370 14..1369	1026/1372 (74%) 1148/1372 (82%)	0.0
Q91Z60	NEUROFASCIN 155 KDA ISOFORM - Rattus norvegicus (Rat), 1174 aa.	1..1042 1..1031	985/1042 (94%) 1008/1042 (96%)	0.0
Q9QVN5	NEUROFASCIN ISOFORM - Rattus sp, 1151 aa.	25..1042 1..1008	965/1019 (94%) 987/1019 (96%)	0.0
Q90924	NEUROFASCIN PRECURSOR - Gallus gallus (Chicken), 1272 aa.	13..1114 14..1120	844/1114 (75%) 949/1114 (84%)	0.0
O94856	KIAA0756 PROTEIN - Homo sapiens (Human), 836 aa (fragment).	193..1050 1..836	830/858 (96%) 833/858 (96%)	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
ig	56..120	11/68 (16%) 43/68 (63%)	0.032
ig	155..215	13/62 (21%) 39/62 (63%)	0.01
ig	278..335	18/61 (30%) 44/61 (72%)	4.3e-11
ig	368..427	13/63 (21%) 44/63 (70%)	2.1e-05
ig	462..520	18/62 (29%) 45/62 (73%)	1.1e-07
ig	553..611	19/60 (32%) 40/60 (67%)	5.9e-09
fn3	630..716	28/88 (32%) 64/88 (73%)	6.6e-14
fn3	729..815	27/92 (29%) 62/92 (67%)	5.9e-07
fn3	827..922	24/97 (25%) 66/97 (68%)	3.7e-07
fn3	934..1026	20/97 (21%) 66/97 (68%)	2.1e-08
fn3	1134..1213	22/85 (26%) 56/85 (66%)	1.5e-08

Example 23.

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

Table 23A. NOV23 Sequence Analysis		
	SEQ ID NO: 61	2497 bp
NOV23a, CG105963-01 DNA Sequence	GCCCCGATGGACGCCGCGTTCCTCCTCGTCCTCGGGCTGTTGGCCCAGAGCCTCTGCC TGTCTTTGGGGGTTCTTGATGGAGGAGGCCACCACCCTGTACCCCTGGCGCCGGGC GCCTGCCCTGAGCCGCGTGCGGAGGGCCTGGGTTCATCCCCCGATCAGCGTATCCGAG AACCACAAGCGTCTCCCTACCCCTGGTTCAGGTGAGCAGGTGGAAGCACCAGTTGG CCAGCGTCATCTCCAGCATCCAGGGCCCCGCGTGGATGAGGAGCCCCGGGGCGTCTT CTCTATCGCCAGTTCACAGGAAGGTCTTCTCAATGCCATGCTGGACCGCGAGAAG	

	ACTGATCGCTTCAGGCTAAGAGGGTTTGGCCCTGGACCTGGGAGGATCCACCCTGGAGG ACCCCACGGACCTGGAGATTGTAGTTGTGGATCAGAATGACAACCGGCCAGCCTTCCT GCAGGAGGCGTTCCTGGCCGCGTGGTGGAGGGTGCAGTCCCAGGTACCTATGTGACC AGGGCAGAGGCCACAGATGCCGACGACCCCGAGACGGACAACGCAGCGCTGCGGTTCT CCATCCTGCAGCAGGGCAGCCCCGAGCTCTTCAGCATCGACGAGCTCACAGGAGAGAT CCGCACAGTGCAAGTGGGGCTGGACCGCGAGGTGGTTCGCGGTGTACAATCTGACCCTG CAGGTGGCGGACATGTCTGGAGACGGCCTCACAGCCACTGCTTCAGCCATCATCACCT TTGATGACATCAATGACAATGCCCCGAGTTCACCAGGGATGAGTTCTTCATGGAGGC CATAGAGGCCGTCAGCGGAGTGGATGTGGGACGCCTGGAAGTGGAGGACAGGGACCTG CCAGGCTCCCCAACTGGGTGGCCAGGTTACCATCCTGGAAGGCGACCCCGATGGGC AGTTCACCATCCGCACGGACCCCAAGACCAACGAGGGTGTCTGTCCATTGTGAAGGC CCTGGACTATGAAAGCTGTGAACACTACGAACTAAAACACACGGGCAGGATAAGACA GAGAACGCACGGGCAGGGCTGAGGGCTGAGCGGGGCCAGGCCAAGGTCCGCGTGCATG TGCAGGACACCAACGAGCCCCCGTGTTCAGGAGAACCCACTTCGGACCAGCCTAGC AGAGGGGGCAGCCCCAGGCACTCTGGTGGCCACCTTCTCTGCCCCGGGACCCTGACACA GAGCAGCTGCAGAGGCTCAGCTACTCCAAGGACTACGACCCGGAAGACTGGCTGCAAG TGGACGCAGCCACTGGCCGGATCCAGACCCAGCACGTGCTCAGCCCGGCGTCCCTTTT CCTCAAGGGCGGCTGGTACAGAGCCATCGTCTTGGCCAGGATGCCTCCAGCCCCGC ACCGCCACCGGCACCCTGTCCATCGAGATCCTGGAGGTGAACGACCATGCACCTGTGC TGGCCCCGCGCCGCGCCGGGCGAGCCTGTGCAGCGAGCCACACCAAGGCCAGGCCTCCT CCTGGGCGCCACGGATGAGGACCTGCCCCCCCCACGGGGCCCCCTTCCACTTCCAGCTG AGCCCCAGGCTCCAGAGCTCGGCCGGAAGTGGAGCCTCAGCCAGGTCAACCTCTCT CCCATCACCGTCTCCACCCAGACCCCCACCTGCCCCATGGCCCCCATTTTCATGTCTGT GGCTCACCAGCTTTTCCCCAGACCCAGCTCCGGAGCCCACAGGCGTGGCCGATGCAGA AACCTCAGGAAGGTGTGTTGTGAATGTGGGAGGGAGGGTGTGGCGGTGCTGGGCTGTG CGGGAGTTCTGACTAGGGGAAGTGGGCTCAGCCTGGGCGCACTGGTCATCGTGCTGGC CAGCGCCCTCCTGCTGCTGGTGTCTGGTCTGCTCGTGCACTCCGGGCGCGGTTCTGG AAGCAGTCTCGGGGCAAGGGGCTGCTGCACGGCCCCCAGGACGACCTTCGAGACAATG TCCTCAACTACGATGAGCAAGGAGGCGGGGAGGAGGACCAGGACGCCTACGACATCAG CCAGCTGCGTCACCCGACAGCGCTGAGCCTGCCTCTGGGACCGCCGCCACTTCGCAGA GATGCCCCGAGGGCCGCTGCACCCCCAGCCACCCCGAGTGCTGCCACCAGCCCCC TGGACATCGCCGACTTCATCAATGATGGCTTGGAGGCTGCAGATAGTGACCCAGTGT GCCGCTTACGACACAGCCCTCATCTATGACTACGAGGGTGACGGCTCGGTGGCGGGG ACGCTGAGCTCCATCCTGTCCAGCCAGGGCGATGAGGACCAGGACTACGACTACCTCA GAGACTGGGGGGCCCCGCTTCGCCCGGCTGGCAGACATGTATGGGCACCCGTGCGGGTT GGAGTACGGGGCCAGATGGGACCACCAGGCCAGGGAGGGTCTTTCTCCTGGGGCACTG CTACCCAGACACAGAGGCCGGACAGCCTGACCCTGGGGCGCAACTGGACATGCCACTC CCC		
	ORF Start: ATG at 7	ORF Stop: TGA at 2464	
	SEQ ID NO: 62	819 aa	MW at 89687.6kD
NOV23a, CG105963-01 Protein Sequence	MDAAFLVLGLLAQSLCLSLGVPGWRRPTTLYPWRRAPALSRVRRRAWVIPPISVSENH KRLPYPLVQVSRWKHQLASVISSIQGPGVDEEPRGVFSIAQFTGKVFLNAMLDREKTD RFRLRGFALDLGGSTLEDPTDLEIVVVDQNDNRPAFLQEAFTGRVLEGAVPGTYVTRA EATDADDPETDNAALRFSILQQGSPELFSIDELTGEIRTVQVGLDREVVAVYNLTQV ADMSGDGLTATASAIITFDDINDNAPEFTRDEFFMEAIEAVSGVDVGRLEVEDRDLPG SPNWVARFTILEGDPDGQFTIRTDPKTNEGVLSIVKALDYESCEHYETKTHGQDKTEN ARAGLRAERGQAKVRVHVQDTNEPPVFQENPLRTSLAEGAPPGTLVATFSARDPDTEQ LQRLSYSKDYDPEDWLQVDAATGRIQTQHVLSPASFLKGGWYRAIVLAQDASQPRTA TGTLSEILEVNDHAPVLAPPPGSLCSEPHQGPGLLLGATDEDLPPHGAPFHFQLSPL RLPELGRNWSLSQVNPLSHHRLHPDPLPHGPHFMSVAHQLFPRPSSGAHRRGRCRNL RKVCECGREGVAVVGCAGVLTRGSGLSLALVIVLASALLLVLLVALRARFWKQ SRGKGLLHGPQDDLRLDNVLYDEQGGGEEDQDAYDISQLRHPTALSLPLGPPPLRRDA PQGRLLHPQPPRVLPTSPLDIADFINDGLEAADS DPSVPYDTALIYDYEKGDSVAGTL SSILSSQGEDQDYDYLRDWGPRFARLADMYGHPGLEYGARWDHQAREGLSPGALLP RHRGRTA		

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

Table 23B. Protein Sequence Properties NOV23a	
PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 22 and 23

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

Table 23C. Geneseq Results for NOV23a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG30224	Novel human diagnostic protein #30215 - Homo sapiens, 814 aa. [WO200175067-A2, 11-OCT-2001]	1..819 1..814	750/820 (91%) 766/820 (92%)	0.0
ABG30224	Novel human diagnostic protein #30215 - Homo sapiens, 814 aa. [WO200175067-A2, 11-OCT-2001]	1..819 1..814	750/820 (91%) 766/820 (92%)	0.0
AAB24089	Human PRO2198 protein sequence SEQ ID NO:79 - Homo sapiens, 814 aa. [WO200053755-A2, 14-SEP-2000]	1..819 1..814	750/820 (91%) 766/820 (92%)	0.0
ABB57233	Mouse ischaemic condition related protein sequence SEQ ID NO:606 - Mus musculus, 906 aa. [WO200188188-A2, 22-NOV-2001]	35..786 149..902	313/767 (40%) 436/767 (56%)	e-152
AAAY70741	Human N-cadherin - Homo sapiens, 906 aa. [WO200021555-A1, 20-APR-2000]	40..786 154..902	311/762 (40%) 435/762 (56%)	e-151

5 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 23D.

Table 23D. Public BLASTP Results for NOV23a				
Protein Accession Number	Protein/Organism/Length	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P55291	Muscle-cadherin precursor (M-cadherin) (Cadherin-15) (Cadherin-14) - Homo sapiens (Human), 814 aa.	1..819 1..814	750/820 (91%) 766/820 (92%)	0.0
P33146	Muscle-cadherin precursor (M-cadherin) (Cadherin-15) (Cadherin-14) - Mus musculus (Mouse), 784 aa.	1..787 1..783	616/791 (77%) 662/791 (82%)	0.0
IJM5CM	M-cadherin - mouse, 730 aa (fragment).	56..787 1..729	576/736 (78%) 620/736 (83%)	0.0
Q8UVQ7	N-CADHERIN - Brachydanio rerio (Zebrafish) (Zebra danio), 893 aa.	39..786 140..889	316/762 (41%) 443/762 (57%)	e-157
Q90275	NEURAL-CADHERIN PRECURSOR (N-CADHERIN) - Brachydanio rerio (Zebrafish) (Zebra danio), 783 aa.	39..786 29..779	315/763 (41%) 442/763 (57%)	e-154

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

Table 23E. Domain Analysis of NOV23a			
Pfam Domain	NOV23a Match Region	Identities/ Similarities for the Matched Region	Expect Value
cadherin	50..143	23/111 (21%) 61/111 (55%)	0.011
cadherin	157..251	38/108 (35%) 74/108 (69%)	8.7e-25
cadherin	265..367	34/107 (32%) 74/107 (69%)	6.2e-18
cadherin	380..473	34/109 (31%) 71/109 (65%)	7.7e-20
Cadherin_C_term	634..788	83/158 (53%) 146/158 (92%)	5.3e-90

Example 24.

