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

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2442739 A1 2002/10/17

(21) 2 442 739

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2002/04/03
(87) Date publication PCT/PCT Publication Date: 2002/10/17
(85) Entrée phase nationale/National Entry: 2003/09/30
(86) N° demande PCT/PCT Application No.: US 2002/010366
(87) N° publication PCT/PCT Publication No.: 2002/081625
(30) Priorités/Priorities: 2001/04/03 (60/281,086) US;
2001/04/05 (60/281,906) US; 2001/04/06 (60/282,020) US;
2001/04/10 (60/282,930) US; 2001/04/12 (60/283,512) US;
2001/04/12 (60/283,444) US; 2001/04/13 (60/283,657) US;
2001/04/13 (60/283,678) US; 2001/04/13 (60/283,710) US;
2001/04/17 (60/284,234) US; 2001/04/19 (60/285,325) US;
2001/04/20 (60/285,381) US; 2001/04/24 (60/286,068) US;
...

(51) Cl.Int.^{7}/Int.Cl.^{7} A61K 38/16

(71) Demandeur/Applicant:
CURAGEN CORPORATION, US

(72) Inventeurs/Inventors:
PADIGARU, MURALIDHARA, US;
SHENOY, SURESH G., US;
KEKUDA, RAMESH, US;
RASTELLI, LUCA, US;
MEZES, PETER S., US;
SMITHSON, GLENNDA, US;
GUO, XIAOJIA, US;
...

(74) Agent: GOUDREAU GAGE DUBUC

(54) Titre : NOUVEAUX ANTICORPS SE LIANT A DES POLYPEPTIDES ANTIGENIQUES, ACIDES NUCLEIQUES CODANT POUR CES ANTIGENES, ET PROCEDES D'UTILISATION DE CEUX-CI

(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.

(21) 2 442 739

(13) A1

(30) Priorités(suite)/Priorities(continued): 2001/04/25 (60/286,292) US; 2001/06/07 (60/296,692) US; 2001/06/26 (60/300,883) US; 2001/08/08 (60/311,003) US; 2001/08/13 (60/311,973) US; 2001/08/16 (60/312,901) US; 2001/09/14 (60/322,283) US; 2001/10/05 (60/327,448) US; 2001/12/31 (60/345,734) US; 2002/01/03 (60/345,755) US; 2002/02/04 (60/354,391) US; 2002/04/02 (10/114,153) US

(72) Inventeurs(suite)/Inventors(continued): GERLACH, VALERIE, US; CASMAN, STACIE J., US; BOLDOG, FERENC L., US; LI, LI, US; ZERHUSEN, BRYAN D., US; TCHERNEV, VELIZAR T., US; GANGOLLI, ESHA A., US; VERNET, CORINE A. M., US; SPYTEK, KIMBERLY A., US; MALYANKAR, URIEL M., US; PATTURAJAN, MEERA, US; MILLER, CHARLES E., US; TAUPIER, RAYMOND J., JR., US; HEYES, MELVYN P., US; JU, JINGFANG, US; PEYMAN, JOHN A., US; CATTERTON, ELINA, US; MACDOUGALL, JOHN R., US; EDINGER, SHLOMIT R., US; STONE, DAVID J., US; MAZUR, ANN, US

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 October 2002 (17.10.2002)

PCT

(10) International Publication Number
WO 02/081625 A3

(51) International Patent Classification[7]: A61K 38/16

(21) International Application Number: PCT/US02/10366

(22) International Filing Date: 3 April 2002 (03.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/281,086	3 April 2001 (03.04.2001)	US
60/281,906	5 April 2001 (05.04.2001)	US
60/282,020	6 April 2001 (06.04.2001)	US
60/282,930	10 April 2001 (10.04.2001)	US
60/283,512	12 April 2001 (12.04.2001)	US
60/283,444	12 April 2001 (12.04.2001)	US
60/283,657	13 April 2001 (13.04.2001)	US
60/283,710	13 April 2001 (13.04.2001)	US
60/283,678	13 April 2001 (13.04.2001)	US
60/284,234	17 April 2001 (17.04.2001)	US
60/285,325	19 April 2001 (19.04.2001)	US
60/285,381	20 April 2001 (20.04.2001)	US
60/286,068	24 April 2001 (24.04.2001)	US
60/286,292	25 April 2001 (25.04.2001)	US
60/296,692	7 June 2001 (07.06.2001)	US
60/300,883	26 June 2001 (26.06.2001)	US
60/311,003	8 August 2001 (08.08.2001)	US
60/311,973	13 August 2001 (13.08.2001)	US
60/312,901	16 August 2001 (16.08.2001)	US
60/322,283	14 September 2001 (14.09.2001)	US
60/327,448	5 October 2001 (05.10.2001)	US
60/345,734	31 December 2001 (31.12.2001)	US
60/345,755	3 January 2002 (03.01.2002)	US
60/354,391	4 February 2002 (04.02.2002)	US
10/114,153	2 April 2002 (02.04.2002)	US

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

US	60/281,086 (CIP)
Filed on	3 April 2001 (03.04.2001)
US	60/281,906 (CIP)
Filed on	5 April 2001 (05.04.2001)
US	60/282,020 (CIP)
Filed on	6 April 2001 (06.04.2001)
US	60/282,930 (CIP)
Filed on	10 April 2001 (10.04.2001)
US	60/283,512 (CIP)
Filed on	12 April 2001 (12.04.2001)
US	60/283,444 (CIP)
Filed on	12 April 2001 (12.04.2001)
US	60/283,657 (CIP)
Filed on	13 April 2001 (13.04.2001)
US	60/283,710 (CIP)
Filed on	13 April 2001 (13.04.2001)
US	60/283,678 (CIP)
Filed on	13 April 2001 (13.04.2001)
US	60/284,234 (CIP)
Filed on	17 April 2001 (17.04.2001)
US	60/285,325 (CIP)
Filed on	19 April 2001 (19.04.2001)
US	60/285,381 (CIP)
Filed on	20 April 2001 (20.04.2001)
US	60/286,068 (CIP)
Filed on	24 April 2001 (24.04.2001)
US	60/286,292 (CIP)
Filed on	25 April 2001 (25.04.2001)
US	60/296,692 (CIP)
Filed on	7 June 2001 (07.06.2001)
US	60/300,883 (CIP)
Filed on	26 June 2001 (26.06.2001)
US	60/311,003 (CIP)
Filed on	8 August 2001 (08.08.2001)
US	60/311,973 (CIP)
Filed on	13 August 2001 (13.08.2001)
US	60/312,901 (CIP)
Filed on	16 August 2001 (16.08.2001)
US	60/322,283 (CIP)
Filed on	14 September 2001 (14.09.2001)
US	60/327,448 (CIP)
Filed on	5 October 2001 (05.10.2001)
US	60/345,755 (CIP)
Filed on	3 January 2002 (03.01.2002)
US	60/354,391 (CIP)
Filed on	2 February 2002 (02.02.2002)
US	Not furnished (CIP)
Filed on	31 December 2001 (31.12.2001)
US	Not furnished (CIP)
Filed on	2 April 2002 (02.04.2002)

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

[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 02/081625 A3

WO 02/081625 A3

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **PADIGARU, Muralidhara** [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). **SHENOY, Suresh, G.** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **KEKUDA, Ramesh** [IN/US]; 1213 Avalon Valley Drive, Danbury, CT 06810 (US). **RASTELLI, Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). **MEZES, Peter, S.** [CA/US]; 7 Clark's lane, Old Lyme, CT 06371 (US). **SMITHSON, Glennda** [US/US]; 125 Michael Drive, Guildford, CT 06435 (US). **GUO, Xiaojia** [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). **GERLACH, Valerie** [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). **CASMAN, Stacie, J.** [US/US]; 17 Peck Street, North Haven, CT 06473 (US). **BOLDOG, Ferenc, L.** [HU/US]; 1687 Hartford Turnpike, North Haven, CT 06473 (US). **LI, Li** [CN/US]; 56 Jerimoth Drive, Branford, CT 06405 (US). **ZERHUSEN, Bryan, D.** [US/US]; 337 Monticello Drive, Branford, CT 06405 (US). **TCHERNEV, Velizar, T.** [BG/US]; 45 Jefferson Road #3-12, Branford, CT 06405 (US). **GANGOLLI, Esha, A.** [IN/US]; 31 Strawberry Hill Road, Madison, CT 06443 (US). **VERNET, Corine, A., M.** [FR/US]; 1739 Foxon Road, Apartment L6, Branford, CT 06471 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street, Number 1, New Haven, CT 06511 (US). **MALYANKAR, Uriel, M.** [IN/US]; 229 Branford Road, Number 330, Branford, CT 06405 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). **MILLER, Charles, E.** [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **HEYES, Melvyn, P.** [GB/US]; 183 Townsend Avenue, Number 3, New Haven, CT 06512 (US). **JU, Jingfang** [US/US]; 391 Rosebud Lane, Orange, CT 06477 (US). **PEYMAN, John, A.** [US/US]; 336 West Rock Avenue, New Haven, CT 06515 (US). **CATTERTON, Elina** [FI/US]; 584 Boston Post Road, Madison, CT 06443 (US). **MACDOUGALL, John, R.** [CA/US]; 117 Russell Street, Hamden, CT 06517 (US). **EDINGER, Shlomit, R.** [US/US]; 766 Edgewood Avenue, New Haven, CT 06515 (US). **STONE, David, J.** [US/US]; 223 Whitethorn Drive, Guilford, CT 06437 (US). **MAZUR, Ann** [US/US]; 35 Burr Road, Bloomfield, CT 06002 (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:
22 May 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	2
CONTENANT LES PAGES	1	À	229

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

JUMBO APPLICATIONS/PATENTS

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

THIS IS VOLUME	1	OF	2
CONTAINING PAGES	1	TO	229

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

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

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

FIELD OF THE INVENTION

The present invention relates to novel antibodies that bind immunospecifically to 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 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.

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 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 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 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 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.

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 46, 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 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.