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

Table 24A. NOV24 Sequence Analysis		
	SEQ ID NO: 63	3617 bp
NOV24a, CG105973-01 DNA Sequence	GAGATGGGACTGCAATAGAAATCCGGGCAGCCCGAAGAGGCACCCAGCGCTCCAGCCA	
	CCAGCTGGGCGCGCCGGGAGTCCCTGGCTCTAGACCAGCCGCGAGGAGGCGCCGCGAG	
	AGAGCTGGTCCCTGCCCGCGGCGGAGGAGGGCTAGAGCCCCTGGGCGAGCCCCCGA	
	GCCGGCTGGGCGGGCGGGCGGGTGGGAGCAGACGCCGGGCACTGTCACCACGGGTGCG	
	CCGAGCGCACCGACCCGGGACACGGGCAGCTGGGGACCGCCAGATTCCACCAGCCCCC	
	CTTGCCCCCGCAGGGGTCTCTCGGCTCGCGCTCCTGGGTAGCAGCCACCCACCGGGGCGG	
	AGGGAGATGTCGCCCCGGGGCCAGCCGCGGTCCCCGGGGAAGCCAGGCGCCGCTGATCG	
	CGCCCCCTCTGCTGCGCCGCGGCCGCGCTGGGGATGTTGCTGTGGTCCCCCGCCTGTCA	
	GGCGTTCAACCTGGACGTGGAAAAGCTCACAGTGTACAGCGGCCCCAAGGGCAGCTAC	
	TTCGGCTACGCCGTGGACTTCCACATACCCGACGCCCGCACAGCGAGTGTCTTGGTGG	
	GGGCGCCCCAAGCCAACACCAGCCAGCCCGATATCGTGGAAGGGGGAGCCGTCTATTA	
	CTGTCTTTGGCCCCGCGGAGGGGTCTGCGCAGTGCAGGCAGATACCGTTTGACACCACC	
	AACAAACAGAAAGATCAGAGTTAATGGAACCAAAGAACCTATCGAGTTCAAATCCAATC	
	AGTGGTTTTGAGCAACAGTGAAAGCTCACAAGGAAAAGTTGTGGCCTGTGCTCCTTT	
	ATATCACTGGAGAACTCTTAAACCGACACCAGAAAAGGACCCAGTTGGCACCTGCTAT	
	GTAGCAATTCAGAACTTCAGCGCCTATGCCGAGTTCTCTCCTTGCCGGAACAGCAATG	
	CTGATCCGGAAGGCCAGGGTTACTGCCAAGCAGGATTTAGTCTGGATTTTATAAGAA	
	TGGAGACCTTATTGTGGGAGGACCTGGGAGTTTCTACTGGCAAGGACAAGTGATCACT	
	GCCAGTGTTGCAGATATCATTGCAAATTAATCATTCAAGGATATCCTCAGGAACTGG	
	CAGGAGAAAAGCAGACGGAAGTGGCTCCAGCTTCCTATGATGACAGTTACCTTGGATA	
	CTCAGTTGCTGCTGGGGAGTTTACTGGGGATTCTCAGCAAGAATTGGTTGCTGGAATT	
	CCAAGAGGAGCACAGAATTTTGGATATGTTTCCATCATTAACTCTACGGATATGACGT	
	TTATTCAGAATTTACGGGAGAACAGATGGCATCTTATTTTGGATATAACGTTGTCTG	
	ATCAGATGTTAACAGTGATGGACTGGATGATGTCCTGGTTGGGGCACCTCTCTTTATG	
	GAACGTGAATTTGAGAGCAACCCAGAGAAAGTAGGGCAAATCTACCTGTATTTGCAAG	
	TGAGCTCTCTCCTCTTCAGAGACCCCCAGATCCTCACTGGCACCCGAGACGTTTGGGAG	
	ATTCGGTAGTGCTATGGCACACTTAGGAGACCTGAACCAAGATGGATAACAATGACATT	
	GCCATCGGAGTGCCCTTTTGCAGGCAAGGATCAAAGAGGCAAAGTGCTCATTTTATAATG	
	GGAACAAAGATGGCTTAAACACCAAGCCTTCCCAAGTTCTGCAAGGAGTGTTGGGCCTC	
	ACATGCTGTCCCTTCCGGATTTGGCTTTACTTTAAGAGGAGATTCAGACATAGACAAG	
	AATGATTACCCAGATTTGATTGTGGGTGCATTTGGAACAGGAAAAGTCGCTGTTTACA	
	GAGCAAGACCGGTTGTGACTGTAGATGCCAGCTTCTGCTGCACCCAATGATTATCAA	
	TCTTGAAAATAAACTTGCCAGGTTCCAGACTCTATGACATCTGCTGCCTGCTTTTCT	
	TTAAGAGTATGTGCATCTGTACAGGCCAGAGCATTGCAAACACAATAGTCTTGATGG	
	CAGAGGTGCAATTAGATTCCCTGAAACAGAAAGGAGCTATTAAACGGACGCTCTTCCT	
	TGATAACCATCAGGCTCATCGCGTCTTCCCTCTTGTGATAAAAAGGCAGAAATCCCAC	
	CAGTGCCAGGATTTTCATCGTTTACCTTCGAGATGAAACTGAATTCCGAGATAAATTAT	
	CTCCAATCAACATTAGTTTGAATTACAGTTTGGACGAATCCACCTTTAAAGAAGGCCT	
	GGAAGTGAAACCAATATTGAACTACTACAGAGAAAACATTGTTAGTGAACAGGCTCAC	
	ATTCTGGTGGACTGTGGAGAAGACAATCTGTGTGTTCTGACTTGAAGCTGTCGGCTA	
	GACCAGATAAGCATCAGGTAATCATTGGAGATGAAAATCACCTTATGCTCATAATAAA	
	TGCAAGAAATGAAGGGGAAGGAGCATATGAAGCTGAACTCTTTGTAATGATACCAGAA	
	GAGGCAGATTATGTTGGAATCGAACGCAACAACAAGGGATTTTCGACCACTGAGCTGTG	
	AGTACAAGATGGAAAATGTAACCAGGATGGTGGTGTGTGACCTTGGGAACCCCTATGGT	
	GTCTGGAACAAATTATTCCTGGGCCTCCGATTTGCAGTTCCACGTCTTGAGAAAACA	
	AACATGAGCATTAACCTTCGATCTCAAATCAGAAGTTCCAACAAGGACAATCCAGACA	
	GCAATTTTGTGAGCCTGCAAATCAACATCACTGCTGTAGCGCAGGTGGAAATAAGAGG	
	AGTGTCACACCCTCCGCAGATTGTTCTGCCATTCTAACTGGGAACCAGAAGAGGAG	
	CCCCACAAAGAGGAGGAGGTTGGACCATTGGTGGAACATATTTATGAGCTGCACAATA	

	TTGGACCAAGTACCATCAGTGACACCATCCTGGAGGTGGGCTGGCCTTTCTCTGCCCC GGATGAATTTCTTCTCTATATTTTCCATATTCAAACCTCTGGGACCTCTGCAGTGCCAA CCAAATCCTAATATCAATCCACAGGATATAAAGCCTGCTGCCTCCCCAGAGGACACCC CTGAGCTCAGCGCCTTTTTTGCGAAACTCTACTATTCTCATCTTGTTCAGGAAGAGGGA TGTACATGTGGTCGAATTCCACAGACAGAGCCCTGCAAAAATACTGAATTGTACAAAT ATCGAGTGTTTACAAATCTCCTGTGCAGTGGGACGACTCGAAGGAGGAGAAAGCGCAG TCCTGAAAGTCAGGTCACGATTATGGGCCCACACCTTCCTCCAGAGAAAAAATGATCC CTATGCTCTTGCATCCCTGGTGTCTTTGAAGTTAAGAAGATGCCTTATACAGATCAG CCAGCAAACTCCCAGAAGGAAGCATAGTAATTAAGACATCAGTTATTTGGGCAACTC CGAATGTTTCCTTCTCAATCCCATTTATGGGTAATAATACTAGCAATACTTCTTGGATT GTTGGTTCTCGCCATTTTAACCTTAGCTTTATGGAAGTGTGGATTCTTTGACAGAGCC AGACCTCCTCAGGAGGACATGACCGACAGGGAACAGCTGACAAATGACAAGACCCCTG AGGCATGACAAGAAAAAAAAAAGAAGACCAAGACCTCAAACACTGGTCCTGTTCAAAG AAAAAGAAAGAACATGAGGCC		
	ORF Start: ATG at 355	ORF Stop: TGA at 3544	
	SEQ ID NO: 64	1063 aa	MW at 117472.3kD
NOV24a, CG105973-01 Protein Sequence	MSPGASRGPRGSQAPLIAPLCCAAAALGMLLWSPACQAFNLDVEKLTIVYSGPKGSYFG YAVDFHIPDARTASVLVGAPKANTSQPDIVEGGAVYYCPWPAEGSAQCRQIPFDTTNN RKIRVNGTKEPIEFKSNQWFGATVKAHKGKVVACAPLYHWRTLKPTPEKDPVGTCTYVA IQNFSAYAEFSPCRNSNADPEGQGYCQAGFSLDFYKNGDLIVGGPGSFYWQQQVITAS VADIIANYSEFKDILRKLAGEKQTEVAPASYDDSYLGYSVAAGEFTGDSQQELVAGIPR GAQNFGYVSIINSTDMTFIQNFTGEQMASYFGYTVVSDVNSDGLDDVLVGAPLFMER EFESNPREVGQIYLYLQVSSLLFRDPQILTGTETFGFRFGSAMAHGLDNLQDGYNDIAI GVPPFAGKDQRGKVLINGNKDGLNTKPSQVLQGVWASHAVPSGFGFTLRGDSDDIDKND YPDLIVGAFGTGKVAVYRARPVVTVDALLLHPMIINLENKTCQVPDSMTSAACFSLR VCASVTGQSIANTIVLMAEVQLDSLKQKGAIKRTLFLDNHQHRVFLVIKQKSHQC QDFIVYLRDETEFRDKLSPINISLNYSLDESTFKEGLEVKPILNYYRENIVSEQAHIL VDCGEDNLCVPDLKLSARPDKHQVIIGDENHMLIINARNEGEGAYEAEFLVMIPEEA DYVGIERNKNGFRPLSCEYKMENVTRMVVCDLGNPMVSGTNYSLGLRFAVPRLEKTNM SINFDLQIRSSNKDNPDSNFVSLQINITAVAQVEIRGVSHPPQIVLPIHNWEPEEEPH KEEEVGPLVEHIYELHNIGPSTISDTILEVGWPFSSARDEFLLYIFHIQTLGPLQCQPN PNINPQDIKPAASPEDTPELSAFLRNSTIPLVLRKRDVHVVEFHRQSPAKILNCTNIE CLQISCAVGRLEGGESAVLKVRSRLWAHTFLQRKNDPYALASLVSEVKKMPYTDQPA KLPEGSIVIKTSVIWATPNVSFSIPLWVILAILLGLLVLAILTLALWKCGFFDRARP PQEDMTDREQLTNDKTPEA		
	SEQ ID NO: 65	3617 bp	
NOV24b, CG105973-02 DNA Sequence	GAGATGGGACTGCAATAGAAATCCGGGCAGCCCGAAGAGGCACCCAGCGCTCCAGCCA CCAGCTGGGCGCGCCGGGAGTCCCTGGCTCTAGACCAGCCGCGAGGAGGCGCCGCGAG AGAGCTGGTCCCTGCCCGCGGCCGGAGGAGGGCTAGAGCCCCTGGGCCAGCCCCCGA GCCGGCTGGGCGGGCGGGCGGGTGGGAGCAGACGCCGGGCACTGTCACCACGGGTGCG CCGAGCGCACCCAGCCGGGACACGGGCAGCTGGGGACCGCCAGATTCCACCAGCCCCC CTTGCCCCGCGAGGGGTCTCGGCTCGCGCTCCTGGGTAGCAGCCACCCACCGGGGCGG AGGGAGATGTCGCCCGGGGCCAGCCGCGGTCCCCGGGGAAGCCAGGCGCCGCTGATCG CGCCCCCTCTGCTGCGCCGCGGCCGCGCTGGGGATGTTGCTGTGGTCCCCCGCCTGTCA GGCGTTCAACCTGGACGTGGAAAAGCTCACAGTGTACAGCGGCCCAAGGGCAGCTAC TTCGGCTACGCCGTGGACTTCCACATACCCGACGCCCCGCACAGCGAGTGTCTTGGTGG GGGCGCCCCAAAGCCAACACCAGCCAGCCCGATATCGTGGAAAGGGGGAGCCGTCTATTA CTGTCCTTGGCCCCGCGGAGGGGTCCGCGCAGTGCAGGCAGATACCGTTTGACACCACC AACAACAGAAAGATCAGAGTTAATGGAACCAAAGAACCTATCGAGTTCAAATCCAATC AGTGGTTTGGAGCAACAGTGAAAGCTCACAAAGGAAAAGTTGTGGCCTGTGCTCCTTT ATATCACTGGAGAACTCTTAAACCGACACCAGAAAAGGACCCAGTTGGCACCTGCTAT GTAGCAATTCAGAACTTCAGCGCCTATGCCGAGTTCTCTCCTTGCCGGAACAGCAATG CTGATCCGGAAGGCCAGGGTTACTGCCAAGCAGGATTTAGTCTGGATTTTTATAAGAA TGGAGACCTTATTGTGGGAGGACCTGGGAGTTTCTACTGGCAAGGACAAGTGATCACT GCCAGTGTTCAGATATCATTGCAAATTACTCATTCAAGGATATCCTCAGGAACTGG CAGGAGAAAAGCAGACGGAAGTGGCTCCAGCTTCCTATGATGACAGTTACCTTGGATA CTCAGTTGCTGCTGGGGAGTTTACTGGGGATTCTCAGCAAGAATTGGTTGCTGGAATT CCAAGAGGAGCACAGAATTTTGGATATGTTTCCATCATTAACCTCTACGGATATGACGT		

	TTATTCAGAATTTACCGGGAGAACAGATGGCATCTTATTTTGGATATACCGTTGTCGT ATCAGATGTTAACAGTGATGGACTGGATGATGTCCTGGTTGGGGCACCTCTCTTTATG GAACGTGAATTTGAGAGCAACCCAGAGAAAGTAGGGCAAATCTACCTGTATTTGCAAG TGAGCTCTCTCTCTCTCAGAGACCCAGATCCTCACTGGCACCGAGACGTTTGGGAG ATTTCGGTAGTGCTATGGCACACTTAGGAGACCTGAACCAAGATGGATACAATGACATT GCCATCGGAGTGCCTTTTTGCAGGCAAGGATCAAAGAGGCAAAGTGCTCATTATATAATG GGAACAAAGATGGCTTAAACACCAAGCCTTCCCAAGTTCTGCAAGGAGTGTTGGGCCTC ACATGCTGTCCCTTCCGGATTTGGCTTTACTTTAAGAGGAGATTGAGACATAGACAAG AATGATTACCCAGATTTGATTGTGGGTGCATTTGGAACAGGAAAAGTCGCTGTTTACA GAGCAAGACCGGTTGTGACTGTAGATGCCAGCTTCTGCTGCACCCAATGATTATCAA TCTTGAAAATAAACTTGCCAGGTTCCAGACTCTATGACATCTGCTGCCTGCTTTTCT TTAAGAGTATGTGCATCTGTACAGGCCAGAGCATTGCAAACACAATAGTCTTGATGG CAGAGGTGCAATTAGATTCCCTGAAACAGAAAGGAGCTATTAAACGGACGCTCTTCCCT TGATAACCATCAGGCTCATCGCGTCTTCCCTCTTGTGATAAAAAGGCAGAAATCCAC CAGTGCCAGGATTTTCATCGTTTACCTTCGAGATGAAACTGAATTCGAGATAAATTAT CTCCAATCAACATTAGTTTGAATTACAGTTTGGACGAATCCACCTTTAAAGAAGGCCT GGAAGTGAAACCAATATTGAACTACTACAGAGAAAACATTGTTAGTGAACAGGCTCAC ATTCTGGTGGACTGTGGAGAAGACAATCTGTGTGTTCTGACTTGAAGCTGTCGGCTA GACCAGATAAGCATCAGGTAATCATTGGAGATGAAAATCACCTTATGCTCATAATAAA TGCAAGAAATGAAGGGGAGGGAGCATATGAAGCTGAACTCTTTGTAATGATACCAGAA GAGGCAGATTATGTTGGAATCGAACGCAACAACAAGGGATTTGACCACTGAGCTGTG AGTACAAGATGGAAAATGTAACCAGGATGGTGGTGTGTGACCTTGGGAACCCATGGT GTCTGGAACAAATTATTTCCCTGGGCCTCCGATTTGCAGTTCCACGTCTTGAGAAAACA AACATGAGCATTAACCTTCGATCTCCAAATCAGAAGTTCCAACAAGGACAATCCAGACA GCAATTTTGTGAGCCTGCAAATCAACATCACTGCTGTAGCGCAGGTGGAAATAAGAGG AGTGTACACCCCTCCGCAGATTGTTCTGCCCATTCATAACTGGGAACCAGAAAGAGGAG CCCCACAAAGAGGAGGAGGTTGGACCATTGGTGGAACATATTTATGAGCTGCACAATA TTGGACCAAGTACCATCAGTGACACCATCCTGGAGGTGGGCTGGCCTTTCTCTGCCCG GGATGAATTTCTTCTCTATATTTTCCATATTTCAAACCTCTGGGACCTCTGCAGTGCCAA CCAAATCCTAATATCAATCCACAGGATATAAAGCCTGCTGCCTCCCCAGAGGACACCC CTGAGCTCAGCGCCTTTTTTGCGAACTCTACTATTCCTCATCTTGTGAGGAAGAGGGA TGTACATGTGGTCAATTCCACAGACAGAGCCCTGCAAAAATACTGAATTGTACAAAT ATCGAGTGTTTACAAATCTCCTGTGCAGTGGGACGACTCGAAGGAGGAGAAAGCGCAG TCCTGAAAGTCAGGTCACGATTATGGGCCACACCTTCTCCAGAGAAAAAATGATCC CTATGCTCTTGCATCCCTGGTGTCTTTGAAGTTAAGAAGATGCCTTATACAGATCAG CCAGCAAACTCCCAGAAGGAAGCATAGCAATTAAGACATCAGTTATTTGGGCAACTC CGAATGTTTCTTCTCAATCCCATTTATGGGTAATAATACTAGCAATACTTCTTGGATT GTTGGTTCTCGCCATTTTAACCTTAGCTTTATGGAAGTGTGGATTCTTTGACAGAGCC AGACCTCCTCAGGAGGACATGACCGACAGGGAACAGCTGACAAATGACAAGACCCCTG AGGCATGACAAGAAAAAAAAGAAGACCAAGACCTCAAACACTGGTCCTGTTCAAAG AAAAAGAAAGAACATGAGGCC		
	ORF Start: ATG at 355	ORF Stop: TGA at 3544	
	SEQ ID NO: 66	1063 aa	MW at 117444.2kD
NOV24b, CG105973-02 Protein Sequence	MSPGASRGPRGSQAPLIAPLCCAAAALGMLLWSPACQAFNLDVEKLTVYSGPKGSYFG YAVDFHIPDARTASVLVGAPKANTSQPDIVEGGAVYYCPWPAEGSAQCRQIPFDTTNN RKIRVNGTKEPIEFKSNQWFGATVKAHKGVVACAPLYHWRTLKPTPEKDPVGTCTYVA IQNFSAYAEFSPCRNSNADPEGQGYCQAGFSLDFYKNGDLIVGGPGSFYWQGQVITAS VADIIANYSFKDILRLKLAGKEQTEVAPASYDDSYLGYSVAAGEFTGDSQQELVAGIPR GAQNFGYVSIINSTDMTFIQNFTGEQMASYFGYTVVSDVNSDGLDDVLVGAPLFMER EFESNPREVGQIYLYLQVSSLLFRDPQILTGTFGRFGSAMAHLDLNQDGYNDIAI GVPFAGKDQRGKVLIIYNGNKDGLNTKPSQVLQGVWASHAVPSGFGFTLRGDSIDKND YFDLIVGAFGTGKVAVYRARPVVTVDQLLLHPMIINLENKTCQVPDSMTSAACFSLR VCASVTGQSIANTIVLMAEVQLDSLKQKGAIKRTLFLDNHQHRVFLVIKQKSHQC QDFIVYLRDETEFRDKLSPINISLNYSLDESTFKEGLEVKPILNYYRENIVSEQAHL VDCGEDNLCVPDLKLSARPDKHQVIIGDENHMLMIINARNEGEGAYEAEFVMIPEEA DYVGIERNKNGFRPLSCEYKMENVTRMVVCDLGNPMVSGTNYSGLRFAVPRLEKTNM SINFDLQIRSSNKDNPDSNFVSLQINITAVAQVEIRGVSHPPQIVLPIHNWEPEEPEPH KEEEVGPLVEHIYELHNIGPSTISDTILEVGWPF SARDEFLLYIFHIQTLGPLQCQPN PNINPQDIKPAASPEDTPELSAFLRNSTIPHLVRKRDVHVVEFHRQSPAKILNCTNIE		

CLQISCAVGRLEGGESAVLKVRSRLWAHTFLQRKNDPYALASLVSFEVKKMPYTDQPA KLPEGSIAIKTSVIWATPNVSFSIPLWVIILAILLGLLVLAILTALWKC GFDRARP PQEDMTDREQLTNDKTPEA
--

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 24B.

Table 24B. Comparison of NOV24a against NOV24b.		
Protein Sequence	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV24b	1..1063	1010/1063 (95%)
	1..1063	1010/1063 (95%)

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

Table 24C. Protein Sequence Properties NOV24a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1125 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 39 and 40

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

Table 24D. Geneseq Results for NOV24a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM39241	Human polypeptide SEQ ID NO 2386 - Homo sapiens, 1035 aa. [WO200153312-A1, 26-JUL-2001]	29..1063 1..1035	1031/1035 (99%) 1031/1035 (99%)	0.0
AAM41027	Human polypeptide SEQ ID NO 5958 - Homo sapiens, 1044 aa. [WO200153312-A1, 26-JUL-2001]	24..1063 1..1044	1024/1046 (97%) 1027/1046 (97%)	0.0
ABG18895	Novel human diagnostic protein #18886 - Homo sapiens, 1061 aa. [WO200175067-A2, 11-OCT-2001]	8..1055 25..1049	500/1052 (47%) 692/1052 (65%)	0.0
ABG18895	Novel human diagnostic protein #18886 - Homo sapiens, 1061 aa. [WO200175067-A2, 11-OCT-2001]	8..1055 25..1049	500/1052 (47%) 692/1052 (65%)	0.0
AAB70508	Tissue remodeling protein alpha 5 beta 1 integrin (VLA-5) protein - Mammalian, 1049 aa. [WO200111086-A2, 15-FEB-2001]	34..1063 37..1049	474/1036 (45%) 667/1036 (63%)	0.0

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

Table 24E. Public BLASTP Results for NOV24a				
Protein Accession Number	Protein/Organism/Length	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P53708	Integrin alpha-8 - Homo sapiens (Human), 1025 aa.	39..1063 1..1025	1020/1025 (99%) 1020/1025 (99%)	0.0
O70304	INTEGRIN ALPHA8 - Mus musculus (Mouse), 1012 aa (fragment).	46..1057 1..1012	910/1012 (89%) 972/1012 (95%)	0.0
P26009	Integrin alpha-8 precursor - Gallus gallus (Chicken), 1044 aa.	27..1063 13..1044	797/1037 (76%) 907/1037 (86%)	0.0
P26008	Integrin alpha-V precursor (Vitronectin receptor alpha subunit) (CD51) - Gallus gallus (Chicken), 1034 aa.	35..1055 16..1022	493/1024 (48%) 678/1024 (66%)	0.0
Q9MZD6	INTEGRIN ALPHA V SUBUNIT PRECURSOR - Bos taurus (Bovine), 1047 aa.	8..1055 12..1035	508/1057 (48%) 692/1057 (65%)	0.0

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

Table 24F. Domain Analysis of NOV24a			
Pfam Domain	NOV24a Match Region	Identities/ Similarities for the Matched Region	Expect Value
FG-GAP	54..117	22/67 (33%) 53/67 (79%)	1.4e-15
FG-GAP	264..317	19/63 (30%) 42/63 (67%)	2.1e-06
FG-GAP	318..383	23/66 (35%) 49/66 (74%)	1.7e-15
FG-GAP	384..443	29/67 (43%) 53/67 (79%)	2e-16
FG-GAP	447..501	22/66 (33%) 41/66 (62%)	7.8e-10
integrin_A	1035..1049	10/15 (67%) 14/15 (93%)	0.00011

Example 25.

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

Table 25A. NOV25 Sequence Analysis			
	SEQ ID NO: 67	1524 bp	
NOV25a, CG106915-01 DNA Sequence	TTTGCCATCATGTTGCGGTTGGTGGCAGCTTGCCCTGAGTCATGTGTGGTGTGCACCA AAGATGTAACCCTCTGTCAACCAGCTAACCTATATAGTAGCAGCCCCTATGACCACGAG GGTTTTAATCATCACCGATGGATATCTCTCCTCTATTGAGAGCACAAACCTGTCTCTC TTGTTTAATCTTGCCCTGCTCTCCCTAAGCAGAAATGGTATCGAGGATGTTTCAGGAAG ATGCCCTGCATGGGCTTACGATGTTGCGGACCTTGTTGCTGGAGCACAACCAAATATC CAGCTCTTCGCTCACTGATCACACCTTCAGCAAGCTTCACAGCCTGCAGGTACTGGTG CTGAGCAATAATGCTCTCCGCACCCTACGAGGGTCTTGTTCCGAAACACAAGCGGCC TGACCCGGCTCCAGCTGGATGGGAATCAGATTACTAATCTCACAGACAGTTCTTTCGG AGGCACGAATCTCCACAGTCTCAGGTATCTGGATTTATCCAACAATTTTATTTCTTAC ATTGGGAAAGATGCCTTCCGGCCCCCTGCCTCAACTACAGGAAGTGGACCTTTCCTGAA ATAGGTTAGCCACATGCCGGATGTGTTTACTCCACTGAAGCAGTTAATCCTTCTGAG CTTAGATAAGAACCAGTGGAGCTGCACCTTGTGATCTCCATCCCCCTTGCTCGGTTTTTA AGAAACTACATTAAGTCTTCTGCTCACACGCTCAGGAATGCCAAGGACCTAAATTGCC AGCCATCTACCGCAGCTGTGGCAGCTGCACAGAGTGTGCTGAGGCTGTCTGAGACCAA CTGTGATTCCAAAGCTCCCAACTTCACTCTGGTTCTAAAGGACAGAAGTCCCCTCCTC CCAGGACCAGATGTGGCCCTGCTGACTGTCCTTGGCTTCGCAGGTGCTGTTGGTCTCA CTTGCCTAGGTTTAGTTGTATTTAACTGGAACTCCACCAAGGCAAAGCAAATGAACA CACATCAGAAAACCTTTGTTGCAGAACCTTCGATGAACCCCTGTGTGCTCATGAGGCA AGAAATTACCACACTAAGGGATACTGCAACTGCCACTTAACTCAGGAAAACGAGATAA AGGTCATGTCCATTGTGGGGTCCAGAAAAGAAATGCCACTTTTACAGGAAAATAGCCA TCAAGCAACATCGGCCTCTGAGTCTGCAACCCCTTGACAGATCATTTAGAAACCTGAAA AAGAAAGACCGTGGGGTAGGCAGCACTTTATTTTGGCAGGATGGTAGATTGCTGCATT CGGAATGTTTCAGAGCCTCCTGGAAATATGAGAGCTTTTAATGAAGCAGGCTTACTTAC AACATATAATCCAAGGAAAGTTCAAAAGCTATGGAATCTTGAGCCTGGAGAAGTCCAG CCTCAAACCTCTGCAACACCATATAATAAGAACAGAAGATATCAGCAGTGACATATTTA GAAGAAGATATGCAACACCCGCTTCAGCCTTGGCAGGAGAAAGTCTTGAGAAGCGTTT AACAAATGAATCATGA		
	ORF Start: ATG at 10	ORF Stop: TGA at 1522	
	SEQ ID NO: 68	504 aa	MW at 56079.2kD
NOV25a, CG106915-01 Protein Sequence	MLRLVAACPESCVVCTKDVTLCQLTYIVAAPMTTRVLIITDGYLSSIESTNLSLLEFN LALLSLSRNGIEDVQEDALHGLTMLRTLLEHNQISSSSLTDHTFSLHSLQVLVLSN NALRTLGRSWFRNTSGLTRLQLDGNQITNLTDSSFGGTNLHSLRYLDLSNNFISYIGK DAFRPLPQLQEVDLSRNRLAHMPDVFTPLKQLILLSLDKNQWSCTCDLHPLARFLRNY IKSSAHTLRNAKDLNCQPSTA AVAAQSVLRLSETNCD SKAPNFTLV LKDRSPLLPGP DVALLTVLGFAGAVGLTCLGLVVFNWKLHQKANEHTSENLC CRTFDEPLCAHEARNY HTKGYCNCHLTQENEIKVMSIVGSRKEMPLLQENSHQATSASESATLDRSFRNLKKKD RGVGSTLFCQDGRLLHSECSEPPGNMRAFNEAGLLT TYNPRKVQKLWNLEPGEVQPQT LQHIIIRTEDISSDIFRRRYATPASALAGESLEKRLTNES		