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

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

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

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

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

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1a	CG56258-01	1	2	Sodium/Calcium Exchanger
1b	CG56258-02	3	4	Sodium/Calcium Exchanger
1c	248057963	5	6	Sodium/Calcium Exchanger
2a	CG59843-01	7	8	Fibropellin III
3a	CG59845-01	9	10	Butyrophilin
4a	CG59871-01	11	12	CVB3 Binding Protein
5a	CG59883-01	13	14	CVB3 Binding Protein
6a	CG59901-01	15	16	Scavenger receptor
7a	CG88748-01	17	18	Cyclic Nucleotide-gated Channel Protein
8a	CG90021-01	19	20	Testicular Metalloprotease (Disintegrin)
9a	CG90709-01	21	22	Ion Transport Protein
9b	CG90709-02	23	24	Ion Transport Protein
9c	CG90709-03	25	26	Ion Transport Protein
9d	CG90709-04	27	28	Ion Transport Protein
10a	CG90739-01	29	30	Neuronal Thread Protein
10b	172390256	31	32	Neuronal Thread Protein
10c	172390440	33	34	Neuronal Thread Protein
10d	172390569	35	36	Neuronal Thread Protein
10e	172390587	37	38	Neuronal Thread Protein
10f	172390603	39	40	Neuronal Thread Protein
10g	172390624	41	42	Neuronal Thread Protein
10h	172390644	43	44	Neuronal Thread Protein
11a	CG91667-01	45	46	Delta-like Homology (dlk1)
11b	CG91667-02	47	48	Delta-like Homology (dlk1)
12a	CG92293-01	49	50	Polyprotein (ovochymase)
12b	CG92293-02	51	52	Polyprotein (ovochymase)
13a	CG92384-01	53	54	Long type PB-Cadherin
14a	CG92455-01	55	56	IGFBP
15a	CG92531-01	57	58	Leucine Rich
16a	CG92715-01	59	60	KIAA0918
16b	CG92715-02	61	62	Leucine Rich Repeat
17a	CG92813-01	63	64	Cadherin Related Tumor Suppressor Precursor
18a	CG92844-01	65	66	Thyroid Hormone Induced Protein B Precursor
18b	174308357	67	68	Thyroid Hormone Induced Protein B Precursor
19a	CG93088-01	69	70	Monocarboxylate Transporter
20a	CG93335-01	71	72	Putative Type II Membrane
21a	CG93345-01	73	74	GPCR
22a	CG93400-01	75	76	GPCR
23a	CG93410-01	77	78	Glutamate Receptor 5 Precursor
23b	188822752	79	80	Glutamate Receptor 5 Precursor
24a	CG93722-01	81	82	Hepsin
25a	CG93858-01	83	84	Fibullin
25b	CG93858-02	85	86	Fibullin
25c	CG56914-03	87	88	Fibullin
26a	CG93871-01	89	90	Fibullin

27a	CG93884-01	91	92	Monocyte Inhibitory Receptor

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 B. 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 46; (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 46, 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 46; (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 46

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 46; (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 46 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 46; (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 46, 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 46 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 46; (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 46 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 46; 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 101 is changed from that selected from the

group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX 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 acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

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

or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "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 1-46, 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 1-46, 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 1-46, 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 1-46, 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 1-46, is one that is sufficiently complementary to the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, 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 1-46, 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 1-46, 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 1-46; or an anti-sense strand nucleotide sequence of SEQ ID NO:2*n*-1,

wherein *n* is an integer between 1-46; or of a naturally occurring mutant of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46.

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. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of a NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, that encodes a polypeptide having a NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, 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 1-46. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46.

In addition to the human NOVX nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to

nucleic acid molecules comprising an open reading frame (ORF) encoding a NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from any one of the human SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 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 probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase “stringent hybridization conditions” refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at

which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60 °C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65 °C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50 °C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to any one of the sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, 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 1-46, 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 1-46, 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/volt) 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 1-46, 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 1-46. 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 1-46, 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 1-46. 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 1-46; more preferably at least about 70% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1-46; still more preferably at least about

80% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1-46; even more preferably at least about 90% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1-46; and most preferably at least about 95% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1-46.

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 1-46, 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 1-46, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into any of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

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

one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

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

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

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46, or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

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

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

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein (*e.g.*, by

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

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

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having

specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX cDNA disclosed herein (*i.e.*, any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

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

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

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

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the

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

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.,* Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res*. 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46, while still encoding a protein

that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

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

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

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

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or

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

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of a NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

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

In an embodiment, the NOVX protein has an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1-46, and retains the functional activity of the protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46, 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 1-46, and retains the functional activity of the NOVX proteins of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or

nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See,* Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46.

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

Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46, whereas a "non-NOVX polypeptide" refers to a polypeptide having an

amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically-active portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at least two biologically-active portions of a NOVX protein. In yet another embodiment, a NOVX fusion protein comprises at least three biologically-active portions of a NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

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

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

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

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different

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

NOVX Agonists and Antagonists

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

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

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

Polypeptide Libraries

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

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

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

NOVX Antibodies

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

An isolated protein of the invention intended to serve as an antigen, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46, 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 polyppeptide or a fragment thereof comprises at least one antigenic epitope. An anti-NOVX antibody of the present invention is said to specifically bind to antigen NOVX when the equilibrium binding constant (K_D) is ≤1 μM, preferably ≤ 100 nM, more preferably ≤ 10 nM, and most preferably ≤ 100 pM to about 1 pM, as measured by assays 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).

Monoclonal Antibodies

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

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

specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

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

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

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

identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

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

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

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

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric

immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

Human Antibodies

Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991);

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

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent

rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

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

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

F_{ab} Fragments and Single Chain Antibodies

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

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the

binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

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

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

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

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (*e.g.* $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies

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

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

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

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

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

Heteroconjugate Antibodies

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

Effector Function Engineering

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

can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

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

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

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

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

Immunoliposomes

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

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

Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention

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 ^{3}H.

Antibody Therapeutics

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

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

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with

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

Pharmaceutical Compositions of Antibodies

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

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

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

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

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

ELISA Assay

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

plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of an analyte mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of an analyte genomic DNA include Southern hybridizations. Procedures for conducting immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulus, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Thory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, *in vivo* techniques for detection of an analyte protein include introducing into a subject a labeled anti-an analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

NOVX Recombinant Expression Vectors and Host Cells

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

to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

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

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

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

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either

fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

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

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

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

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

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

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

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type

specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

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

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

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

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

introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

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

Transgenic NOVX Animals

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

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences, *i.e.*, any one of SEQ ID NO:2*n*-1, wherein *n* is an integer

between 1-46, 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 1-46), 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 1-46, can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX

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

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

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

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use

of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

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

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

preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

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

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

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

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

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

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

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

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

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

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

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

Screening and Detection Methods

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

detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

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

Screening Assays

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

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

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

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

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

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

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX

target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention. In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

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

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

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

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

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

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants.

Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

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

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

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

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

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

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

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) map

their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue typing); and (*iii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

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

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g.,* D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day

using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

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

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

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

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

sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

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

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

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If coding sequences, such as those of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

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

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

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

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

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the

biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test

subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

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

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

Prognostic Assays

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

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder

associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

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

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating

nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

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

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

array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.,* Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

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

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to

an exemplary embodiment, a probe based on a NOVX sequence, *e.g.*, a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79.

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

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

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

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

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

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

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

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.*, NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (The disorders include 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.) In conjunction with such

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

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

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome pregnancy zone protein precursor enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a

metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

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

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of

expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

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

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include 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 Ostoeodystrophy, 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 Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

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

Prophylactic Methods

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

Therapeutic Methods

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

NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situ*ations in which NOVX is abnormally downregulated and/or in which increased NOVX activity 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).

Determination of the Biological Effect of the Therapeutic

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

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

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.

As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders,

Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

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

EXAMPLES

Example A: Polynucleotide and Polypeptide Sequences, and Homology Data

Example 1.

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

Table 1A. NOV1 Sequence Analysis

	SEQ ID NO: 1	2813 bp
NOV1a, CG56258-01 DNA Sequence	TCTCGTGT**ATG**GCGTGGTTAAGGTTGCAGCCTCTCACCTCTGCCTTCCTCCATTTTGG GCTGGTTACCTTTGTGCTCTTCCTGAATGGTCTTCGAGCAGAGGCTGGTGGCTCAGGG GACGTGCCAAGCACAGGGCAGAACAATGAGTCCTGTTCAGGGTCATCGGACTGCAAGG AGGGTGTCATCCTGCCAATCTGGTACCCGGAGAACCCTTCCCTTGGGGACAAGATTGC CAGGGTCATTGTCTATTTTGTGGCCCTGATATACATGTTCCTTGGGGTGTCCATCATT GCTGACCGCTTCATGGCATCTATTGAAGTCATCACCTCTCAAGAGAGGGAGGTGACAA TTAAGAAACCCAATGGAGAAACCAGCACAACCACTATTCGGGTCTGGAATGAAACTGT CTCCAACCTGACCCTTATGGCCCTGGGTTCCTCTGCTCCTGAGATACTCCTCTCTTTA ATTGAGGTGTGTGGTCATGGGTTCATTGCTGGTGATCTGGGACCTTCTACCATTGTAG GGAGTGCAGCCTTCAACATGTTCATCATCATTGGCATCTGTGTCTACGTGATCCCAGA CGGAGAGACTCGCAAGATCAAGCATCTACGAGTCTTCTTCATCACCGCTGCTTGGAGT ATCTTTGCCTACATCTGGCTCTATATGATTCTGGCAGTCTTCTCCCCTGGTGTGGTCC AGGTTTGGGAAGGCCTCCTCACTCTCTTCTTCTTTCCAGTGTGTGTCCTTCTGGCCTG GGTGGCAGATAAACGACTGCTCTTCTACAAATACATGCACAAAAAGTACCGCACAGAC AAACACCGAGGAATTATCATAGAGACAGAGGGTGACCACCCTAAGGGCATTGAGATGG ATGGGAAAATGATGAATTCCCATTTTCTAGATGGGAACCTGGTGCCCCTGGAAGGGAA GGAAGTGGATGAGTCCCGCAGAGAGATGATCCGGATTCTCAAGGATCTGAAGCAAAAA CACCCAGAGAAGGACTTAGATCAGCTGGTGGAGATGGCCAATTACTATGCTCTTTCCC ACCAACAGAAGAGCCGTGCCTTCTACCGTATCCAAGCCACTCGTATGATGACTGGTGC AGGCAATATCCTGAAGAAACATGCAGCAGAACAAGCCAAGAAGGCCTCCAGCATGAGC GAGGTGCACACCGATGAGCCTGAGGACTTTATTTCCAAGGTCTTCTTTGACCCATGTT CTTACCAGTGCCTGGAGAACTGTGGGGCTGTACTCCTGACAGTGGTGAGGAAAGGGGG AGACATGTCAAAGACCATGTATGTGGACTACAAAACAGAGGATGGTTCTGCCAATGCA GGGGCTGACTATGAGTTCACAGAGGGCACGGTGGTTCTGAAGCCAGGAGAGACCCAGA AGGAGTTCTCCGTGGGCATAATTGATGACGACATTTTTGAGGAGGATGAACACTTCTT TGTAAGGTTGAGCAATGTCCGCATAGAGGAGGAGCAGCCAGAGGAGGGGATGCCTCCA GCAATATTCAACAGTCTTCCCTTGCCTCGGGCTGTCCTAGCCTCCCCTTGTGTGGCCA CAGTTACCATCTTGGATGATGACCATGCAGGCATCTTCACTTTTGAATGTGATACTAT TCATGTCAGTGAGAGTATTGGTGTTATGGAGGTCAAGGTTCTGCGGACATCAGGTGCC CGGGGTACAGTCATCGTCCCCTTTAGGACAGTAGAAGGGACAGCCAAGGGTGGCGGTG AGGACTTTGAAGACACATATGGGGAGTTGGAATTCAAGAATGATGAAACTGTGAAAAC TCTTCAGGTGAAGATAGTTGATGACGAGGAATATGAGAAAAAGGATAATTTCTTCATT GAGCTGGGCCAGCCCCAGTGGCTTAAGCGAGGGATTTCAGCTCTGCTACTCAATCAAG GGGATGGGGACAGGAAGCTAACAGCCGAGGAGGAGGAGGCTCGGAGGATAGCAGAGAT GGGCAAGCCAGTTCTTGGGGAGAACTGCCGGCTGGAGGTCATCATCGAGGAGTCATAT GATTTTAAGAACACGGTGGATAAACTCATCAAGAAAACGAACTTGGCCTTGGTAATTG GGACCCATTCATGGAGGGAGCAGTTTTTAGAGGCAATTACGGTGAGCGCAGGGGACGA GGAGGAGGAGGAGGACGGGTCCCGGGAGGAGCGGCTGCCGTCGTGCTTTGACTACGTG ATGCACTTCCTGACGGTGTTCTGGAAGGTGCTCTTCGCCTGTGTGCCCCCCACCGAGT ACTGCCACGGCTGGGCCTGCTTTGGTGTCTCCATCCTGGTCATCGGCCTGCTCACCGC CCTCATTGGGGACCTCGCCTCCCACTTCGGCTGCACCGTTGGCCTCAAGGACTCTGTC AATGCTGTTGTCTTCGTTGCCCTGGGCACCTCCATCCCTGACACGTTCGCCAGCAAGG TGGCGGCGCTGCAGGACCAGTGCGCCGACGCGTCCATCGGCAACGTGACCGGCTCCAA CGCGGTGAACGTGTTCCTTGGCCTGGGCGTCGCCTGGTCTGTGGCCGCCGTGTACTGG GCGGTGCAGGGCCGCCCCTTCGAGGTGCGCACTGGCACGCTGGCCTTCTCCGTCACGC	