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

Table 25B. Protein Sequence Properties NOV25a	

analysis:	probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.2622 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV25a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 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
AAU83655	Human PRO protein, Seq ID No 128 - Homo sapiens, 473 aa. [WO200208288-A2, 31-JAN-2002]	3..237 22..283	79/264 (29%) 121/264 (44%)	5e-19
AAB49891	Human PRO526 protein sequence - Homo sapiens, 473 aa. [WO200070050-A1, 23-NOV-2000]	3..237 22..283	79/264 (29%) 121/264 (44%)	5e-19
AAB50908	Human PRO526 protein - Homo sapiens, 473 aa. [WO200073452-A2, 07-DEC-2000]	3..237 22..283	79/264 (29%) 121/264 (44%)	5e-19
AAU04589	Human Nogo receptor - Homo sapiens, 473 aa. [WO200151520-A2, 19-JUL-2001]	3..237 22..283	79/264 (29%) 121/264 (44%)	5e-19
AAU12362	Human PRO526 polypeptide sequence - Homo sapiens, 473 aa. [WO200140466-A2, 07-JUN-2001]	3..237 22..283	79/264 (29%) 121/264 (44%)	5e-19

5 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
Q9BGY6	HYPOTHETICAL 56.5 KDA PROTEIN - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 510 aa.	1..504 1..504	478/504 (94%) 481/504 (94%)	0.0
Q961X3	GH01279P - Drosophila melanogaster (Fruit fly), 615 aa.	45..233 313..522	68/212 (32%) 103/212 (48%)	3e-21
Q9N0E3	UNNAMED PROTEIN PRODUCT - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 473 aa.	3..237 22..283	82/264 (31%) 125/264 (47%)	1e-19
Q9VZ84	CG7509 PROTEIN - Drosophila melanogaster (Fruit fly), 633 aa.	45..233 313..540	69/230 (30%) 103/230 (44%)	6e-19
Q9BZR6	NOGO RECEPTOR - Homo sapiens (Human), 473 aa.	3..237 22..283	79/264 (29%) 121/264 (44%)	1e-18

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
LRR	58..81	7/25 (28%) 19/25 (76%)	0.3
LRR	108..131	10/25 (40%) 19/25 (76%)	0.11
LRR	132..155	8/25 (32%) 18/25 (72%)	0.7
LRR	158..181	11/25 (44%) 19/25 (76%)	0.00021
LRR	182..204	10/25 (40%) 18/25 (72%)	0.093
LRRCT	214..270	15/63 (24%) 42/63 (67%)	0.046

Example 26.

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

Table 26A. NOV26 Sequence Analysis		
	SEQ ID NO: 69	3757 bp
NOV26a, CG106924-01 DNA Sequence	CTCTTTGCCATCATGTTGCGGTTGGTGGCAGCTTGCCCTGAGTCATGTGTGGTGTGCA CCAAAGATGTAACCCCTCTGTCCACCAGCTAACCTATATAGTAGCAGCCCCCTATGACCAC GAGGGTTTTAATCATCACCGATGGATATCTCTCCTCTATTGAGAGCACAAACCTGTCT CTCTTGTTTAATCTTGCCCTGCTCTCCCTAAGCAGAAATGGTATCGAGGATGTTTCAGG AAGATGCCCTGCATGGGCTTACGATGTTGCGGACCTTGTGCTGGAGCACAAACCAAT ATCCAGCTCTTCGCTCACTGATCACACCTTCAGCAAGCTTCACAGCCTGCAGGTAAGT GTGCTGAGCAATAATGCTCTCCGCACCCCTACGAGGGTCTTGGTTCCGAAACACAAGCG GCCTGACCCGGCTCCAGCTGGATGGGAATCAGATTACTAATCTCACAGACAGTTCTTT CGGAGGCACGAATCTCCACAGTCTCAGGTATCTGGATTTATCCAACAATTTTATTTCC TACATTGGGAAAGATGCCTTCCGGCCCCCTGCCTCAACTACAGGAAGTGGACCTTTCCC GAAATAGGTTAGCCACATGCCGGATGTGTTTACTCCACTGAAGCAGTTAATCCTTCT GAGCTTAGATAAGAACCAGTGGAGCTGCACTTGTGATCTCCATCCCCCTTGCTCGGTTT TTAAGAACTACATTAAGTCTTCTGCTCACACGCTCAGGAATGCCAAGGACCTAAATT GCCAGCCATCTACCGCAGCTGTGGCAGCTGCACAGAGTGTGCTGAGGCTGTCTGAGAC CAACTGTGATTCCAAAGCTCCCAACTTCACTCTGGTTCTAAAGGACAGAAGTCCCCTC CTCCCAGGACCAGATGTGGCCCTGCTGACTGTCTTGGCTTCGCAGGTGCTGTTGGTC TCACTTGCCTAGGTTTAGTTGTATTAACTGGAACTCCACCAAGGCAAAGCAAATGA ACACACATCAGAAAACCTTTGTTGCAGAACCCTTCGATGAACCCCTGTGTGCTCATGAG GCAAGAAATTACCACACTAAGGGATACTGCAACTGCCACTTAACCTCAGGAAAACGAGA TAAAGGTCATGTCCATTGTGGGGTCCAGAAAAGAAATGCCACTTTTACAGGAAATAG CCATCAAGCAACATCGGCCTCTGAGTCTGCAACCCTTGACAGATCATTTAGAAACCTG AAAAAGAAAGACCGTGGGGTAGGCAGCACTTTATTTTGCCAGGATGGTAGATTGCTGC ATTCGGAATGTTTCAGAGCCTCCTGGAAATATGAGAGCTTTTAATGAAGCAGGCTTACT TACAACATATAATCCAAGGAAAGTTCAAAAGCTATGGAATCTTGAGCCTGGAGAAGTC CAGCCTCAAACCTCTGCAACACCATATAATAAGAACAGAAGATATCAGCAGTGACATAT TTAGAAGAAGATATGCAACACCCGCTTCAGCCTTGGCAGGAGAAAGTCTTGAGAAGCG TTTAACAATGAATCATGGCAGCCTCCAATAGAAAAAGAAGACAATGGCTTACACCCT CACAGGCAAAGACATTTTATTACAAGCTCATCATCCAAGCCTTGTGAGCCTGAGGAAC ACTATGTACAAAATATCGTACAAAAAATAGATCAAAATATGATGATCCTTGTGGACT GTTAAAACAGAGCAAACCTAGGTATTTTCAGCCAAACAATTCCTTATCTGTAAATAT GTGCCCTGTGAGCAATTTGAAGATTACATGAAAGAAAAGAAGCCAAATCGTAGACAAC ACTCAAAGCCTGAGAAAGAGCAAATCCAATTAACAGTGAATAGAAAAATTTCTTAT GAGTGAGGACAACATAGATTTATCAGGATTATCAACAAAACCAAGAAAGCATATTCC CCAAAGAGGGTTATCTTCCATGATCCTGATTTAGTAGAAATAAATAGGTTCGATGATGT CACCCAAAATATCAACCCCTTGGAACGACAGAAAAATCAAAGTAACCAACTGACTAA GTTGGATGTTAAAAAATTTAGCAACACTGGGGAGAGAAACAAAGGAGAAAAATGGTTT ACTAATTCATGGGTTCTGAAAAGGAAGAGAACCCTCAGTCTGACCTCAAAGGGAAAA TTAAAGGACAAAACCTTAAATTTAAATTTACATCCTTTTAGAAAAGTCAGAGTCCATCC AGAAAAATCCTTGTCAAGTCTCCCAAAGCAATGCAAGCAGGTATTGTTGCCTCCTAAG AAATTATCCAAAACCTTCTGAGACAGAAGCCAAAATAAATACTGTGTGTTCTGCAGATT TTCTTCAACAGTCAGAGAGTAGCAACTATGTTAGACTCACTTCAAAGAGGCTGCCTCT GAAACATGACTCAAAGCAGACCCCATATTATCAACGAAACACTAAACGTGCCCCCTG CTCAGTGCTAACAACCTTGCCTGTAGTCAACCAGAGCTCTATAGAAAGCAGCTGTTACT CAGCTGGCCACATTCCTGATGGAAACACATCAAAATTGCCCAACCTACACCCACTGA TGCTGAGCACAGGCACTCACATTCTCAATTCTCAACTGAGCAAATGGAAGATGCAACT CAGCTTGAATCAAAGTGCTTAGTTATTTAGCAACTACTTGGGAAAATACAGGAAGTG ATGTTTTTACCATTCCAACATTCCAGGAGGGCTACTGACCAAGGGACAACGGAGTCCAC TGAGCACATGGGACAGAATGTATCAAAGACCAGTGAGTTAAATCAGTTTTCTTTGTCC CCGAGGAATCAAACACAACCTTTTAGATGCTCACAAGACTGACAGCTACAACAAGGAAT	

	ACACTTTAGACCAAAATGAAGGCTTACAACACAGAGAGCAAAATTCAAGTCATGCACA GCTTGAAAATAAAGAAAAAACATTAATGACAAAACCCCAAATACCACATCAAATTGTG GAAAATTGTATTATGGATAAGGAAGAAAATGATGTAGAAAAAAACTTTCAAAAACAG AACTTATGATTCCTCTCTCATTCCCCAAACACAATCCAAGAACAACCTATCATTTAT GAAGACAAATTCAATTCCATACCAAAATAGAATAGAACTTCCCAAGGATATCAGTACT TCTCCTGTTAGTAGTCAAGCCGTTTGGCACCTAACCAATAGTAGCGAAAAAGGAATTG ACAGCACAAATGCATTGCCCAGAAATGACGGCACTGAAGCACTAGAGATAAAAATAGT AGGGAAAGAAGAGAAAAATATGCTTGATGAAAGCAAGACAGATTCTAGTATGTAACT CAGATCTCACAAATGACCTTAAAAGGCATCACAAAAGAAAGGCAGCAAACTTGGGAAA ATGGAACAAGTGAAAAATATATATTACATGATGCAAGCTCTGCCGAGGAGACCATTAC AGCTAAAGATTTAAGTATCACAGTTCCCATGAAACCCAAAATAGAATACTTTGCAGT GAAGTAGATCCTGAAGTTAACAGTAATGTACATAATTTTAGAGAAGTTCAAAATATTC AACCAGATAAAGATAGGGCACATAAAGAAGGCGCAATGACAGTGGAGACACATGAAGC GCTTTCCTTCTTACCAGGGTTAAAAGACAGTTTTTGAGGCAGAAAATGAGGTGTTTTTA GTTCTTAGCAGAATAAATGAAGCTGAAAACCTCTGCTCCAAAACCTGTACTGTATCCAC CATCTGCTGAATATGCTACTACATCACCTTTAGAAACAGAATAAA		
	ORF Start: ATG at 13	ORF Stop: TAA at 3754	
	SEQ ID NO: 70	1247 aa	MW at 140902.2kD
NOV26a, CG106924-01 Protein Sequence	MLRLVAACPESCVVCTKDVTLCHQLTYIVAAPMTTRVLIITDGYLSSIESTNLSLLEN LALLSLSRNGIEDVQEDALHGLTMLRTLLEHNQISSSSLTDHTFSKLHSLQVLVLSN NALRTLGRGSWFRNTSGLTRLQLDGNQITNLTDSSFGGTNLHSLRYLDLSNNFISYIGK DAFRPLPQLQEVDLSRNRLAHMPDVFTPLKQLILLSLDKNQWSCTCDLHPLARFLRNY IKSSAHTLRNAKDLNCQPSTA AVAAAQSVLRLSETNCD SKAPNFTLV LKDRSPLLPGP DVALLTVLGFAGAVGLTCLGLVVFNWKLHQKANEHTSENLCRTFDEPLCAHEARNY HTKGYCNCHLTQENEIKVMSIVGSRKEMPLLQENSHQATSASESATLDRSFRNLKKKD RGVGSTLFCQDGRLLHSECSEPPGNMRAFNEAGLLTTYNPRKVQKLWNLEPGEVQPQT LQHHIIRTEDISSDIFRRRYATPASALAGESLEKRLTNESWQPPIEKEDNGLHPRQR HFITSSSSSKPCEPEEHYVQNIQKNRSKYDDPCGLLKQSKPRYFQPNNSLICKYVPC QFEDYMKEKKPNRRQHSEKEQIQINSAIEKFLMSEDNIDLSGLSTKTKKAYSPPRV IFHDPDLVEINRSMSPKISTPWKRQKNQSNQLTKLDVKKFSNTGERNKGEKWFNSW VLKRKRTPQSDLKGKIKGQNLKLNHPFRKVRVHPEKSLSSLPKQCKQVLLPPKKLSK TSETEAKINTVCSADFLQQSESSNYVRLTSKRLPLKHDSKQTPYYQRNTKRAPLLSAN NLRVNVQSSIESSCYSAGHIPDGNTSKLPQPTPTDAEHRHSHSQFSTEQMEDATQLES KVLSYLATTWENTGSDVLPFQHSRRATDQGTTESTEHMGQNVSKTSELNQFSLSPRNQ TQLLDAHKTDSDYNKEYTLDQNEGLQHREQNSSSHAQLENKEKTLMTKPQIPHQIVENCI MDKEENDVEKKLSKTETYDSSLIPQTQSKNNLSFMKTNSIPYQNRIELPKDISTSPVS SQAVWHLTNSSEKGIDSTNALPRNDGTEALEIKIVGKEEKNMLDESKTDSMLTQISQ MTLKGITKERQQTWENGTSKYILHDASSAETITAKDLSITSSHETQNRILCSEVDP EVNSNVHNFREVQNIQPKDRAHKEGAMTVETHEALSFLPGLKDSFEAENEVFLVPSR INEAENSAPKPVLYPPSAEYATTSLETE		
	SEQ ID NO: 71	645 bp	
NOV26b, 210062144 DNA Sequence	GGATCCGCCCTGCTCTCCCTAAGCAGAAATGGTATCGAGGATGTTTCAGGAAGATGCCC TGCATGGGCTTACGATGTTGCGGACCTTGTTGCTGGAGCACAAACCAATATCCAGCTC TTCGCTCACTGATCACACCTTCAGCAAGCTTCACAGCCTGCAGGTACTGGTGCTGAGC AATAATGCTCTCCGCACCCTACGAGGGTCTTGGTTCCGAAACACAAGCGGCCTGACCC GGCTCCAGCTGGATGGGAATCAGATTACTAATCTCACAGACAGTTCTTTCGGAGGCAC GAATCTCCACAGTCTCAGGTATCTGGATTTATCCAACAATTTTATTTCCTACATTGGG AAAGATGCCTTCCGGCCCCCTGCCTCAACTACAGGAAGTGGACCTTTCCCGAAATAGGT TAGCCACATGCCGGATGTGTTTACTCCACTGAAGCAGTTAATCCTTCTGAGCTTAGA TAAGAACCAGTGGAGCTGCACTTGTGATCTCCATCCCCTTGCTCGGTTTTTAAGAAAC TACATTAAGTCTTCTGCTCACACGCTCAGGAATGCCAAGGACCTAAATTGCCAGCCAT CTACCGCAGCTGTGGCAGCTGCACAGAGTGTGCTGAGGCTGTCTGAGACCAACTGTGA TCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 72	215 aa	MW at 23982.8kD
NOV26b,	GSALLSLSRNGIEDVQEDALHGLTMLRTLLEHNQISSSSLTDHTFSKLHSLQVLVLS		