	TCTTCACCGTCTTCGCCTTCGTGGGCATTGCCGTGCTGCTGTACCGGCGCCGGCCGCA CATCGGCGGCGAGCTGGGCGGCCCGCGCGGACCCAAGCTCGCCACCACCGCGCTCTTC CTGGGCCTCTGGCTCCTGTACATCCTCTTCGCCAGCCTGGAGGCGTACTGCCACATCC GGGGCTTC**TAG**GGCCTCGCGCAGAGACTC		
	ORF Start: ATG at 9	ORF Stop: TAG at 2793	
	SEQ ID NO: 2	928 aa	MW at 102900.1kD
NOV1a, CG56258-01 Protein Sequence	MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGV ILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKK PNGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSA AFNMFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVW EGLLTLFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGK MMNSHFLDGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQ KSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQ CLENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEF SVGIIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVT ILDDDHAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDF EDTYGELEFKNDETVKTLQVKIVDDEEYEKKDNFFIELGQPQWLKRGISALLLNQGDG DRKLTAEEEEARRIAEMGKPVLGENCRLEVIIEESYDFKNTVDKLIKKTNLALVIGTH SWREQFLEAITVSAGDEEEEEDGSREERLPSCFDYVMHFLTVFWKVLFACVPPTEYCH GWACFGVSILVIGLLTALIGDLASHFGCTVGLKDSVNAVVFVALGTSIPDTFASKVAA LQDQCADASIGNVTGSNAVNVFLGLGVAWSVAAVYWAVQGRPFEVRTGTLAFSVTLFT VFAFVGIAVLLYRRRPHIGGELGGPRGPKLATTALFLGLWLLYILFASLEAYCHIRGF		
	SEQ ID NO: 3	2840 bp	
NOV1b, CG56258-02 DNA Sequence	GTCTCTGGCCTATCAGGAGGACAACTGGTGCTGCAATAGAAGCCAGTGGCTAAGTCTC GTGT**ATG**GCGTGGTTAAGGTTGCAGCCTCTCACCTCTGCCTTCCTCCATTTTGGGCTG GTTACCTTTGTGCTCTTCCTGAATGGTCTTCGAGCAGAGGCTGGTGGCTCAGGGGACG TGCCAAGCACAGGGCAGAACAATGAGTCCTGTTCAGGGTCATCGGACTGCAAGGAGGG TGTCATCCTGCCAATCTGGTACCCGGAGAACCCTTCCCTTGGGGACAAGATTGCCAGG GTCATTGTCTATTTTGTGGCCCTGATATACATGTTCCTTGGGGTGTCCATCATTGCTG ACCGCTTCATGGCATCTATTGAAGTCATCACCTCTCAAGAGAGGGAGGTGACAATTAA GAAACCCAATGGAGAAACCAGCACAACCACTATTCGGGTCTGGAATGAAACTGTCTCC AACCTGACCCTTATGGCCCTGGGTTCCTCTGCTCCTGAGATACTCCTCTCTTTAATTG AGGTGTGTGGTCATGGGTTCATTGCTGGTGATCTGGGACCTTCTACCATTGTAGGGAG TGCAGCCTTCAACATGTTCATCATCATTGGCATCTGTGTCTACGTGATCCCAGACGGA GAGACTCGCAAGATCAAGCATCTACGAGTCTTCTTCATCACCGCTGCTTGGAGTATCT TTGCCTACATCTGGCTCTATATGATTCTGGCAGTCTTCTCCCCTGGTGTGGTCCAGGT TTGGGAAGGCCTCCTCACTCTCTTCTTCTTTCCAGTGTGTGTCCTTCTGGCCTGGGTG GCAGATAAACGACTGCTCTTCTACAAATACATGCACAAAAAGTACCGCACAGACAAAC ACCGAGGAATTATCATAGAGACAGAGGGTGACCACCCTAAGGGCATTGAGATGGATGG GAAAATGATGAATTCCCATTTTCTAGATGGGAACCTGGTGCCCCTGGAAGGGAAGGAA GTGGATGAGTCCCGCAGAGAGATGATCCGGATTCTCAAGGATCTGAAGCAAAAACACC CAGAGAAGGACTTAGATCAGCTGGTGGAGATGGCCAATTACTATGCTCTTTCCCACCA ACAGAAGAGCCGTGCCTTCTACCGTATCCAAGCCACTCGTATGATGACTGGTGCAGGC AATATCCTGAAGAAACATGCAGCAGAACAAGCCAAGAAGGCCTCCAGCATGAGCGAGG TGCACACCGATGAGCCTGAGGACTTTATTTCCAAGGTCTTCTTTGACCCATGTTCTTA CCAGTGCCTGGAGAACTGTGGGGCTGTACTCCTGACAGTGGTGAGGAAAGGGGGAGAC ATGTCAAAGACCATGTATGTGGACTACAAAACAGAGGATGGTTCTGCCAATGCAGGGG CTGACTATGAGTTCACAGAGGGCACGGTGGTTCTGAAGCCAGGAGAGACCCAGAAGGA GTTCTCCGTGGGCATAATTGATGACGACATTTTTGAGGAGGATGAACACTTCTTTGTA AGGTTGAGCAATGTCCGCATAGAGGAGGAGCAGCCAGAGGAGGGGATGCCTCCAGCAA TATTCAACAGTCTTCCCTTGCCTCGGGCTGTCCTAGCCTCCCCTTGTGTGGCCACAGT TACCATCTTGGATGATGACCATGCAGGCATCTTCACTTTTGAATGTGATACTATTCAT GTCAGTGAGAGTATTGGTGTTATGGAGGTCAAGGTTCTGCGGACATCAGGTGCCCGGG GTACAGTCATCGTCCCCTTTAGGACAGTAGAAGGGACAGCCAAGGGTGGCGGTGAGGA CTTTGAAGACACATATGGGGAGTTGGAATTCAAGAATGATGAAACTGTCAAAACAATT CACATCAAGGTAATTGATGATGAGGCATATGAGAAAAACAAGAATTACTTCATTGAGA TGATGGGCCCCCGCATGGTGGATATGAGTTTTCAGAAAGCGCTCCTGTTATCTCCAGA CAGGAAGCTGACTATGGAAGAAGAGGAGGCCAAGAGGATAGCAGAGATGGGAAAGCCA		

	GTATTGGGTGAACACCCCAAACTAGAAGTCATCATTGAAGAGTCCTATGAGTTCAAGA CTACGGTGGACAAACTGATCAAGAAGACAAACCTGGCCTTGGTTGTGGGGACCCATTC CTGGAGGGACCAGTTCATGGAGGCCATCACCGTCAGTGCAGCAGGGGATGAGGATGAG GATGAATCCGGGGAGGAGAGGCTGCCCTCCTGCTTTGACTACGTCATGCACTTCCTGA CTGTCTTCTGGAAGGTGCTGTTTGCCTGTGTGCCCCCCACAGAGTACTGCCACGGCTG GGCCTGCTTCGCCGTCTCCATCCTCATCATTGGCATGCTCACCGCCATCATTGGGGAC CTGGCCTCGCACTTCGGCTGCACCATTGGTCTCAAAGATTCAGTCACAGCTGTTGTTT TCGTGGCATTTGGCACCTCTGTCCCAGATACGTTTGCCAGCAAAGCTGCTGCCCTCCA GGATGTATATGCAGACGCCTCCATTGGCAACGTGACGGGCAGCAACGCCGTCAATGTC TTCCTGGGCATCGGCCTGGCCTGGTCCGTGGCCGCCATCTACTGGGCTCTGCAGGGAC AGGAGTTCCACGTGTCGGCCGGCACACTGGCCTTCTCCGTCACCCTCTTCACCATCTT TGCATTTGTCTGCATCAGCGTGCTCTTGTACCGAAGGCGGCCGCACCTGGGAGGGGAG CTTGGTGGCCCCCGTGGCTGCAAGCTCGCCACAACATGGCTCTTTGTGAGCCTGTGGC TCCTCTACATACTCTTTGCCACACTAGAGGCCTATTGCTACATCAAGGGGTTC**TAA**		
	ORF Start: ATG at 63	ORF Stop: TAA at 2838	
	SEQ ID NO: 4	925 aa	MW at 102802.3kD
NOV1b, CG56258-02 Protein Sequence	MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGV ILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKK PNGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSA AFNMFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVW EGLLTLFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGK MMNSHFLDGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQ KSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQ CLENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEF SVGIIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVT ILDDDHAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDF EDTYGELEFKNDETVKTIHIKVIDDEAYEKNKNYFIEMMGPRMVDMSFQKALLLSPDR KLTMEEEEAKRIAEMGKPVLGEHPKLEVIIEESYEFKTTVDKLIKKTNLALVVGTHSW RDQFMEAITVSAAGDEDEDESGEERLPSCFDYVMHFLTVFWKVLFACVPPTEYCHGWA CFAVSILIIGMLTAIIGDLASHFGCTIGLKDSVTAVVFVAFGTSVPDTFASKAAALQD VYADASIGNVTGSNAVNVFLGIGLAWSVAAIYWALQGQEFHVSAGTLAFSVTLFTIFA FVCISVLLYRRRPHLGGELGGPRGCKLATTWLFVSLWLLYILFATLEAYCYIKGF		
	SEQ ID NO: 5	2685 bp	
NOV1c, 248057963 DNA Sequence	**GGA**TCCGAGGCTGGTGGCTCAGGGGACGTGCCAAGCACAGGGCAGAACAATGAGTCCT GTTCAGGGTCATCGGACTGCAAGGAGGGTGTCATCCTGCCAATCTGGTACCCGGAGAA CCCTTCCCTTGGGGACAAGATTGCCAGGGTCATTGTCTATTTTGTGGCCCTGATATAC ATGTTCCTTGGGGTGTCCATCATTGCTGACCGCTTCATGGCATCTATTGAAGTCATCA CCTCTCAAGAGAGGGAGGTGACAATTAAGAAACCCAATGGAGAAACCAGCACAACCAC TATTCGGGTCTGGAATGAAACTGTCTCCAACCTGACCCTTATGGCCCTGGGTTCCTCT GCTCCTGAGATACTCCTCTCTTTAATTGAGGTGTGTGGTCATGGGTTCATTGCTGGTG ATCTGGGACCTTCTACCATTGTAGGGAGTGCAGCCTTCAACATGTTCATCATCATTGG CATCTGTGTCTACGTGATCCCAGACGGAGAGACTCGCAAGATCAAACATCTACGAGTC TTCTTCATCACCGCTGCTTGGAGTATCTTTGCCTACATCTGGCTCTATATGATTCTGG CAGTCTTCTCCCCTGGTGTGGTCCAGGTTTGGGAAGGCCTCCTCACTCTCTTCTTCTT TCCAGTGTGTGTCCTTCTGGCCTGGGTGGCAGATAAACGACTGCTCTTCTACAAATAC ATGCACAAAAAGTACCGCACAGACAAACACCGAGGAATTATCATAGAGACAGAGGGTG ACCACCCTAAGGGCATTGAGATGGATGGGAAAATGATGAATTCCCATTTTCTAGATGG GAACCTGGTGCCCCTGGAAGGGAAGGAAGTGGATGAGTCCCGCAGAGAGATGATCCGG ATTCTCAAGGATCTGAAGCAAAAACACCCAGAGAAGGACTTAGATCAGCTGGTGGAGA TGGCCAATTACTATGCTCTTTCCCACCAACAGAAGAGCCGCGCCTTCTACCGTATCCA AGCCACTCGTATGATGACTGGTGCAGGCAATATCCTGAAGAAACATGCAGCAGAACAA GCCAAGAAGGCCTCCAGCATGAGCGAGGTGCACACCGATGAGCCTGAGGACTTTATTT CCAAGGTCTTCTTTGACCCATGTTCTTACCAGTGCCTGGAGAACTGTGGGGCTGTACT CCTGACAGTGGTGAGGAAAGGGGGAGACATGTCAAAGACCATGTATGTGGACTACAAA ACAGAGGATGGTTCTGCCAATGCAGGGGCTGACTATGAGTTCACAGAGGGCACGGTGG TTCTGAAGCCAGGAGAGACCCAGAAGGAGTTCTCCGTGGGCATAATTGATGACGACAT TTTTGAGGAGGATGAACACTTCTTTGTAAGGTTGAGCAATGTCCGCATAGAGGAGGAG CAGCCAGAGGAGGGGATGCCTCCAGCAATATTCAACAGTCTTCCCTTGCCTCGGGCTG		