210062144 Protein Sequence	NNALRTLRGSWFRNTSGLTRLQLDGNQITNLTDSSEGGTNLHSLRYLDLSNNFISYIG KDAFRPLPQLQEVDLSRNRLAHMPDVFTPLKQLILLSLDKNQWSCTCDLHPLARFLRN YIKSSAHTLRNAKDLNCQPSTAATAVAAAQSVLRLSETNCDLE
-------------------------------	---

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

Table 26B. Comparison of NOV26a against NOV26b.		
Protein Sequence	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV26b	60..270 3..213	199/211 (94%) 199/211 (94%)

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

Table 26C. Protein Sequence Properties NOV26a	
PSort analysis:	0.8524 probability located in mitochondrial inner membrane; 0.6000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome); 0.2622 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 26D. Geneseq Results for NOV26a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU83655	Human PRO protein, Seq ID No 128 - Homo sapiens, 473 aa. [WO200208288-A2, 31-JAN-2002]	3..237 22..283	79/264 (29%) 121/264 (44%)	1e-18
AAB49891	Human PRO526 protein sequence - Homo sapiens, 473 aa. [WO200070050-A1, 23-NOV-2000]	3..237 22..283	79/264 (29%) 121/264 (44%)	1e-18
AAB50908	Human PRO526 protein - Homo sapiens, 473 aa. [WO200073452-A2, 07-DEC-2000]	3..237 22..283	79/264 (29%) 121/264 (44%)	1e-18
AAU04589	Human Nogo receptor - Homo sapiens, 473 aa. [WO200151520-A2, 19-JUL-2001]	3..237 22..283	79/264 (29%) 121/264 (44%)	1e-18
AAU12362	Human PRO526 polypeptide sequence - Homo sapiens, 473 aa. [WO200140466-A2, 07-JUN-2001]	3..237 22..283	79/264 (29%) 121/264 (44%)	1e-18

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
Q9BGY6	HYPOTHETICAL 56.5 KDA PROTEIN - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 510 aa.	1..504 1..504	478/504 (94%) 481/504 (94%)	0.0
Q961X3	GH01279P - Drosophila melanogaster (Fruit fly), 615 aa.	45..233 313..522	68/212 (32%) 103/212 (48%)	9e-21
Q9N0E3	UNNAMED PROTEIN PRODUCT - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 473 aa.	3..237 22..283	82/264 (31%) 125/264 (47%)	3e-19
Q9VZ84	CG7509 PROTEIN - Drosophila melanogaster (Fruit fly), 633 aa.	45..233 313..540	69/230 (30%) 103/230 (44%)	1e-18
Q9BZR6	NOGO RECEPTOR - Homo sapiens (Human), 473 aa.	3..237 22..283	79/264 (29%) 121/264 (44%)	3e-18

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
LRR	58..81	7/25 (28%) 19/25 (76%)	0.3
LRR	108..131	10/25 (40%) 19/25 (76%)	0.11
LRR	132..155	8/25 (32%) 18/25 (72%)	0.7
LRR	158..181	11/25 (44%) 19/25 (76%)	0.00021
LRR	182..204	10/25 (40%) 18/25 (72%)	0.093
LRRCT	214..270	15/63 (24%) 42/63 (67%)	0.046

Example 27.

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

Table 27A. NOV27 Sequence Analysis			
	SEQ ID NO: 73		2358 bp
NOV27a, CG106942-01 DNA Sequence	GACTTCCTGGCTCGCCAGCCCCTTCCTTCCGGAGCCTGACCCGGGCCCCGGGCGACCTC CCCGCGCGCTTCCCGGCCGCTGCCAGGGGGTAGAGCGGGCGCAGCCGATCACTACCT GACGGCCTTTTGGCGGCCTGGCCGGGCTGTGCAGGGTGGTAGGGCAAGACGCGCGGC TCCCAATTCTCCCCGGCGCCTTCGCCGGCCCCGGGCTTCTCGCGCTCCGCTCCGGGCT GCACCGAGTTGGGCCGGCGCGCCGCGTTGGTGTTCGCCGCGGGCGGCAGCTCAGAGTC TCCAGGTTGGGGCGGGCCTGGGCCGCACGGCTCCTCCACCCAGGTGACGCTGAGCAGG CTCAGGGTGAAGCCCAGGGAGATGCCACGGCCACGGGCCCTGCGGGCCGCAGCACCG ACAGCAGCAGCGATGCCCGCATGGCGCCCCGGACCGCGGGTCCCCGGCCCCGGCGAACC CCCAGAGCAGCCAGAGGAGTCTCCGAGGGGGCGGGACCGGGGAGGGGGCGGATCCGGA GGGCTCGGGCCCCGCGGGCGGGCCCCGCTCCCTCCCCGCAGAGCAGAGCCAGCGGCCCG AGCCGAATCCCCGGAGCCGCGCCTCGATTCCCCTCCAGCAGCTGCTCTGGGCTGCGCA GGGTTCTTGCGCTCGGCACTGGAGCCTCAGCCGCGGGCCGCAGCTGTCCGACGTGTAC TGCAAGGGCCCCGCCCCGGGGTGGGGTCTCGGGCTCTCGCTACCGGAGAGGGAGGAG AAGGGGGAGGTTAAAGGGGAAGGACCCCCGGAAGTGCCCCCTCCTCAGTGCGGGAGAG GGAGACGCCGGGGGCGGAGTCCCCTGCCTCCCGCGGCGTGGTTGGTGCGTCCCATGTG ACGTCAGAAGCAGCCCCGCCCTGCCTGGATGGTGCGCCCTGAGTGACGTCAGGAGCAG AGGCCGGAGCTGTCCATCAGCACCAAAGGCCGCGGGCGGGCTCAGGGCATGGGGCCGC GGTTCTGGGGCGGCCCCGAGCCCCGGCTCCTGCGCCTTCCCCTTCCTCAGGCCCAGCCC GAGTTCCCGGACGCCGCGGGACTGGAGTGCCAGCCGGTGTGGACGTGGAGCGGCGCC GCCACCGCGCCGACACCATTTCTCTCCGGCCCAGCAGCCCCCTTCCTCGCACGACGGAC TTTCCCTGGACCCCAGCACTATGCCGGGGACTGTGGCAACACTGCGGTTCCAGCTGCT GCCCCCTGAGCCAGATGATGCCTTCTGGGGTGCACCTTGTGAACAGCCCCCTGGAGCGC AGGTACCAGGCACTGCCGGCCCTCGTCTGCATCATGTGCTGTTTGTGGAGTCGTCT ACTGCTTCTTCGGTTACCGCTGCTTCAAGGCAGTGCTCTTTCTCACTGGGTTGCTGTT TGGCTCGGTGGTCATCTTCCTCCTCTGCTACCGAGAGCGGGTGCTAGAGACACAGCTG AGTGCTGGGGCGAGCGCGGGCATCGCTCTGGGCATCGGGCTGCTCTGCGGGCTGGTGG CCATGCTAGTGCGCAGCGTGGGCCTCTTCTGGTGGGGCTGCTGCTCGGCCTGCTGCT CGCAGCTGCTGCCCTGCTGGGCTCCGCACCCTACTACCAGCCAGGCTCCGTGTGGGGT CCACTGGGGCTGTTGCTGGGGGGCGGCCTGCTCTGTGCCCTGCTCACTCTGCGCTGGC CCCGCCCCTCACCACCCTGGCCACCGCCGTGACTGGTGCTGCGCTGATCGCCACTGC CGCTGACTACTTCGCCGAGCTGCTACTGCTGGGGCGCTACGTGGTGGAGCGACTCCGG GCTGCTCCTGTGCCCCCACTCTGCTGGCGAAGCTGGGCCCTGCTGGCACTCTGGCCCC TGCTCAGCCTGATGGGCGTTCTGGTGCAGTGGAGGGTGACAGCTGAGGGGGACTCCCA CACGGAAGTGGTCATCAGCCGGCAGCGCCGACGCGTGCAACTGATGCGGATTCGGCAG CAGGAAGATCGCAAGGAGAAAAGGCGGAAAAGAGACCTCCTCGGGCTCCCCTCAGAG GTCCCCGGGCTCCTCCCAGGCCTGGGCCACCAGATCCTGCTTATCGGCGCAGGCCAGT GCCCATCAAACGCTTCAATGGAGACGTCTCTCCCCGAGCTATATCCAGAGCTTCCGA GACCGGCAGACCGGGAGCTCCCTGAGCTCCTTCATGGCCTCACCCACAGATGCGGACT ATGAGTATGGGTCCCGGGGACCTCTGACAGCCTGCTCAGGCCCCCAGTGCGGGTATA GCCATATCTGTCTGTCTAGACTCTGCAGTCACCAGCTCTGACAGCTCGAGGAGGCCGG TAGGCTGCAATCAGCTTCCGGTTTGGTGGTCCTTCCCA		
	ORF Start: ATG at 977		ORF Stop: TAG at 2261
	SEQ ID NO: 74		428 aa MW at 46672.9kD
NOV27a, CG106942-01 Protein Sequence	MGPRFWGGPSPGSCAFPLRPSPSSRTPRDWSASRCWTWSGAATAPTPFSPAQQPPSS HDGLSLDPSTMPGTVALTRFQLLPPEPDFAFWGAPCEQPLERRYQALPALVCIMCCLF GVVYCFEGYRCFKAVLFLTGLLFGSVVIFLLCYRERVLETQLSAGASAGIALGIGLLC GLVAMLVRSVGLFLVGLLLGLLLAAALLGSAPYYQPGSVWGPLGLLLGGGLLCALLT		

	LRWPRPLTTLATAVTGAALIATAADYFAELLLLGRYVVERLRAAPVPPPLCWRSWALLA LWPLLSLMGVLVQWRVTAEGDSHTEVVISRQRRRVQLMRIRQQEDRKEKRRKKRPPRA PLRGPRAPPRPGPPDPAYRRRPVPIKRFNGDVLSPSYIQSFRDRQTGSSLSSFMASPT DADYEYGSRGPLTACSGPPVRV
--	---

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

Table 27B. Protein Sequence Properties NOV27a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.2400 probability located in nucleus
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV27a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 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
AAM34044	Peptide #8081 encoded by probe for measuring placental gene expression - Homo sapiens, 124 aa. [WO200157272-A2, 09-AUG-2001]	79..188 19..124	60/111 (54%) 76/111 (68%)	2e-25
AAM20130	Peptide #6564 encoded by probe for measuring cervical gene expression - Homo sapiens, 124 aa. [WO200157278-A2, 09-AUG-2001]	79..188 19..124	60/111 (54%) 76/111 (68%)	2e-25
AAM73861	Human bone marrow expressed probe encoded protein SEQ ID NO: 34167 - Homo sapiens, 124 aa. [WO200157276-A2, 09-AUG-2001]	79..188 19..124	60/111 (54%) 76/111 (68%)	2e-25
AAM61147	Human brain expressed single exon probe encoded protein SEQ ID NO: 33252 - Homo sapiens, 124 aa. [WO200157275-A2, 09-AUG-2001]	79..188 19..124	60/111 (54%) 76/111 (68%)	2e-25
ABB24734	Protein #6733 encoded by probe for measuring heart cell gene expression - Homo sapiens, 124 aa. [WO200157274-A2, 09-AUG-2001]	79..188 19..124	60/111 (54%) 76/111 (68%)	2e-25

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
Q9VWD2	CG14234 PROTEIN - Drosophila melanogaster (Fruit fly), 381 aa.	103..392 43..329	82/299 (27%) 143/299 (47%)	3e-20
Q9CRG1	2010003B14RIK PROTEIN - Mus musculus (Mouse), 556 aa (fragment).	112..305 286..490	53/208 (25%) 95/208 (45%)	6e-07
Q9NS93	SEVEN TRANSMEMBRANE PROTEIN TM7SF3 - Homo sapiens (Human), 570 aa.	115..305 303..504	50/205 (24%) 88/205 (42%)	4e-05
Q9NUS4	CDNA FLJ11169 FIS, CLONE PLACE1007282 - Homo sapiens (Human), 570 aa.	115..305 303..504	50/205 (24%) 88/205 (42%)	4e-05
O28838	Hypothetical protein AF1434 - Archaeoglobus fulgidus, 199 aa.	107..304 14..188	51/201 (25%) 82/201 (40%)	7e-04

Example 28.