	TCCTAGCCTCCCCTTGTGTGGCCACAGTTACCATCTTGGATGATGACCATGCAGGCAT CTTCACTTTTGAATGTGATACTATTCATGTCAGTGAGAGTATTGGTGTTATGGAGGTC AAGGTTCTGCGGACATCAGGTGCCCGGGGTACAGTCATCGTCCCCTTTAGGACAGTAG AAGGGACAGCCAAGGGTGGCGGTGAGGACTTTGAAGACACATATGGGGAGTTGGAATT CAAGAATGATGAAACTGTGAAAACCATAAGGGTTAAAATAGTAGATGAGGAGGAATAC GAAAGGCAAGAGAATTTCTTCATTGCCCTTGGTGAACCGAAATGGATGGAACGTGGAA TATCAGATGTGACAGACAGGAAGCTGACTATGGAAGAAGAGGAGGCCAAGAGGATAGC AGAGATGGGAAAGCCAGTATTGGGTGAACACCCCAAACTAGAAGTCATCATTGAAGAG TCCTATGAGTTCAAGACTACGGTGGACAAACTGATCAAGAAGACAAACCTGGCCTTGG TTGTGGGGACCCATTCCTGGAGGGACCAGTTCATGGAGGCCATCACCGTCAGTGCAGC AGGGGATGAGGATGAGGATGAATCCGGGGAGGAGAGGCTGCCCTCCTGCTTTGACTAC GTCATGCACTTCCTGACTGTCTTCTGGAAGGTGCTGTTTGCCTGTGTGCCCCCCACAG AGTACTGCCACGGCTGGGCCTGCTTCGCCGTCTCCATCCTCATCATTGGCATGCTCAC CGCCATCATTGGGGACCTGGCCTCGCACTTCGGCTGCACCATTGGTCTCAAAGATTCA GTCACAGCTGTTGTTTTCGTGGCATTTGGCACCTCTGTCCCAGATACGTTTGCCAGCA AAGCTGCTGCCCTCCAGGATGTATATGCAGACGCCTCCATTGGCAACGTGACGGGCAG CAACGCCGTCAATGTCTTCCTGGGCATCGGCCTGGCCTGGTCCGTGGCCGCCATCTAC TGGGCTCTGCAGGGACAGGAGTTCCACGTGTCGGCCGGCACACTGGCCTTCTCCGTCA CCCTCTTCACCATCTTTGCATTTGTCTGCATCAGCGTGCTCTTGTACCGAAGGCGGCC GCACCTGGGAGGGGAGCTTGGTGGCCCCCGTGGCTGCAAGCTCGCCACAACATGGCTC TTTGTGAGCCTGTGGCTCCTCTACATACTCTTTGCCACACTAGAGGCCTATTGCTACA TCAAGGGGTTCCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 6	895 aa	MW at 99385.0kD
NOV1c, 248057963 Protein Sequence	GSEAGGSGDVPSTGQNNESCSGSSDCKEGVILPIWYPENPSLGDKIARVIVYFVALIY MFLGVSIIADRFMASIEVITSQEREVTIKKPNGETSTTTIRVWNETVSNLTLMALGSS APEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFIIIGICVYVIPDGETRKIKHLRV FFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFFFPVCVLLAWVADKRLLFYKY MHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFLDGNLVPLEGKEVDESRREMIR ILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQATRMMTGAGNILKKHAAEQ AKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLTVVRKGGDMSKTMYVDYK TEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEEDEHFFVRLSNVRIEEE QPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIHVSESIGVMEV KVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETVKTIRVKIVDEEEY ERQENFFIALGEPKWMERGISDVTDRKLTMEEEEAKRIAEMGKPVLGEHPKLEVIIEE SYEFKTTVDKLIKKTNLALVVGTHSWRDQFMEAITVSAAGDEDEDESGEERLPSCFDY VMHFLTVFWKVLFACVPPTEYCHGWACFAVSILIIGMLTAIIGDLASHFGCTIGLKDS VTAVVFVAFGTSVPDTFASKAAALQDVYADASIGNVTGSNAVNVFLGIGLAWSVAAIY WALQGQEFHVSAGTLAFSVTLFTIFAFVCISVLLYRRRPHLGGELGGPRGCKLATTWL FVSLWLLYILFATLEAYCYIKGFLE		

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

Table 1B. Comparison of NOV1a against NOV1b and NOV1c.

Protein Sequence	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV1b	1..928 1..925	833/929 (89%) 866/929 (92%)
NOV1c	30..928 2..893	808/899 (89%) 844/899 (93%)

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

Table 1C. Protein Sequence Properties NOV1a	
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 31 and 32

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

Table 1D. Geneseq Results for NOV1a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV1a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAM47745	Human natrium(+)-calcium(2+) exchanger form 3 protein, HNCX3 - Homo sapiens, 927 aa. [WO200183744-A2, 08-NOV-2001]	1..928 1..927	862/929 (92%) 900/929 (96%)	0.0
AAB41497	Human ORFX ORF1261 polypeptide sequence SEQ ID NO:2522 - Homo sapiens, 952 aa. [WO200058473-A2, 05-OCT-2000]	48..928 74..952	701/890 (78%) 788/890 (87%)	0.0
AAM26102	Peptide #139 encoded by probe for measuring placental gene expression - Homo sapiens, 609 aa. [WO200157272-A2, 09-AUG-2001]	1..593 11..608	420/606 (69%) 496/606 (81%)	0.0
AAM13701	Peptide #135 encoded by probe for measuring cervical gene expression - Homo sapiens, 609 aa. [WO200157278-A2, 09-AUG-2001]	1..593 11..608	420/606 (69%) 496/606 (81%)	0.0
AAM53461	Human brain expressed single exon probe encoded protein SEQ ID NO: 25566 - Homo sapiens, 609 aa. [WO200157275-A2, 09-AUG-2001]	1..593 11..608	420/606 (69%) 496/606 (81%)	0.0

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

Table 1E. Public BLASTP Results for NOV1a				
Protein Accession Number	**Protein/Organism/Length**	**NOV1a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96QG1	SODIUM/CALCIUM EXCHANGER SCL8A3 - Homo sapiens (Human), 924 aa.	1..928 1..924	866/929 (93%) 903/929 (96%)	0.0
Q96QG2	SODIUM/CALCIUM EXCHANGER SCL8A3 - Homo sapiens (Human), 925 aa.	1..928 1..925	857/930 (92%) 892/930 (95%)	0.0
P70549	Sodium/calcium exchanger 3 precursor (Na(+)/Ca(2+)-exchange protein 3) - Rattus norvegicus (Rat), 927 aa.	1..928 1..927	848/929 (91%) 895/929 (96%)	0.0
AAL39160	SODIUM/CALCIUM EXCHANGER - Mus musculus (Mouse), 928 aa.	1..928 1..928	837/929 (90%) 879/929 (94%)	0.0
Q9UPR5	Sodium/calcium exchanger 2 precursor (Na(+)/Ca(2+)-exchange protein 2) - Homo sapiens (Human), 921 aa.	48..928 43..921	701/890 (78%) 788/890 (87%)	0.0

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

Table 1F. Domain Analysis of NOV1a			
Pfam Domain	**NOV1a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Filo_VP35: domain 1 of 1	184..199	7/16 (44%) 10/16 (62%)	6.1
Na_Ca_Ex: domain 1 of 2	110..257	35/153 (23%) 120/153 (78%)	1.2e-32
Glycos_transf_4: domain 1 of 1	760..910	33/215 (15%) 95/215 (44%)	5.7
Na_Ca_Ex: domain 2 of 2	764..912	55/152 (36%) 130/152 (86%)	2.1e-48

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: 7	2277 bp
NOV2a, CG59843-01 DNA Sequence	CCGGCTCCCGCGCCCTCCCGGCCGGCC**ATG**CAGCCCCGCCGCGCCCAGGCGCCCGGTG CGCAGCTGCTGCCCGCGCTGGCCCTGCTGCTGCTGCTGCTCGGAGCGGGGCCCCGAGG CAGCTCCCTGGCCAACCCGGTGCCCGCCGCGCCCCTGTCTGCGCCCGGGCCGTGCGCC GCGCAGCCCTGCCGGAATGGGGGTGTGTGCACCTCGCGCCCTGAGCCGGACCCGCAGC ACCCGGCCCCCGCCGGCGAGCCTGGCTACAGCTGCACCTGCCCCGCCGGGATCTCCGG CGCCAACTGCCAGCTTGTTGCAGATCCTTGTGCCAGCAACCCTTGTCACCATGGCAAC TGCAGCAGCAGCAGCAGCAGCAGCAGCGATGGCTACCTCTGCATTTGCAATGAAGGCT ATGAAGGTCCCAACTGTGAACAGGCACTTCCCAGTCTCCCAGCCACTGGCTGGACCGA ATCCATGGCACCCCGACAGCTTCAGCCTGTTCCTGCTACTCAGGAGCCTGACAAAATC CTGCCTCGCTCTCAGGCAACGGTGACACTGCCTACCTGGCAGCCGAAAACAGGGCAGA AAGTTGTAGAAATGAAATGGGATCAAGTGGAGGTGATCCCAGATATTGCCTGTGGGAA TGCCAGTTCTAACAGCTCTGCGGGTGGCCGCCTGGTATCCTTTGAAGTGCCACAGAAC ACCTCAGTCAAGATTCGGCAAGATGCCACTGCCTCACTGATTTTGCTCTGGAAGGTCA CGGCCACAGGATTCCAACAGTGCTCCCTCATAGATGGACGAAGTGTGACCCCCCTTCA GGCTTCAGGGGGACTGGTCCTCCTGGAGGAGATGCTCGCCTTGGGGAATAATCACTTT ATTGGTTTTGTGAATGATTCTGTGACTAAGTCTATTGTGGCTTTGCGCTTAACTCTGG TGGTGAAGGTCAGCACCTGTGTGCCGGGGGAGAGTCACGCAAATGACTTGGAGTGTTC AGGAAAAGGAAAATGCACCACGAAGCCGTCAGAGGCAACTTTTTCCTGTACCTGTGAG GAGCAGTACGTGGGTACTTTCTGTGAAGAATACGATGCTTGCCAGAGGAAACCTTGCC AAAACAACGCGAGCTGTATTGATGCAAATGAAAAGCAAGATGGGAGCAATTTCACCTG TGTTTGCCTTCCTGGTTATACTGGAGAGCTTTGCCAGTCCAAGATTGATTACTGCATC CTAGACCCATGCAGAAATGGAGCAACATGCATTTCCAGTCTCAGTGGATTCACCTGCC AGTGTCCAGAAGGATACTTCGGATCTGCTTGTGAAGAAAAGGTGGACCCCTGCGCCTC GTCTCCGTGCCAGAACAACGGCACCTGCTATGTGGACGGGGTACACTTTACCTGCAAC TGCAGCCCGGGCTTCACAGGGCCGACCTGTGCCCAGCTTATTGACTTCTGTGCCCTCA GCCCCTGTGCTCATGGCACGTGCCGCAGCGTGGGCACCAGCTACAAATGCCTCTGTGA TCCAGGTTACCATGGCCTCTACTGTGAGGAGGAATATAATGAGTGCCTCTCCGCTCCA TGCCTGAATGCAGCCACCTGCAGGGACCTCGTTAATGGCTATGAGTGTGTGTGCCTGG CAGAATACAAAGGAACACACTGTGAATTGTACAAGGATCCCTGCGCTAACGTCAGCTG TCTGAACGGAGCCACCTGTGACAGCGACGGCCTGAATGGCACGTGCATCTGTGCACCC GGGTTTACAGGTGAAGAGTGCGACATTGACATAAATGAATGTGACAGTAACCCCTGCC ACCATGGTGGGAGCTGCCTGGACCAGCCCAATGGTTATAACTGCCACTGCCCGCATGG TTGGGTGGGAGCAAACTGTGAGATCCACCTCCAATGGAAGTCCGGGCACATGGCGGAG AGCCTCACCAACATGCCACGGCACTCCCTCTACATCATCATTGGAGCCCTCTGCGTGG CCTTCATCCTTATGCTGATCATCCTGATCGTGGGGATTTGCCGCATCAGCCGCATTGA ATACCAGGGTTCTTCCAGGCCAGCCTATGAGGAGTTCTACAACTGCCGCAGCATCGAC AGCGAGTTCAGCAATGCCATTGCATCCATCCGGCATGCCAGGTTTGGAAAGAAATCCC GGCCTGCAATGTATGATGTGAGCCCCATCGCCTATGAAGATTACAGTCCTGATGACAA ACCCTTGGTCACACTGATTAAAACTAAAGATTTG**TAA**TCTTTTTTTGGATTATTTTTC AAAAAGATGAGATAC	
	ORF Start: ATG at 28	ORF Stop: TAA at 2239
	SEQ ID NO: 8	737 aa	MW at 78473.7kD
NOV2a, CG59843-01 Protein Sequence	MQPRRAQAPGAQLLPALALLLLLLGAGPRGSSLANPVPAAPLSAPGPCAAQPCRNGGV CTSRPEPDPQHPAPAGEPGYSCTCPAGISGANCQLVADPCASNPCHHGNCSSSSSSS DGYLCICNEGYEGPNCEQALPSLPATGWTESMAPRQLQPVPATQEPDKILPRSQATVT LPTWQPKTGQKVVEMKWDQVEVIPDIACGNASSNSSAGGRLVSFEVPQNTSVKIRQDA TASLILLWKVTATGFQQCSLIDGRSVTPLQASGGLVLLEEMLALGNNHFIGFVNDSVT	

	KSIVALRLTLVVKVSTCVPGESHANDLECSGKGKCTTKPSEATFSCTCEEQYVGTFCE EYDACQRKPCQNNASCIDANEKQDGSNFTCVCLPGYTGELCQSKIDYCILDPCRNGAT CISSLSGFTCQCPEGYFGSACEEKVDPCASSPCQNNGTCYVDGVHFTCNCSPGFTGPT CAQLIDFCALSPCAHGTCRSVGTSYKCLCDPGYHGLYCEEEYNECLSAPCLNAATCRD LVNGYECVCLAEYKGTHCELYKDPCANVSCLNGATCDSDGLNGTCICAPGFTGEECDI DINECDSNPCHHGGSCLDQPNGYNCHCPHGWVGANCEIHLQWKSGHMAESLTNMPRHS LYIIIGALCVAFILMLIILIVGICRISRIEYQGSSRPAYEEFYNCRSIDSEFSNAIAS IRHARFGKKSRPAMYDVSPIAYEDYSPDDKPLVTLIKTKDL

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

Table 2B. Protein Sequence Properties NOV2a

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 35 and 36

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
AAU29042	Human PRO polypeptide sequence #19 - Homo sapiens, 737 aa. [WO200168848-A2, 20-SEP-2001]	1..737 1..737	737/737 (100%) 737/737 (100%)	0.0
AAB01313	Human PRO299 polypeptide - Homo sapiens, 737 aa. [WO200032776-A2, 08-JUN-2000]	1..737 1..737	737/737 (100%) 737/737 (100%)	0.0
AAY17822	Human PRO299 protein sequence - Homo sapiens, 737 aa. [WO9928462-A2, 10-JUN-1999]	1..737 1..737	737/737 (100%) 737/737 (100%)	0.0
AAW39257	Human membrane protein - Homo sapiens, 737 aa. [JP10036395-A, 10-FEB-1998]	1..737 1..737	737/737 (100%) 737/737 (100%)	0.0
AAW39256	Human partial mature membrane protein - Homo sapiens, 612 aa. [JP10036395-A, 10-FEB-1998]	27..638 1..612	612/612 (100%) 612/612 (100%)	0.0

In a BLAST search of public sequence datbases, 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
BAB72175	TRANSMEMBRANE PROTEIN BET - Mus musculus (Mouse), 737 aa.	1..737 1..737	666/737 (90%) 694/737 (93%)	0.0
AAH24766	HYPOTHETICAL 49.8 KDA PROTEIN - Homo sapiens (Human), 459 aa (fragment).	279..737 1..459	459/459 (100%) 459/459 (100%)	0.0
AAH22636	HYPOTHETICAL 42.4 KDA PROTEIN - Mus musculus (Mouse), 389 aa (fragment).	349..737 1..389	371/389 (95%) 382/389 (97%)	0.0
Q9NTF1	HYPOTHETICAL 27.8 KDA PROTEIN - Homo sapiens (Human), 252 aa (fragment).	486..737 1..252	252/252 (100%) 252/252 (100%)	e-158
Q9UDM2	WUGSC:H_NH0150O02.1 PROTEIN - Homo sapiens (Human), 192 aa (fragment).	384..575 1..192	192/192 (100%) 192/192 (100%)	e-123

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

Table 2E. Domain Analysis of NOV2a

Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value
EGF: domain 1 of 10	48..91	17/50 (34%) 34/50 (68%)	0.038
EGF: domain 2 of 10	98..132	16/47 (34%) 27/47 (57%)	1.5e-05
Vinculin: domain 1 of 1	225..248	11/29 (38%) 17/29 (59%)	6.5
EGF: domain 3 of 10	307..347	12/51 (24%) 28/51 (55%)	1.1
EGF: domain 4 of 10	353..389	14/47 (30%) 28/47 (60%)	5.9e-07

EGF: domain 5 of 10	396..427	16/47 (34%) 26/47 (55%)	2.8e-07
metalthio: domain 1 of 1	398..458	17/70 (24%) 29/70 (41%)	5.7
EGF: domain 6 of 10	434..465	16/47 (34%) 25/47 (53%)	2.2e-06
Keratin_B2: domain 1 of 1	343..496	39/194 (20%) 80/194 (41%)	0.72
EGF: domain 7 of 10	472..502	15/47 (32%) 25/47 (53%)	3.2e-07
EGF: domain 8 of 10	509..540	13/47 (28%) 23/47 (49%)	2.2e-06
EGF: domain 9 of 10	547..578	15/47 (32%) 23/47 (49%)	0.00048
DSL: domain 1 of 1	509..578	17/73 (23%) 44/73 (60%)	3.4
EGF: domain 10 of 10	585..616	16/47 (34%) 27/47 (57%)	2.7e-07
Rhabd_glycop: domain 1 of 1	638..684	9/50 (18%) 31/50 (62%)	1.5