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

Table 28A. NOV28 Sequence Analysis		
	SEQ ID NO: 75	2177 bp
NOV28a, CG107513-01 DNA Sequence	ATTCCCCCTTGCCGACCCACATACACCATGAAGAGGTGCAGATCGGACGAGCTGCAGCA ACAACAGGGCGAGGAGGATGGAGCTGGGCTGGAAGATGCCGCTTCCCACCTGCCGGGC GCGGACCTCCGGCCTGGGGAGACCACGGGTGCTAACTCTGCTGGCGGGCCAACTTCAG ACGCCGGCGCTGCCGCGGCCCAACCCAGGTCCCCGAAGCAAGCCTCCTGATTTTAA GAAAATCCAGCAGCTGTCAGAGGGCTCCATGTTTGGCCACGGTCTGAAGCACCTGTTTC CACAGCCGCCGTCGGTCTCGGGAAAGGGAGCACCAGACGTCTCAGGATTCCCAGCAGC ATCAGCAGCAGCAGGGTATGTCCGACCATGACTCCCCAGATGAGAAGGAGCGCTCTCC GGAGATGCATCGCGTCTCCTACGCCATGTCCCTGCACGACCTGCCCGCCCGGCCACC GCCTTCAACCGCGTGCTGCAGCAGATCCGCTCCCGGCCCTCCATCAAGCGGGGCGCCA GCCTGCACAGCAGCAGTGGGGGCGGCAGCAGCGGGAGCAGCAGCCGGCGCACCAAGAG TAGCTCCCTGGAGCCCCAGCGTGGCAGCCCTCACCTGCTGCGCAAGGCCCCCCAGGAC AGCAGCCTGGCCGCCATCCTGCACCAGCACCAGTGCCGTCCCCGCTCTTCTCTCCACCA CCGACACTGCTCTGCTGCTGGCCGACGGCAGCAACGTGTACCTCCTGGCTGAGGAGGC CGAAGGCATCGGGGACAAGGTGGATAAGGGAGACCTGGTGGCCCTGAGCCTCCCCGCC GGCCATGGTGACACCGACGGCCCCATCAGCCTGGACGTGCCCGATGGGGCACCGGACC CCCAGCGGACCAAGGCCGCCATTGACCACCTGCACCAGAAGATCCTGAAGATCACCGA GCAGATCAAGATTGAGCAGGAGGCTCGCGACGACAATGTGGCAGAGTATCTGAAACTG GCCAACAACGCGGACAAGCAGCAGGTGTCACGCATCAAGCAAGTGTTTCGAGAAGAAGA	

	ACCAGAAGTCAGCCCAGACCATCGCCCAGCTGCACAAGAAGCTGGAGCACTACCGCCG GCGCCTGAAGGAGATTGAGCAGAACGGGCCCTCGCGGCAGCCCAAGGACGTGCTGCGG GACATGCAGCAGGGGCTGAAGGACGTGGGCGCCAACGTGCGCGCAGGCATCAGCGGCT TTGGGGGCGGCGTGGTGGAGGGCGTCAAGGGCAGCCTCTCTGGCCTCTCACAGGCCAC CCACACCGCCGTGGTGTCCAAGCCCCGGGAGTTTGCCAGCCTCATCCGCAACAAGTTT GGCAGTGCTGACAACATCGCCCACCTGAAGGACCCCCTGGAAGATGGGCCCCCTGAGG AGGCAGCCCCGGGCACTGAGCGGCAGTGCCACACTCGTCTCCAGCCCCAAGTATGGCAG CGATGATGAGTGCTCCAGCGCCACGCTCAGCTCAGCCGGGGCAGGCAGCAACTCTGGG GCTGGGCCTGGTGGGGCGCTGGGGAGCCCTAAGTCCAATGCACTGTATGGTGCTCCTG GAAACCTGGATGCTCTGCTGGAAGAGCTACGGGAGATCAAGGAGGGACAGTCTCACCT GGAGGACTCCATGGAAGACCTGAAGACTCAGCTGCAGAGGGACTACACCTACATGACC CAGTGCCTGCAGGAGGAGCGCTACAGGTACGAGCGGCTGGAGGAGCAGCTCAACGACC TGACTGAGCTTCATCAGAACGAGATGACGAACCTGAAGCAGGAGCTGGCCAGCATGGA GGAGAAGGTGGCCTACCAGTCCTATGAGAGGGCACGGGACATCCAGGAGGCCGTGGAG TCCTGCCTGACCCGGGTACCAAGCTGGAGCTGCAGCAGCAACAGCAGCAGGTGGTAC AGCTGGAGGGCGTGGAGAATGCCAACGCGCGGGCGCTGCTGGGCAAGTTCATCAACGT GATCCTGGCGCTCATGGCCGTGCTGCTGGTGTTCGTGTCCACCATCGCCAACCTTCATC ACGCCCCTCATGAAGACACGCCTGCGCATCACCAGCACCACCCTCCTGGTCCTCGTCC TGTTCTCTCTCTGGAAGCACTGGGACTCCCTCACCTACCTCCTGGAGCACGTGTTGCT GCCCAGCTGAGTGGCCAGCCACACCAACCCT		
	ORF Start: ATG at 27	ORF Stop: TGA at 2154	
	SEQ ID NO: 76	709 aa	MW at 77503.9kD
NOV28a, CG107513-01 Protein Sequence	MKRCRSDELQQQQGEEDGAGLEDAASHLPGADLRPGETTGANSAGGPTSDAGAAAAPN PGPRSKPPDLKKIQQLSEGSFMFGHGLKHLFHSRRRSREREHQTSQDSQQHQQQQGMSD HDSPDEKERSPEMHRVSYAMSLHDLPARPTAFNRVLQQIRSRPSIKRGASLHSSSGGG SSGSSSRRTKSSSLEPQRGSPHLLRKAPQDSSLAAILHQHQCRPRSSSTTDTALLLAD GSNVYLLAEEAEGIGDKVDKGDVLVALSLPAGHGDTDGPISLDVPDGPDPQRTKAAID HLHQKILKITEQIKIEQEARDNVAEYLKLANNADKQQVSRKQVFEKKNQKSAQTIA QLHKKLEHYRRRLKEIEQNGPSRQPKDVLDRMQQLKDVGANVRAGISGFGGGVVEGV KGSLSGLSQATHAVVSKPREFASLIRNKFGSADNIAHLKDPLEDGPPEEAARALSGS ATLVSSPKYGSDDDECSSATLSSAGAGSNSGAGPGGALGSPKSNALYGAPGNLDALLEE LREIKEGQSHLEDSMEDLKTQLQRDYTYMTQCLQEERYRYERLEEQLNDLTELHQNEM TNLKQELASMEEKVAYQSYERARDIQEAVESCLTRVTKLELQQQQQQQVVQLEGVENAN ARALLGKFINVILALMAVLLVFVSTIANFITPLMKTRLRITSTLLVLVLFLWLKHWL SLTYLLEHVLLPS		

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

Table 28B. Protein Sequence Properties NOV28a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 28C. Geneseq Results for NOV28a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AA Y94907	Human secreted protein clone cal06_19x protein sequence SEQ ID NO:20 - Homo sapiens, 653 aa. [WO200009552-A1, 24-FEB-2000]	57..703 13..646	359/672 (53%) 438/672 (64%)	e-178
AAM78708	Human protein SEQ ID NO 1370 - Homo sapiens, 477 aa. [WO200157190-A2, 09-AUG-2001]	249..708 26..476	258/466 (55%) 326/466 (69%)	e-134
AAU28090	Novel human secretory protein, Seq ID No 259 - Homo sapiens, 446 aa. [WO200166689-A2, 13-SEP-2001]	258..708 4..445	254/457 (55%) 320/457 (69%)	e-132
AAM40705	Human polypeptide SEQ ID NO 5636 - Homo sapiens, 369 aa. [WO200153312-A1, 26-JUL-2001]	352..703 13..362	224/355 (63%) 267/355 (75%)	e-117
AAM38919	Human polypeptide SEQ ID NO 2064 - Homo sapiens, 331 aa. [WO200153312-A1, 26-JUL-2001]	379..703 2..324	209/328 (63%) 248/328 (74%)	e-108

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 28D.

Table 28D. Public BLASTP Results for NOV28a				
Protein Accession Number	Protein/Organism/Length	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O75069	Hypothetical protein KIAA0481 (Cerebral protein-11) (hucep-11) - Homo sapiens (Human), 650 aa (fragment).	69..709 10..650	638/641 (99%) 640/641 (99%)	0.0
Q9ULS5	Hypothetical protein KIAA1145 - Homo sapiens (Human), 467 aa (fragment).	250..708 17..466	258/465 (55%) 325/465 (69%)	e-134
AAH26867	SIMILAR TO KIAA1145 PROTEIN - Mus musculus (Mouse), 446 aa.	258..708 4..445	249/457 (54%) 316/457 (68%)	e-129
O94876	Hypothetical protein KIAA0779 - Homo sapiens (Human), 320 aa (fragment).	393..703 1..313	202/314 (64%) 241/314 (76%)	e-105
Q9VI21	CG1021 PROTEIN - Drosophila melanogaster (Fruit fly), 638 aa.	288..675 252..638	188/409 (45%) 248/409 (59%)	1e-84

Example 29.

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

Table 29A. NOV29 Sequence Analysis			
	SEQ ID NO: 77	664 bp	
NOV29a, CG107533-02 DNA Sequence	ATCGCCCCCTCCTGCGCTAGCGGAGGTGATCGCCGCGGCGATGCCGGAGGAGGGTTTCGG GCTGCTCGGTGCGGCGCAGGCCCTATGGGTGCGTCTTGCCTGCGGGCTGCTTTGGTCCCAT GGTCGCGGGGCTTGGTGATCTGCCTCGTGGTGTGCATCCAGCGCTTCGCACAGGCTCAG CAGCAGCTGCCGCTCGAGTCACTTGGGGACCTCAGCAGGACCCCAGGCTATACTGGCA GGGGGGCCCAGCACTGGGCCGCTCCTTCCTGCATGGACCAGAGCTGGACAAGGGGCAG CTACGTATCCATCGTGATGGCATCTACATGGTACACATCCAGGTGACGCTGGCCATCT GCTCCTCCACGACGGCCTCCAGGCACCACCCACCACCTGGCCGTGGGAATCTGCTC TCCCGCCTCCCGTAGCATCAGCCTGCTGCGTCTCAGCTTCCACCAAGGTTGTACCATT GCCTCCAGCGCCTGACGCCCTGGCCCGAGGGGACACACTCTGCACCAACCTCACTG GGACACTTTTGCCTTCCCGAAACACTGATGAGACCTTCTTTGGAGTGCAGTGGGTGCG CCCCTGACCACTGCTGCTGATTAGGGTTTTTTAAATTTTATTTTATTTTATTTAAGTT CAAGAGAAAAAGTGTACACACAGGGG		
	ORF Start: ATG at 40	ORF Stop: TGA at 334	
	SEQ ID NO: 78	98 aa	MW at 10705.2kD
NOV29a,	MPEEGSGCSVRRRPYGCVLRAALVPLVAGLVICLVVCIQRFAQAQQQLPLESLGDLR TPGYTGRGAQHWAAPSCMDQSWTRGSYVSIVMASTWYTSR		

CG107533-02 Protein Sequence	
---------------------------------	--

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

Table 29B. Protein Sequence Properties NOV29a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 45 and 46

5 A search of the NOV29a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 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
AAR50121	CD27L - Homo sapiens, 193 aa. [WO9405691-A, 17-MAR-1994]	1..54 1..54	54/54 (100%) 54/54 (100%)	7e-25
AAW41180	CD27 ligand - Homo sapiens, 216 aa. [US5716805-A, 10-FEB-1998]	39..54 62..77	16/16 (100%) 16/16 (100%)	0.16
AAR50122	sCD27L-3 - Homo sapiens, 216 aa. [WO9405691-A, 17-MAR-1994]	39..54 62..77	16/16 (100%) 16/16 (100%)	0.16
AAR53971	CD27-L type II transmembrane protein - Mammalia, 216 aa. [WO9410308-A, 11-MAY-1994]	39..54 62..77	16/16 (100%) 16/16 (100%)	0.16
AAG26041	Zea mays protein fragment SEQ ID NO: 30347 - Zea mays subsp. mays, 172 aa. [EP1033405-A2, 06-SEP-2000]	24..72 6..59	19/55 (34%) 26/55 (46%)	2.4

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
Q96J57	TUMOR NECROSIS FACTOR (LIGAND) SUPERFAMILY, MEMBER 7 - Homo sapiens (Human), 193 aa.	1..54 1..54	54/54 (100%) 54/54 (100%)	2e-24
P32970	CD27 ligand (CD27-L) (CD70 antigen) - Homo sapiens (Human), 193 aa.	1..54 1..54	54/54 (100%) 54/54 (100%)	2e-24
Q9KFY7	HYPOTHETICAL PROTEIN BH0329 - Bacillus halodurans, 423 aa.	39..98 240..300	18/61 (29%) 28/61 (45%)	7.4
Q9RC64	UNKNOWN - Bacillus halodurans, 262 aa.	39..98 79..139	18/61 (29%) 28/61 (45%)	7.4
O34255	PURL PROTEIN - Wolinella succinogenes, 331 aa (fragment).	70..93 8..31	11/24 (45%) 15/24 (61%)	9.7

Example 30.