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: 9	813 bp	
NOV3a, CG59845-01 DNA Sequence	TTGGGGAGAAGATTAAAAAATGCTAGAAGC**ATG**CCAACATATAAAATTCCAAGTCAAT GGTCAAACAGTGCACTCGCATCTCTCACATCACCCACTTGGCCCTCTTCCAAATGTAC TTTACTTCCTCTTCTCATCCTTTCGTTGCTGTCTCCCTTCACAGAACAATTCATAGTG AATAGCTTAAAGAGGCCAATCTTGGCTCCACTGGGTGGAAAAGTTGAGCTCAGTTGCC AGCTGTCTCCACCACAAAGCGCAGAACACATGGAAATACGCTGGTTCCAGAGTCACTA CACACGACCTGTTTACCTGTATAAGGATGGTAAAGACCTGTATGGAGAAACTATCTCC AAGTATGTGGAGCGGACAGAGCTCCTGAAAGAAGCCATTGGAGAAGGTAAAGTGACCC TCAGGATCCTTAATGTCAGTGCTGATGATGACGGGCAGTACCACTGCTTCTTCAAAGA CAGAAATGTCTATGAAGAGTCCATCACAGAAGTGAAGGTCTCAGATAAACTGTTTCCA TGGAATTCTATCTGGATACTGATTCTGGTTGCAATCTTGGCTGTTCTGCTATTCTTCA TTATGTTGGGAACTGTGTTCCTTTGGAGGAGGAGAGGCACTCTGCGTTTTAGAGTTTC CAGTTTTTCTGTTCTGTTTTTTCCCCATCTTTGTGGTTTTATCTACCGACTTGCATGT ACAAAGCTTCAACTCATCCTCCTGTCTGGGCCCCCACTTTTAATTCTCATTCTTTGTT ATGCATACAGTCTCAAGCCTTTC**TAG**GATATTACCAGGGCAGTTGACTGCCTTTAACT G		
	ORF Start: ATG at 31	ORF Stop: TAG at 778	
	SEQ ID NO: 10	249 aa	MW at 28550.3kD

NOV3a, CG59845-01 Protein Sequence	MPTYKIPSQWSNSALASLTSPTWPSSKCTLLPLLILSLLSPFTEQFIVNSLKRPILAP LGGKVELSCQLSPPQSAEHMEIRWFQSHYTRPVYLYKDGKDLYGETISKYVERTELLK EAIGEGKVTLRILNVSADDDGQYHCFFKDRNVYEESITEVKVSDKLFPWNSIWILILV AILAVLLFFIMLGTVFLWRRRGTLRFRVSSFSVLFFPHLCGFIYRLACTKLQLILLSG PPLLILILCYAYSLKPF

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

Table 3B. Protein Sequence Properties NOV3a

PSort analysis:	0.8000 probability located in mitochondrial inner membrane; 0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 46 and 47

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
AAY44236	Human myelin oligodendrocyte glycoprotein - Homo sapiens, 247 aa. [WO9960021-A2, 25-NOV-1999]	15..242 1..228	103/233 (44%) 147/233 (62%)	5e-45
AAW37543	Human myelin oligodendrocyte glycoprotein - Homo sapiens, 247 aa. [WO9735879-A1, 02-OCT-1997]	15..242 1..228	102/233 (43%) 147/233 (62%)	1e-44
AAR71360	Human MOG - Homo sapiens, 247 aa. [WO9507096-A, 16-MAR-1995]	15..242 1..228	102/233 (43%) 147/233 (62%)	1e-44
AAR70182	Human myelin oligonucleotide glycoprotein (MOG) - Homo sapiens, 247 aa. [WO9506727-A, 09-MAR-1995]	15..242 1..228	102/233 (43%) 147/233 (62%)	1e-44
AAR71361	Human truncated MOG - Homo sapiens, 203 aa. [WO9507096-A, 16-MAR-1995]	15..209 1..197	90/198 (45%) 129/198 (64%)	4e-40

In a BLAST search of public sequence datbases, 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**
Q9BGS7	HYPOTHETICAL 28.2 KDA PROTEIN - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 247 aa.	15..242 1..228	102/233 (43%) 148/233 (62%)	3e-44
Q96KU9	BA145L22.1.1 (MYELIN/OLIGODENDROCYTE GLYCOPROTEIN (MOG) BETA 1 (ISOFORM 1)) - Homo sapiens (Human), 252 aa.	15..242 1..228	102/233 (43%) 147/233 (62%)	3e-44
Q16653	Myelin-oligodendrocyte glycoprotein precursor - Homo sapiens (Human), 247 aa.	15..242 1..228	102/233 (43%) 147/233 (62%)	3e-44
A55717	myelin/oligodendrocyte glycoprotein precursor - mouse, 247 aa.	15..242 1..228	98/233 (42%) 139/233 (59%)	8e-40
CAB89269	BA145L22.1.6 (MYELIN/OLIGODENDROCYTE GLYCOPROTEIN (MOG), ISOFORM 6) - Homo sapiens (Human), 208 aa.	15..209 1..197	90/198 (45%) 129/198 (64%)	1e-39

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

Table 3E. Domain Analysis of NOV3a			
Pfam Domain	**NOV3a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ig: domain 1 of 1	60..143	13/85 (15%) 54/85 (64%)	0.00063
ATP-synt_B: domain 1 of 1	162..177	7/16 (44%) 14/16 (88%)	8.6

5

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: 11	1536 bp	
NOV4a, CG59871-01 DNA Sequence	**ATG**TGGGAGGACGGGCTGCCCCACCCCCGCGCGCAGGAACCCTGGTTTAGCTTAAGG GATGGAGGCGGGGACCCCTGCGCAGGCTTGCGGCGTGGGAGGCGGCCGCCCGCGACCT ACGACGCCGCGCGCCGGGAGGCTGAGAGTTCGGCGCCGGGAGGGTCCCGGGGACAGAA GAGCGCCTCGCCCGGTTGCCAAGGCAACCCCACGCGGCTGGAGAAGCCGGCGCTCGCA GCCCGGCCCGGGCCGCTGCCGGAAGTGACGCGAGTTCACCTGCCGAGCGGGGGCTGGG AGGAGGGGCGGAGGGTGCAGAGGTGCCGCCGCCGCCGCGAGCCAGTCGGGAGCGCGCG AGGCGCGGGGAGCCTGGGCACCAGGAGCGAGAGCCGCCTACCTGCAGCCGCCGCCCAC GGCACGGCAGCCACCATGGCGCTCCTGCTGTGCTTCGTGCTCCTGTGCGGAGTAGTGG ATTTCGCCAGAAGTTTGAGTATCACTACTCCTGAAGAGATGATTGAAAAAGCCAAAGG GGAAACTGCCTATCTGCCATGCAAATTTACGCTTAGTCCCGAAGACCAGGGACCGCTG GACATCGAGTGGCTGATATCACCAGCTGATAATCAGAAGGTGGATCAAGTGATTATTT TATATTCTGGAGACAAAATTTATGATGACTACTATCCAGATCTGAAAGGCCGAGTACA TTTTACGAGTAATGATCTCAAATCTGGTGATGCATCAATAAATGTAACGAATTTACAA CTGTCAGATATTGGCACATATCAGTGCAAAGTGAAAAAAGCTCCTGGTGTTGCAAATA AGAAGATTCATCTGGTAGTTCTTGTTAAGCCTTCAGGTGCGAGATGTTACGTTGATGG ATCTGAAGAAATTGGAAGTGACTTTAAGATAAAATGTGAACCAAAAGAAGGTTCACTT CCATTACAGTATGAGTGGCAAAAATTGTCTGACTCACAGAAAATGCCCACTTCATGGT TAGCAGAAATGACTTCATCTGTTATATCTGTAAAAAATGCCTCTTCTGAGTACTCTGG GACATACAGCTGTACAGTCAGAAACAGAGTGGGCTCTGATCAGTGCCTGTTGCGTCTA AACGTTGTCCCTCCTTCAAATAAAGCTGGACTAATTGCAGGAGCCATTATAGGAACTT TGCTTGCTCTAGCGCTCATTGGTCTTATCATCTTTTGCTGTCGTAAAAAGCGCAGAGA AGAAAAATATGAAAAGGAAGTTCATCACGATATCAGGGAAGATGTGCCACCTCCAAAG AGCCGTACGTCCACTGCCAGAAGCTACATCGGCAGTAATCATTCATCCCTGGGGTCCA TGTCTCCTTCCAACATGGAAGGATATTCCAAGACTCAGTATAAACAAGTACCAAGTGA AGACTTTGAACGCACTCCTCAGAGTCCGACTCTCCCACCTGCTAAGGTAGCTGCCCCT AATCTAAGTCGAATGGGCGCGATTCCTGTGATGATTCCCGCACAGAGCAAGGATGGGT CTATAGTA**TAG**AGCCTCCATACATCTCA		
	ORF Start: ATG at 2	ORF Stop: TAG at 1517	
	SEQ ID NO: 12	505 aa	MW at 54859.8kD
NOV4a, CG59871-01 Protein Sequence	MWEDGLPHPRAQEPWFSLRDGGGDPCAGLRRGRRPPATYDAARREAESSAPGGSRGQK SASPGCQGNPTRLEKPALAARPGPLPEVTRVHLPSGGWEEGRRVQRCRRRREPVGSAR GAGSLGTRSESRLPAAAAHGTAATMALLLCFVLLCGVVDFARSLSITTPEEMIEKAKG ETAYLPCKFTLSPEDQGPLDIEWLISPADNQKVDQVIILYSGDKIYDDYYPDLKGRVH FTSNDLKSGDASINVTNLQLSDIGTYQCKVKKAPGVANKKIHLVVLVKPSGARCYVDG SEEIGSDFKIKCEPKEGSLPLQYEWQKLSDSQKMPTSWLAEMTSSVISVKNASSEYSG TYSCTVRNRVGSDQCLLRLNVVPPSNKAGLIAGAIIGTLLALALIGLIIFCCRKKRRE EKYEKEVHHDIREDVPPPKSRTSTARSYIGSNHSSLGSMSPSNMEGYSKTQYKQVPSE DFERTPQSPTLPPAKVAAPNLSRMGAIPVMIPAQSKDGSIV		

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

Table 4B. Protein Sequence Properties NOV4a	
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

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
AAB47270	Human CAR - Homo sapiens, 365 aa. [US6245966-B1, 12-JUN-2001]	141..505 1..365	364/365 (99%) 364/365 (99%)	0.0
AAW57212	Human coxsackievirus and adenovirus receptor - Homo sapiens, 365 aa. [WO9811221-A2, 19-MAR-1998]	141..505 1..365	364/365 (99%) 364/365 (99%)	0.0
AAW69697	Human coxsackievirus and Ad2 and Ad5 receptor HCAR protein - Homo sapiens, 365 aa. [WO9833819-A1, 06-AUG-1998]	141..505 1..365	364/365 (99%) 364/365 (99%)	0.0
AAB50930	Human PRO5723 protein - Homo sapiens, 352 aa. [WO200073452-A2, 07-DEC-2000]	141..483 1..343	339/343 (98%) 339/343 (98%)	0.0
AAB65294	Human PRO5723 protein sequence SEQ ID NO:505 - Homo sapiens, 352 aa. [WO200073454-A1, 07-DEC-2000]	141..483 1..343	339/343 (98%) 339/343 (98%)	0.0