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

Table 30A. NOV30 Sequence Analysis		
	SEQ ID NO: 79	2840 bp
NOV30a, CG107562-01 DNA Sequence	CGGCAGTGGCAGGAGCCGCCTTTCCGATTCCCTACGATGCGGGTGCTGAGCTATGGCA AAGGGCAGCGAAGTGACGAGCGAGACCCGCGTACGACTGTGAAAGCCACCTGGAGCCA CCTTGCCGGGATTGTACCTGCAGGCAGAAAGTCTTCCTACGACCGTCTTTTCCCTTAG AGGCACCAGAATCCCTGTAACCATTCATCCAGGTGTTGAGAAGATATGTAGCAGCCGA GCACCCATCTTTTGACACCGTCCTCTGAAATCAGCTTTGGAGATGCTTTCACCTCTGTC CGTCTTCTGCAGCAGCCAGGCAGAGTGCCGACTCCTTCACAGCCGTGAGGAACTCTTC AGGCTCCAGAAGCTCTTAAACCTGATCTACAATGGAAAAAATTCTTTTTTATCTGTTT CTCATTGGCATAGCAGTGAAAGCTCAGATCTGTCCAAAGCGTTGTGTCTGTCAGATTT TGTCTCCTAATCTTGCAACCCCTTGTGCCAAGAAAGGGCTTTTATTTGTTCCACCAA CATTGACAGAAGAACTGTGGAAGTGCAGTGGCAGACAATTTTGTTACAAATATTAAA AGGAAAGATTTTGCCAATATGACCAGCTTGGTGGACCTGACTCTATCCAGGAATACAA TAAGTTTTATTACACCTCATGCTTTCGCTGACCTACGAAATTTGAGGGCTTTGCATTT GAATAGCAACAGATTGACTAAAATTACAAATGATATGTTTCAGTGGTCTTTCCAATCTT CATCATTTGATACTGAACAACAATCAGCTGACTTTAATTTCTCTACAGCGTTTGATG ATGTCTTCGCCCTTGAGGAGCTGGATCTGTCTATAATAATCTAGAAACCATTCCTTG GGATGCTGTTGAGAAGATGGTTAGCTTGCATACCCTTAGTTTGGATCACAATATGATT GATAACATTCCTAAGGGGACCTTCTCCCATTTGCACAAGATGACTCGGTTAGATGTGA CATCAAATAAATTGCAGAAGCTACCACCTGACCCTCTCTTTCAGCGAGCTCAGGTACT	

	AGCAACCTCAGGAATCATAAGCCCATCTACTTTTGCATTAAGTTTGGTGGAAACCCC TTGCATTGCAATTGTGAATTGTTGTGGTTGAGGCGTCTGTCCAGAGAAGATGACTTAG AGACCTGTGCTTCTCCTCCACTTTTAACTGGCCGCTACTTTTGGTCAATTCCTGAAGA AGAGTTTTTGTGTGAGCCTCCTCTCATTAAGTCTCGTCATACACATGAGATGAGAGTCCTG GAGGGACAAAGGGCAACACTGAGGTGCAAAGCCAGGGGAGACCCTGAGCCTGCAATTC ACTGGATTTCTCCTGAAGGGAAGCTTATTTCAAATGCAACAAGATCTCTGGTGTATGA TAACGGAACACTTGACATTCTTATCACAAGTGTAAAGGATACAGGTGCTTTTACCTGC ATTGCTTCCAATCCTGCTGGGGAAGCAACACAAATAGTGGATCTTCATATAATTAAGC TCCCTCACTTACTAAATAGTACAAACCATATCCATGAGCCTGATCCTGGTTCTTCAGA TATCTCAACTTCTACCAAGTCAGGTTCTAATACAAGCAGTAGTAATGGTGATACTAAA TTGAGTCAAGATAAAATTGTGGTGGCAGAAGCTACATCATCAACGGCACTACTTAAAT TTAATTTTCAAAGAAATATCCCTGGAATACGTATGTTTCAAATCCAGTACAATGGTAC TTATGATGACACCCTTGTTTACAGGATGATACCTCCTACGAGCAAACTTTTCTGGTC AATAATCTGGCTGCTGGAACATATGTATGACTTGTGTGTCTTGGCCATATATGATGATG GCATCACTTCCCTCACTGCCACAAGAGTCGTGGGTTCATCCAGTTTACTACGGAACA GGATTATGTGCGTTGCCATTTTCATGCAGTCCCAGTTTTTGGGAGGCACCATGATTATT ATTATTGGTGGAAATCATTGTAGCATCTGTGCTGGTATTCATCATTATTCTGATGATCC GGTATAAGGTTTGCAACAATAATGGGCAACACAAGGTCACCAAGGTTAGCAATGTTTA TTCCCAAACCTAACGGGGCTCAAATACAAGGCTGTAGTGTAACGCTGCCCCAGTCCGTG TCCAAACAAGCTGTGGGACACGAAGAGAATGCCAGTGTTGTAAAGCTACCAGTGACA ATGTGATTCAATCTTCAGAACTTGTTCGAGTCAAGACTCCTCTACCACTACCTCTGC TTTGCCCTCCTTCTGACTTCAAGCACTTCTGTGTCCCAAAGCAGAAAAGAAAGACT GGCACAAGCCAAAGTACAGAACCACAGAATGAAGCCGTCACAAATGTTGAATCCCAA ACACTAACAGGAACAACCTCAACTGCCTTGCAGTTAGCTAGCCGTCCTCCCGATTCTGT CACAGAGGGGGCCACGTCTAAAAGAGCACATATAAAGCCAAGTAAGTTTATCACTTTG CCTGCTGAGAGATCCGGAGCAAGGCACAAGTACTCCCTCAATGGAGAATTAAAGGAAT ACTATTGTTATATTAACCTCGCCGAACACATGTGGACTGTTTCCTAAAAGAAGCATGTC TATGAATGTGATGTTTATTTCAGTCTGACTGTTCTGATGGTCATAGTGGAAGGCAACT CTCAAATTCTGAGGGACTACTGGAAAGCTCTGTGTAATTTATAATTTCTTTTTCATGA AAAATCATTTTGAGAACTCACATAGAAGATTGGAATTTGCAATTCCAATGCTGTGTAT AAATCAACCTTCTCAGATGCTTTGCTGACTAATGTTGACCAGATTGTCCAGGAAAC		
	ORF Start: ATG at 380	ORF Stop: TGA at 2678	
	SEQ ID NO: 80	766 aa	MW at 84690.6kD
NOV30a, CG107562-01 Protein Sequence	MEKILFYLFILIGIAVKAQICPKRCVCQILSPNLATLCAKKGLLFVPPNIDRRTVELRL ADNFVTNIKRKDFANMTSLVDLTLNRNTISFITPHAFADLRNLRALHLNSNRLTKITN DMFSGLSNLHLILNNNQLTLISSTAFDDVFALEELDLSYNNLETIPWDAVEKMOVSLH TSLDHNMIDNIPKGTFSHLHKMTRLVDVTSNKLOKLPPDPLFQRAQVLATSGIISPST FALSFGGNPLHCNCELLWLRRLSREDDLETASPPLLTGRYFWSIPEEEFLCEPPLIT RHTHEMRVLEGQRATLRCKARGDPEPAIHWISPEGKLISNATRSVLYDNGTLDILITT VKDTGAFTCIASNPAGEATQIVDLHIIKLPHLLNSTNHIHEPDGSSDISTSTKSGSN TSSSNGDTKLSQDKIVVAEATSSTALLKFNFQRNIPGIRMFQIQYNGTYDDTLVYRMI PPTSKTFLVNNLAAGTMYDLCVLAIYDDGITSLATRVVGC IQFTTEQDYVRCHF MQS QFLGGTMIIIIIGGIIIVASVLVFIILMIRYKVCNNNGQHKVTKVSNVYSQTNGAQIQG CSVTL PQSVSKQAVGHEENAQCCKATSDNVIQSSETCSSQDSSTTTTSALPPSWTSSTS VSQKQKRKTGTPSTEPQNEAVTNVESQNTNRNNSTALQLASRPPDSVTEGPTSKRAH IKPSKFITLPAERSGARHKYSLNGELKEYCYINS PNTCGLFPKRSMNMVFIQSDC SDGHSGKATLKF		
	SEQ ID NO: 81	2388 bp	
NOV30b, CG107562-02 DNA Sequence	GCTCTTAAACCTGATCTACAATGGAAAAAATTCTTTTTTATCTGTTTCTCATTGGCAT AGCAGTGAAAGCTCAGATCTGTCCAAAGCGTTGTGTCTGTCTCAGATTTTGTCTCCTAAT CTTGCAACCCTTTGTGCCAAGAAAGGGCTTTTATTTGTTCCACCAAACATTGACAGAA GAACTGTGGAAGTGCAGTTGGCAGACAATTTTGTACAAATATTAAGGAAAGATTT TGCCAATATGACCAGCTTGGTGGACCTGACTCTATCCAGGAATACAATAAGTTTATT ACACCTCATGCTTTCGCTGACCTACGAAATTTGAGGGCTTTGCATTTGAATAGCAACA GATTGACTAAAATTACAAATGATATGTTTCAGTGGTCTTTCCAATCTTCATCATTGAT ACTGAACAACAATCAGCTGACTTTAATTTCTCTACAGCGTTTGATGATGTCTTCGCC CTTGAGGAGCTGGATCTGTCCCTATAATAATCTAGAAACCATTCTTGGGATGCTGTTG AGAAGATGGTTAGCTTGCATACCCTTAGTTTGGATCACAATATGATTGATAACATTCC		

	TAAGGGGACCTTCTCCCATTTGCACAAGATGACTCGGTAGATGTGACATCAAATAAA TTGCAGAAGCTACCACCTGACCCTCTCTTTTCAGCGAGCTCAGGTACTAGCAACCTCAG GAATCATAAGCCCATCTACTTTTGCATTAAAGTTTGGTGGAAACCCCTTGCAATTGCAA TTGTGAATTGTTGTGGTTGAGGCGTCTGTCCAGAGAAGATGACTTAGAGACCTGTGCT TCTCCTCCACTTTTAACTGGCCGCTACTTTTGGTCAATTCCTGAAGAAGAGTTTTTGT GTGAGCCTCCTCTCATTACTCGTCATACACATGAGATGAGAGTCCTGGAGGGACAAAG GGCAACACTGAGGTGCAAAGCCAGGGGAGACCCTGAGCCTGCAATTCCTGGAATTTCT CCTGAAGGGAAGCTTATTTCAAATGCAACAAGATCTCTGGTGTATGATAACGGAACAC TTGACATTCTTATCACAACTGTAAAGGATACAGGTGCTTTTACCTGCATTGCTTCCAA TCCTGCTGGGGAAGCAACACAAATAGTGGATCTTCATATAATTAAGCTCCCTCACTTA CTAAATAGTACAAACCATATCCATGAGCCTGATCCTGGTTCCTTCAGATATCTCAACTT CTACCAAGTCAGGTTCTAATAACAAGCAGTAGTAATGGTGATACTAAATTGAGTCAAGA TAAAATTGTGGTGGCAGAAGCTACATCATCAACGGCACTACTTAACTTTACTTTTCAA AGAACTATCCCTGGAATACGTATGTTTCAAATCCAGTACAATGGTACTTATGATGACA CCCTTGTTTACAGGATGATACCTCCTACGAGCAAACTTTTCTGGTCAATAATCTGGC TGCTGGAACATATGTATGACTTGTGTGTCTTGGCCATATATGATGATGGCATCACTTCC CTCACTGCCACAAGAGTCGTGGGTTGCATCCAGTTTACTACGGAACAGGATTATGTGC GTTGCCATTTTCATGCAGTCCCAGTTTTTGGGAGGCACCATGATTATTATTATTGGTGG AATCATTGTAGCATCTGTGCTGGTATTCATCATTATTCTGATGATCCGGTATAAGGTT TGCAACAATAATGGGCAACACAAGGTCACCAAGGTTAGCAATGTTTATTCCCAAACCTA ACGGGGCTCAAATACAAGGCTGTAGTGTAACGCTGCCCCAGTCCGTGTCCAAACAAGC TGTGGGACACGAAGAGATTGCCAGTGTTGTAAAGCTACCAGTGACAATGTGATTCAA TCTTCAGAACTTGTTTCGAGTCAGGACTCCTCTACCACTACCTCTGCTTTGCCCTCCTT CCTGGACTTCAAGCACTTCTGTGTCCCAAAGCAGAAAAGAAAGACTGGCACAAAGCC AAGTACAGAACCACAGAATGAAGCCGTCACAAATGTTGAATCCCAAACACTAACAGG AACAACTCAACTGCCTTGCAGTTAGCTAGCCGTCTCCCGATTCTGTACAGAGGGGC CCACGTCTAAAAGAGCACATATAAAGCCAAATGCTTTGCTGACTAATGTTGACCAGAT TGTCAGAGGAAACACAGAGGCTGGAGTTAATCTGAAGAGCACCCTTCTCCTCTCTCTC CTGAAAAAATTTGCCACTGATATTTTACTGGATAAAATTCAAAAATGTTTCAATTCA CAAAGGCTAATTGTTGAAGTGGTGTCTGTAAGAAATTTGTCTACAGGAGCCAAGGTGA AAGTCTCTGATGACGGCGGAAGTGGCTCCATTAGACCATGGTTCATCCTCTTTTAAAA ACAAATTTTT		
	ORF Start: ATG at 21	ORF Stop: TGA at 2178	
	SEQ ID NO: 82	719 aa	MW at 79402.7kD
NOV30b, CG107562-02 Protein Sequence	MEKILFYLFLLIGIAVKAQICPKRCVCQILSPNLATLCAKKGLLFVPPNIDRRRTVELRL ADNFVTNIKRKDFANMTSLVDLTLSRNTISFITPHAFADLRNLRLHLNSNRLTKITN DMFSGLSNLHHLILNNNQLTLISSTAFDDVFAL EELDSLNNLETIPWDAVEKMOVSLH TSLSDHNMIDNIPKGTFSHLHKMTRLDVTSNKLQKLPPDPLFQRAQVLATSGIISPST FALSFGGNPLHCNCELLWLRLSREDDLETCASPPLLTGRYFWSIPEEEFLCEPPLIT RHTHEMRVLEGQRATLRCKARGDPEPAIHWISPEGKLISNATRSVYDNGTLDILITT VKDTGAFTCIASNPAGEATQIVDLHI IKLPHLLNSTNHIHEPDGSSDISTSTKSGSN TSSSNGDTKLSQDKIVVAEATSSTALLNFTFQRTIPGIRMFQIQYNGTYDDTLVYRMI PPTSKTFLVNNLAAGTMYDLCLVLAIDYDGGITSLTATRVVGC IQFTTEQDYVRCHFMQS QFLGGTMII IGGIIVASVLVFI IILMIRYKVCNNNGQHKVTKVSNVYSQTNGAQIQG CSVTLPQSVSKQAVGHEEIAQCCKATSDNVIQSSETCSSQDSSTTTTSALPPSWTSSTS VSQKQKRKTGTPSTEPQNEAVTNVESQNTNRNNSTALQLASRPDSVTEGPTSKRAH IKPNALLTNVDQIVQETQRLELI		
	SEQ ID NO: 83	1545 bp	
NOV30c, 210086373 DNA Sequence	GGATCCCAGATCTGTCCAAAGCGTTGTGTCTGTCAGATTTTGTCTCCTAATCTTGCAA CCCTTTGTGCCAAGAAAGGGCTTTTATTTGTTCCACCAAACATTGACAGAAGAACTGT GGAAGTGCAGTTGGCAGACAATTTTGTACAAATATTTAAAGGAAAGATTTTGCCAAT ATGACCAGCTTGGTGGACCTGACTCTATCCAGGAATACAATAAGTTTTATTACACCTC ATGCTTTTCGCTGACCTACGAAATTTGAGGGCTTTGCATTTGAATAGCAACAGATTGAC TAAAATTACAAATGATATGTTTCAAGTGGTCTTTCCAATCTTCATCATTTGATACTGAAC AACAAATCAGCTGACTTTAATTTCTCTACAGCGTTTGATGATGTCTTCGCCCTTGAGG AGCTGGATCTGTCTATAATAATCTAGAAACCATTCTTGGGATGCTGTTGAGAAGAT GGTTAGCTTGCATACCCTTAGTTTGGATCACAATATGATTGATAACATTCCTAAGGGG ACCTTCTCCCATTTGCACAAGATGACTCGGTAGATGTGACATCAAATAAATTGCAGA		