In a BLAST search of public sequence datbases, 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**
P78310	Coxsackievirus and adenovirus receptor precursor (Coxsackievirus B-adenovirus receptor) (hCAR) (CVB3 binding protein) - Homo sapiens (Human), 365 aa.	141..505 1..365	364/365 (99%) 364/365 (99%)	0.0
Q9UKV4	COXSACKIE AND ADENOVIRUS RECEPTOR PROTEIN - Homo sapiens (Human), 344 aa (fragment).	141..479 1..339	338/339 (99%) 338/339 (99%)	0.0
AAK57804	COXSACKIE VIRUS AND ADENOVIRUS RECEPTOR BCAR - Bos taurus (Bovine), 365 aa.	141..505 1..365	331/365 (90%) 345/365 (93%)	0.0
P97792	Coxsackievirus and adenovirus receptor homolog precursor (mCAR) - Mus musculus (Mouse), 365 aa.	141..505 1..365	327/365 (89%) 344/365 (93%)	0.0
Q9DBJ8	COXSACKIEVIRUS AND ADENOVIRUS RECEPTOR - Mus musculus (Mouse), 366 aa.	141..505 1..366	327/366 (89%) 344/366 (93%)	0.0

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

Table 4E. Domain Analysis of NOV4a			
Pfam Domain	**NOV4a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ig: domain 1 of 2	174..262	13/90 (14%) 62/90 (69%)	0.0054
ig: domain 2 of 2	295..354	11/62 (18%) 46/62 (74%)	1.5e-05
Adeno_E3_CR2: domain 1 of 1	372..417	15/50 (30%) 24/50 (48%)	4.9

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: 13	1302 bp	
NOV5a, CG59883-01 DNA Sequence	**ATG**AACATGAAATTGCAGATGTCTGTTTTTGACATACTGATTAAAATTCCTTTGGACA CGTATCCAGAAGTGGGATTGATGGATCATAGGTGCGCCAGGCGCGGGGAGCCTAGGAC CTGGAGCGAGAGCCGCCTACCTGCAGCCGCCGCCCACGGCACGGCAGCCACCGTGGCG CTCCTGCTGCGCTTCGTGCTCCTGTGCAGAGTCGCGGATTTCATCAGAGGTTGGAGTA TCACTACTCCTGAGCAGATGATTGAAAAAGCCAAAGGGGAAACTGCCTATCTGCCATG CAAATTTACGCTTAGTCCTGAAGACCAGGGACCACTGGACATCGAGTGGCTGATATCA CCAGCTGATAATCAGAAGGTGGATCAAGTGATTATTTTATATTCTGGAGACAAAAATT ATGATGACTACTATCCAGATCTGAAAGGCCGAGTACATTTTAAGAGTAATGATCTCAA ATCTGGTGATGCATCAATAAATGTAACGAATTTTCAGCTGTCAGATATTGGCACAGAT CAGTGCAAAGTGAAAAGAGCTCCTGGTGTTGCAAATAGGAAGATTCAGCTGGTAGTTC TTGGTAAGCCTTCAGGTACAAGATGTTACGTTGATGGATCAGAAGAAATTGGAAGTGA CTTTAAATTAAAATGTGAACCAAAAGAAGGTTCACTTCCATTACAGTATGAGTGGCAA AAATTGTCTGACTCACAGAAAATGCCCACTTCATGGTTAGCAGAAATGACTTCATCTG TTATATCTGTAAAAAAAAATGCTTCTTCTGAGTACTCTGGGACATACAGCTGTACAAT CAGAAACAGAGTGGGCTCTGATCAGTGCCTGTTGCGTGTAAACGTTGTCCCTCCTTCA AATAAAGCTGGACTAATTGCAGGAGCCATTATAGGAACTTTGCTTGCTCTAGTGCTCA TTGGTCTTATCATCTTTTGCTGTCGTAAAAAGCGCAGAGAAGAAAAATATGAAAAGGA AGTTCATCACGATATCAAGGAAGATGTGCCGCCTCCAAAGAGCCACACGTCCACTGCC AGAAGCTACATAGGCAGTAATCATTCATCCCTGGGATCCATATCTCCTTCCAACATGG AAGGATATTCCAAGACTCAGTATAAACAAGTACCAAGTGAAGACTTTGAACGCACTCC TCAGAGTCCGACTCTCCCACCTGCTAAGGTAGCTGCCCCTAATCTAAGTCGAATGGGC GCGATTCCTGTGATGATTCCCGCACAGAGCAAGGATGGGTCTATAGTA**TAG**AGCCTCC ATACATCTCATCTGTGCTCTCCGTGT		
	ORF Start: ATG at 1	ORF Stop: TAG at 1267	
	SEQ ID NO: 14	422 aa	MW at 46596.9kD
NOV5a, CG59883-01 Protein Sequence	MNMKLQMSVFDILIKIPLDTYPEVGLMDHRCARRGEPRTWSESRLPAAAAHGTAATVA LLLRFVLLCRVADFIRGWSITTPEQMIEKAKGETAYLPCKFTLSPEDQGPLDIEWLIS PADNQKVDQVIILYSGDKNYDDYYPDLKGRVHFKSNDLKSGDASINVTNFQLSDIGTD QCKVKRAPGVANRKIQLVVLGKPSGTRCYVDGSEEIGSDFKLKCEPKEGSLPLQYEWQ KLSDSQKMPTSWLAEMTSSVISVKKNASSEYSGTYSCTIRNRVGSDQCLLRVNVVPPS NKAGLIAGAIIGTLLALVLIGLIIFCCRKKRREEKYEKEVHHDIKEDVPPPKSHTSTA RSYIGSNHSSLGSISPSNMEGYSKTQYKQVPSEDFERTPQSPTLPPAKVAAPNLSRMG AIPVMIPAQSKDGSIV		

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

Table 5B. Protein Sequence Properties NOV5a	
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:	Cleavage site between residues 23 and 24

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**
AAB47270	Human CAR - Homo sapiens, 365 aa. [US6245966-B1, 12-JUN-2001]	57..422 1..365	340/366 (92%) 349/366 (94%)	0.0
AAW57212	Human coxsackievirus and adenovirus receptor - Homo sapiens, 365 aa. [WO9811221-A2, 19-MAR-1998]	57..422 1..365	340/366 (92%) 349/366 (94%)	0.0
AAW69697	Human coxsackievirus and Ad2 and Ad5 receptor HCAR protein - Homo sapiens, 365 aa. [WO9833819-A1, 06-AUG-1998]	57..422 1..365	340/366 (92%) 349/366 (94%)	0.0
AAW57213	Mouse coxsackievirus and adenovirus receptor - Mus sp, 376 aa. [WO9811221-A2, 19-MAR-1998]	57..422 1..365	316/366 (86%) 338/366 (92%)	0.0
AAB50930	Human PRO5723 protein - Homo sapiens, 352 aa. [WO200073452-A2, 07-DEC-2000]	57..400 1..343	315/344 (91%) 324/344 (93%)	0.0

In a BLAST search of public sequence datbases, 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**
P78310	Coxsackievirus and adenovirus receptor precursor (Coxsackievirus B-adenovirus receptor) (hCAR) (CVB3 binding protein) - Homo sapiens (Human), 365 aa.	57..422 1..365	340/366 (92%) 349/366 (94%)	0.0
AAK57804	COXSACKIE VIRUS AND ADENOVIRUS RECEPTOR BCAR - Bos taurus (Bovine), 365 aa.	59..422 3..365	323/364 (88%) 341/364 (92%)	0.0
P97792	Coxsackievirus and adenovirus receptor homolog precursor (mCAR) - Mus musculus (Mouse), 365 aa.	57..422 1..365	317/366 (86%) 339/366 (92%)	0.0

Q9DBJ8	COXSACKIEVIRUS AND ADENOVIRUS RECEPTOR - Mus musculus (Mouse), 366 aa.	57..422 1..366	317/367 (86%) 339/367 (91%)	0.0
Q9R066	COXSACKIE-ADENOVIRUS-RECEPTOR HOMOLOG - Rattus norvegicus (Rat), 358 aa (fragment).	57..415 1..358	314/359 (87%) 332/359 (92%)	0.0

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

Table 5E. Domain Analysis of NOV5a

Pfam Domain	NOV5a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ig: domain 1 of 2	90..178	12/90 (13%) 62/90 (69%)	37
ig: domain 2 of 2	211..271	11/63 (17%) 43/63 (68%)	0.014
Adeno_E3_CR2: domain 1 of 1	289..334	15/50 (30%) 24/50 (48%)	3.2

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: 15	2412 bp
NOV6a, CG59901-01 DNA Sequence	**ATG**GCCCGGGCCAGGGCCGGGGCGCTGCTGGCGCTTTGGGTGCTCGGGGCCGCCGCGC ATCCGCAGTGCCTGGACTTCAGGCCGCCCTTCCGGCCGACGCAGCCGCTGCGCCTCTG CGCGCAGTACTCGGACTTCGGCTGCTGCGATGAGGGGCGCGACGCCGAGCTGACCCGC CGCTTCTGGGCCCTGGCGAGCCGCGTGGACGCCGCCGAGTGGGCCGCGTGCGCCGGCT ACGCGAGGGACCTGCTGTGCCAGTCCGTGGAGTGGACAGACATGCAAAGAGATAATGA AGTCCTAGCCAAGCTGACTGGCTGGAGCGCCCCTGGCGACGGAGCAGTGACTGCTGTA GAGAACTCACCCTCTCTGGACTACCCTGGTCTGGGCACCACATTTACGTCGTGTGAAT GCTCGCCGTATGCAGCCCACCTCTATGACGCCGAGGACCCATTCACGCCCCTGCGCAC GGTGCCCGGGCTCTGCCAGGATTACTGCCTGGACATGTGGCATAAGTGCCGGGGGCTG TTCCGTCACCTGTCAACTGACCAGGAGCTCTGGGCGCTGGAGGGCAACCTTGCCAGGT TCTGCCGCTACCTGTCCCTGGATGACACGGACTACTGCTTCCCTTACCTGCTGGTCAA CAAGAACCTCAACTCAAACCTGGGCCACGTGGTAGCCGATGCCAAGGGCTGCCTGCAG CTGTGCCTGGAGGAGGTGGCCAACGGGCTGCGCAACCCCGTGGCCATGGTCCATGCCA GGGATGGCACCCACCGCTTCTTCGTGGCCGAGCAGGTGGGGCTGGTGTGGGCCTACCT GCCCGACCGCTCGAGGCTGGGGAAGCCTTTCCTGAACATCAGCCGGGTGGTGCTCACC TCGCCCTGGGAGGGTGACGAGCGTGGCTTCCTGGGCATTGCCTTCCACCCCAGCTTCC AGCACAACCGCAGGCTCTACGTCTACTACTCAGTGGGTATCCGCAGCAGTGAGTGGAT	2412 bp