	AGCTACCACCTGACCCTCTCTTTTCAGCGAGCTCAGGTACTAGCAACCTCAGGAATCAT AAGCCCATCTACTTTTGCATTAAGTTTTGGTGGAAACCCCTTGCATTGCAATTGTGAA TTGTTGTGGTTGAGGCGTCTGTCCAGAGAAGATGACTTAGAGACCTGTGCTTCTCCTC CACTTTTAACTGGCCGCTACTTTTGGTCAATTCCCTGAAGAAGAGTTTTTGTGTGAGCC TCCTCTCATTACTCGTCATACACATGAGATGAGAGTCCTGGAGGGACAAAGGGCAACA CTGAGGTGCAAAGCCAGGGGAGACCCTGAGCCTGCAATTCACTGGATTTCTCCTGAAG GGAAGCTTATTTCAAATGCAACAAGATCTCTGGTGTATGATAACGGAACACTTGACAT TCTTATCACAACGTGTAAGGATACAGGTGCTTTTACCTGCATTGCTTCCAATCCTGCT GGGGAAGCAACACAAATAGTGGATCTTCATATAATTAAGCTCCCTCACTTACTAAATA GTACAAACCATATCCATGAGCCTGATCCTGGTTCTTCAGATATCTCAACTTCTACCAA GTCAGGTTCTAATACAAGCAGTAGTAATGGTGATACTAAATTGAGTCAAGATAAAATT GTGGTGGCAGAAGCTACATCATCAACGGCACTACTTAAATTTAATTTTCGAAGAAATA TCCCTGGAATACGTATGTTTCAAATCCAGTACAATGGTACTTATGATGACACCCTTGT TTACAGAATGATACCTCCTACGAGCAAACTTTTCTGGTCAATAATCTGGCTGCTGGA ACTATGTATGACTTGTGTGTCTTGGCCATATATGATGATGGCATCACTTCCCTCACTG CCACAAGAGTCGTGGGTTGCATCCAGTTTACTACGGAACAGGATTATGTGCGTTGCCA TTTCATGCAGTCCCAGTTTTTGGGAGGCACCCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 84	515 aa	MW at 57372.8kD
NOV30c, 210086373 Protein Sequence	GSQICPKRCVCQILSPNLATLCAKKGLLFVPPNIDRRTVELRLADNFVTNIKRKDFAN MTSLVDLTLRNTISFITPHAFADLRNLRLHLNSNRLTKITNDMFSGLSNLHHLILN NNQLTLISSTAFDDVFALEELDSLNNLETIPWDAVEKMVSLHTLSLDHNMIDNIPKG TFSHLHKMTRLDVDTSNKLQKLPPDPLFQRAQVLATSGIISPSTFALSFGGNPLHCNCE LLWLRRLSREDDLETASPPLLTGRYFWSIPEEEFLCEPPLITRHTHEMRVLEGQRAT LRCKARGDPEPAIHWISPEGKLI SNATRSLVYDNGTLDILITTVKDTGAFTCIASNPA GEATQIVDLHI IKLPHLLNSTNHIHEPDGSSDISTSTKSGSNTSSSNGDTKLSQDKI VVAEATSSSTALLKFNFRRNIPGIRMFQIQYNGTYDDTLVYRMI PPTSKTFLVNNLAAG TMYDLCVLAIYDDGITSLTATRVVGC IQFTTEQDYVRCHFMQSQFLGGTLE		
	SEQ ID NO: 85	1545 bp	
NOV30d, 210086403 DNA Sequence	GGATCCCAGATCTGTCCAAAGCGTTGTGTCTGTCAGATTTTGTCTCCTAATCTTGCAA CCCTTTGTGCCAAGAAAGGGCTTTTATTTGTTCCACCAACATTGACAGAAGAACTGT GGAACCTGCGGTTGGCAGACAATTTTGTACAAATATTTAAAGGAAAGATTTTGCCAAT ATGACCAGCTTGGTGGACCTGACTCTATCCAGGAATACAATAAGTTTTATTACACCTC ATGCTTTCGCTGACCTACGAAATTTGAGGGCTTTGCATTTGAATAGCAACAGATTGAC TAAATTACAAATGATATGTTTCAGTGGTCTTTCCAATCTTCATCATTTGATACTGAAC ACAATCAGCTGACTTTAATTTCTCTACAGCGTTTGATGATGTCTTCACCCTTGAGG AGCTGGATCTGTCTTATAATAATCTAGAAACCATTCTTGGGATGCTGTTGAGAAGAT GGTTAGCTTGCATACCCTTAGTTTGGATCACAATATGATTGATAACATTCCTAAGGGG ACCTTCTCCCATTTGCACAAGATGACTCGGTTAGATGTGACATCAAATAAATTGCAGA AGCTACCACCTGACCCTCTCTTTCAGCGAGCTCAGGTACTAGCAACCTCAGGAATCAT AAGCCCATCTACTTTTGCATTAAGTTTTGGTGGAAACCCCTTGCATTGCAATTGTGAA TTGTTGTGGTTGAGGCGTCTGTCCAGAGAAGATGACTTAGAGACCTGTGCTTCTCCTC CACTTTTAACTGGCCGCTACTTTTGGTCAATTCCCTGAAGAAGAGTTTTTGTGTGAGCC TCCTCTCATTACTCGTCATACACATGAGATGAGAGTCCTGGAGGGACAAAGGGCAACA CTGAGGTGCAAAGCCAGGGGAGACCCTGAGCCTGCAATTCACTGGATTTCTCCTGAAG GGAAGCTTATTTCAAATGCAACAAGATCTCTGGTGTATGATAACGGAACACTTGACAT TCTTATCACAACGTGTAAGGATACAGGTGCTTTTACCTGCATTGCTTCCAATCCTGCT GGGGAAGCAACACAAATAGTGGATCTTCATATAATTAAGCTCCCTCACTTACTAAATA GTACAAACCATATCCATGAGCCTGATCCTGGTTCTTCAGATATCTCAACTTCTACCAA GTCAGGTTCTAATACAAGCAGTAGTAATGGTGATACTAAATTGAGTCAAGATAAAATT GTGGTGGCAGAAGCTACATCATCAACGGCACTACTTAAATTTAATTTTCAAAGAAATA TCCCTGGAATACGTATGTTTCAAATCCAGTACAATGGTACTTATGATGACACCCTTGT TTACAGAATGATACCTCCTACGAGCAAACTTTTCTGGTCAATAATCTGGCTGCTGGA ACTATGTATGACTTGTGTGTCTTGGCCATATATGATGATGGCATCACTTCCCTCACTG CCACAAGAGTCGTGGGTTGCATCCAGTTTACTACGGAACAGGATTATGTGCGTTGCCA TTTCATGCAGTCCCAGTTTTTGGGAGGCACCCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	

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 190

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 190

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

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

CLAIMS

What is claimed is:

1. An isolated polypeptide comprising the mature form of an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 73.
2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 73.
3. An isolated polypeptide comprising an amino acid sequence which is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 73.
4. An isolated polypeptide, wherein the polypeptide comprises an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 73.
5. The polypeptide of claim 1 wherein said polypeptide is naturally occurring.
6. A composition comprising the polypeptide of claim 1 and a carrier.
7. A kit comprising, in one or more containers, the composition of claim 6.
8. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein the therapeutic comprises the polypeptide of claim 1.
9. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

- (a) providing said sample;
- (b) introducing said sample to an antibody that binds immunospecifically to polypeptide; and
- (c) determining the presence or amount of antibody bound to said polypeptide, thereby determining the presence or amount of polypeptide in said sample.

10. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

- a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and
- b) comparing the expression of said polypeptide in the sample of step (a) to the expression of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,

wherein an alteration in the level of expression of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

11. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:

- (a) introducing said polypeptide to said agent; and
- (b) determining whether said agent binds to said polypeptide.

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

13. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:

- (a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;
- (b) contacting the cell with a composition comprising a candidate substance; and

- (c) determining whether the substance alters the property or function ascribable to the polypeptide;

whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition in the absence of the substance, the substance is identified as a potential therapeutic agent.

14. A method for screening for a modulator of activity of or of latency or predisposition to a pathology associated with the polypeptide of claim 1, said method comprising:
 - (a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein said test animal recombinantly expresses the polypeptide of claim 1;
 - (b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a); and
 - (c) comparing the activity of said polypeptide in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator activity of or latency or predisposition to, a pathology associated with the polypeptide of claim 1.
15. The method of claim 14, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.
16. A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of claim 1 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.
17. A method of treating or preventing a pathology associated with the polypeptide of claim 1, the method comprising administering the polypeptide of claim 1 to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

18. The method of claim 17, wherein the subject is a human.
19. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 73 or a biologically active fragment thereof.
20. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 73.
21. The nucleic acid molecule of claim 20, wherein the nucleic acid molecule is naturally occurring.
22. A nucleic acid molecule, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 73.
23. An isolated nucleic acid molecule encoding the mature form of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 73.
24. An isolated nucleic acid molecule comprising a nucleic acid selected from the group consisting of 2n-1, wherein n is an integer between 1 and 73.
25. The nucleic acid molecule of claim 20, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 73, or a complement of said nucleotide sequence.
26. A vector comprising the nucleic acid molecule of claim 20.

27. The vector of claim 26, further comprising a promoter operably linked to said nucleic acid molecule.
28. A cell comprising the vector of claim 26.
29. An antibody that immunospecifically binds to the polypeptide of claim 1.
30. The antibody of claim 29, wherein the antibody is a monoclonal antibody.
31. The antibody of claim 29, wherein the antibody is a humanized antibody.
32. The antibody of claim 29, wherein the antibody is a fully human antibody.
33. The antibody of claim 29, wherein the dissociation constant for the binding of the polypeptide to the antibody is less than 1×10^{-9} M.
34. The antibody of claim 29, wherein the antibody neutralizes an activity of the polypeptide.
35. A method of treating or preventing a NOVX-associated disorder, the method comprising administering to a subject in which such treatment or prevention is desired the antibody of claim 29 in an amount sufficient to treat or prevent the pathology in the subject.
36. The method of claim 35, wherein the subject is human.
37. A method for determining the presence or amount of the nucleic acid molecule of claim 20 in a sample, the method comprising:
 - (a) providing said sample;
 - (b) introducing said sample to a probe that binds to said nucleic acid molecule; and
 - (c) determining the presence or amount of said probe bound to said nucleic acid molecule,thereby determining the presence or amount of the nucleic acid molecule in said sample.

38. The method of claim 37 wherein presence or amount of the nucleic acid molecule used as a marker for cell or tissue type.
39. The method of claim 38 wherein the cell or tissue type is cancerous.
40. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the nucleic acid molecule of claim 20 in a first mammalian subject, the method comprising:
- a) measuring the level of expression of the nucleic acid in a sample from the first mammalian subject; and
 - b) comparing the level of expression of said nucleic acid in the sample of step (a) to the level of expression of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of expression of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

41. A method of producing the polypeptide of claim 1, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 73.
42. The method of claim 41 wherein the cell is a bacterial cell.
43. The method of claim 41 wherein the cell is an insect cell.
44. The method of claim 41 wherein the cell is a yeast cell.
45. The method of claim 41 wherein the cell is a mammalian cell.
46. A method of producing the polypeptide of claim 2, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell

comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 73.

47. The method of claim 46 wherein the cell is a bacterial cell.
48. The method of claim 46 wherein the cell is an insect cell.
49. The method of claim 46 wherein the cell is a yeast cell.
50. The method of claim 46 wherein the cell is a mammalian cell.