	CCGCATCAGCGAGTTCAGAGTCTCCGAGGATGACGAGAACGCCGTGGACCACAGCTCT GAGAGGATAATCCTGGAGGTCAAAGAACCAGCCTCAAACCACAACGGGGGCCAGCTGC TTTTCGGGGATGACGGGTACCTCTACATCTTCACTGGAGATGGCGGGATGGCCGGAGA CCCCTTTGGGACATTTGGAAATGCCCAAAACAAGTATGTTCAGCTTTTGATTGGCTTG TGGGTTGGTCTCCATATCCCTGGGCTTCTCATACTCTTCCAGAGGTCGGCGCTGCTGG GCAAGGTGCTGCGCATCGACGTGGACCCCGAGGTCTACGCCCTAGGCGTGCGCAACAT GTGGCGCTGCTCCTTCGACCGTGGCGACCCCTCCTCGGGCACTGGCCGCGGGCGCCTC TTCTGCGGCGACGTGGGCCAGAACAAGTTCGAGGAGGTGGACGTGGTGGAGCGCGGCG GCAACTATGGCTGGCGCGCGCGCGAAGGGTTCGAGTGCTACGACCGCAGCCTGTGCGC CAACACCTCTCTCAATGACTTGCTGCCGATTTTCGCCTACCCGCACACGGTTGGCAAG TCGGTCACAGGGGGCTACGTGTACCGGGGCTGCGAGTACCCCAACCTGAACGGCCTCT ACATTTTTGGGGATTTCATGAGCGGGCGTCTGATGTCCCTCCAAGAGAACCCAGGGAC AGGCCAGTGGCAGTACAGTGAGATCTGCATGGGCCACGGCCAGACCTGTGAGTTCCCA GGCCTCATCAACAACTACTACCCGTACATCATCTCCTTCGGGGAGGACGAGGCCGGGG AGCTGTACTTCATGTCGACAGGGGAGCCGAGTGCCACAGCTCCACGCGGAGTTGTCTA CAAAATAATTGACGCATCCAGAGTTCATCCCGAAGACACGGAGCACCCCGCGGCCTAC AGCGCGGGCGCCCACGCGGGCGCCCGCCGAGGGCGCCCCACGGCCGCTCCCCCCGCGC CAACCCCGCGGCCAGCGCGGCCCACCCAGCAGCCAGGGAGCCGGAGGGGCGGCGGGCG GCGGCGGGGGCGGCTGAACTCGGCGAGCCGGGCGTTCCGGGATGGCGAGGTGCGCCTG GTGCGGCCCGCGGGCCTGAGCTCTGGCAGCGGGCGCGTGGAGGTGTTCGTGGGCGGAC GCTGGGGCACCGTGTGCGACGACTCCTGGAACATCAGCGGCGCCGCCGTCGTGTGTCG CCAGCTGGGGTTTGCCTACGCCGTGCGCGCCGTCAAGAGAGCCGAGTTCGGCCAGGGC GGCTCGCTGCCCATTCTGCTGGACGATGTGCGCTGCGCGGGCTGGGAGCGGAACCTGC TGGAGTGCCAGCACAACGGCGTGGGCACCCACAACTGCGAGCACGACGAGGATGCGGG CGTCGTGTGCAGCCACCAGAACCCCGACCTG**TAG**
	ORF Start: ATG at 1 | ORF Stop: TAG at 2410
	SEQ ID NO: 16 | 803 aa | MW at 88653.7kD
NOV6a, CG59901-01 Protein Sequence	MARARAGALLALWVLGAAAHPQCLDFRPPFRPTQPLRLCAQYSDFGCCDEGRDAELTR RFWALASRVDAAEWAACAGYARDLLCQSVEWTDMQRDNEVLAKLTGWSAPGDGAVTAV ENSPSLDYPGLGTTFTSCECSPYAAHLYDAEDPFTPLRTVPGLCQDYCLDMWHKCRGL FRHLSTDQELWALEGNLARFCRYLSLDDTDYCFPYLLVNKNLNSNLGHVVADAKGCLQ LCLEEVANGLRNPVAMVHARDGTHRFFVAEQVGLVWAYLPDRSRLGKPFLNISRVVLT SPWEGDERGFLGIAFHPSFQHNRRLYVYYSVGIRSSEWIRISEFRVSEDDENAVDHSS ERIILEVKEPASNHNGGQLLFGDDGYLYIFTGDGGMAGDPFGTFGNAQNKYVQLLIGL WVGLHIPGLLILFQRSALLGKVLRIDVDPEVYALGVRNMWRCSFDRGDPSSGTGRGRL FCGDVGQNKFEEVDVVERGGNYGWRAREGFECYDRSLCANTSLNDLLPIFAYPHTVGK SVTGGYVYRGCEYPNLNGLYIFGDFMSGRLMSLQENPGTGQWQYSEICMGHGQTCEFP GLINNYYPYIISFGEDEAGELYFMSTGEPSATAPRGVVYKIIDASRVHPEDTEHPAAY SAGAHAGARRGRPTAAPPAPTPRPARPTQQPGSRRGGGRRRGRLNSASRAFRDGEVRL VRPAGLSSGSGRVEVFVGGRWGTVCDDSWNISGAAVVCRQLGFAYAVRAVKRAEFGQG GSLPILLDDVRCAGWERNLLECQHNGVGTHNCEHDEDAGVVCSHQNPDL

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

Table 6B. Protein Sequence Properties NOV6a

PSort analysis:	0.4600 probability located in plasma membrane; 0.2073 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 20 and 21

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 6C.

Table 6C. Geneseq Results for NOV6a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU12201	Human PRO1779 polypeptide sequence - Homo sapiens, 724 aa. [WO200140466-A2, 07-JUN-2001]	20..629 42..605	343/636 (53%) 410/636 (63%)	0.0
AAB25594	Protein encoded by human secreted protein gene #1 - Homo sapiens, 724 aa. [WO200029435-A1, 25-MAY-2000]	20..629 42..605	343/636 (53%) 410/636 (63%)	0.0
AAB94773	Human protein sequence SEQ ID NO:15860 - Homo sapiens, 529 aa. [EP1074617-A2, 07-FEB-2001]	223..629 2..410	269/432 (62%) 319/432 (73%)	e-159
AAB25576	Protein encoded by human secreted protein gene #1 - Homo sapiens, 529 aa. [WO200029435-A1, 25-MAY-2000]	223..629 2..410	269/432 (62%) 319/432 (73%)	e-159
AAY97561	Mouse Hedgehog interacting protein sequence - Mus musculus, 700 aa. [WO200074706-A1, 14-DEC-2000]	93..631 50..593	183/615 (29%) 272/615 (43%)	5e-59

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

Table 6D. Public BLASTP Results for NOV6a

Protein Accession Number	Protein/Organism/Length	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96JK4	KIAA1822 PROTEIN - Homo sapiens (Human), 533 aa (fragment).	299..803 1..533	464/556 (83%) 466/556 (83%)	0.0
Q91638	GENE 5 PROTEIN - Xenopus laevis (African clawed frog), 995 aa.	5..626 39..606	346/648 (53%) 419/648 (64%)	0.0

Q9H8A0	CDNA FLJ13840 FIS, CLONE THYRO1000783, MODERATELY SIMILAR TO XENOPUS LAEVIS TAIL-SPECIFIC THYROID HORMONE UP-REGULATED (GENE 5) MRNA - Homo sapiens (Human), 529 aa.	223..629 2..410	269/432 (62%) 319/432 (73%)	e-159
Q9D2G9	4930507C10RIK PROTEIN - Mus musculus (Mouse), 497 aa.	248..629 1..383	260/407 (63%) 299/407 (72%)	e-148
Q96BT4	SIMILAR TO HYPOTHETICAL PROTEIN FLJ13840 - Homo sapiens (Human), 256 aa.	223..475 2..256	168/278 (60%) 195/278 (69%)	1e-91

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

Table 6E. Domain Analysis of NOV6a

Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
SRCR: domain 1 of 1	699..797	49/115 (43%) 73/115 (63%)	6.2e-25

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: 17	2111 bp
NOV7a, CG88748-01 DNA Sequence	GTCCTCTGTGACCTTCTTCTTGGCCCTCTTCCTGGCAGATAAGTGCTCTGGTTGTACA TGGAGG**ATG**ACCGAAAAAACCAATGGTGTGAAGAGCTCCCCAGCCAATAATCACAACC ATCATGCACCTCCTGCCATCAAGGCCAATGGCAAAGATGACCACAGGACAAGCAGCAG GCCACACTCTGCAGCTGACGATGACACCTCCTCAGAACTGCAGAGGCTGGCAGACGTG GATGCCCCACAGCAGGGAAGGAGTGGCTTCCGCAGGATAGTTCGCCTGGTGGGGATCA TCAGAGAATGGGCCAACAAGAATTTCCGAGAGGAGGAACCTAGGCCTGACTCATTCCT CGAGCGTTTTCGTGGGCCTGAACTCCAGACTGTGACCACACAGGAGGGGGATGGCAAA GGCGACAAGGATGGCGAGGACAAAGGCACCAAGAAGAAATTTGAACTATTTGTCTTGG ACCCAGCTGGGGATTGGTACTACTGCTGGCTATTTGTCATTGCCATGCCCGTCCTTTA CAACTGGTGCCTGCTGGTGGCCAGAGCCTGCTTCAGTGACCTACAGAAAGGCTACTAC CTGGTGTGGCTGGTGCTGGATTATGTCTCAGATGTGGTCTACATTGCGGACCTCTTCA TCCGATTGCGCACAGGTTTCCTGGAGCAGGGGCTGCTGGTCAAAGATACCAAGAAACT GCGAGACAACTACATCCACACCCTGCAGTTCAAGCTGGATGTGGCTTCCATCATCCCC ACTGACCTGATCTATTTTGCTGTGGACATCCACAGCCCTGAGGTGCGCTTCAACCGCC TGCTGCACTTTGCCCGCATGTTTGAGTTCTTTGACCGGACAGAGACACGCACCAACTA CCCTAACATCTTCCGCATCAGCAACCTTGTCCTCTACATCTTGGTCATCATCCACTGG AATGCCTGCATCTATTATGCCATCTCCAAATCCATAGGCTTTGGGGTCGACACCTGGG	2111 bp

	TTTACCCAAACATCACTGACCCTGAGTATGGCTACCTGGCTAGGGAATACATCTATTG CCTTTACTGGTCCACACTGACTCTCACTACCATTGGGGAGACACCACCCCCTGTAAAG GATGAGGAGTACCTATTTGTCATCTTTGACTTCCTGATTGGCGTCCTCATCTTTGCCA CCATCGTGGGAAATGTGGGCTCCATGATCTCCAACATGAATGCCACCCGGGCAGAGTT CCAGGCTAAGATCGATGCCGTGAAACACTACATGCAGTTCCGAAAGGTCAGCAAGGGG ATGGAAGCCAAGGTCATTAGGTGGTTTGACTACTTGTGGACCAATAAGAAGACAGTGG ATGAGCGAGAAATTCTCAAGAATCTGCCAGCCAAGCTCAGGGCTGAGATAGCCATCAA TGTCCACTTGTCCACACTCAAGAAAGTGCGCATCTTCCATGATTGTGAGGCTGGCCTG CTGGTAGAGCTGGTACTGAAACTCCGTCCTCAGGTCTTCAGTCCTGGGGATTACATTT GCCGCAAAGGGGACATCGGCAAGGAGATGTACATCATTAAGGAGGGCAAACTGGCAGT GGTGGCTGATGATGGTGTGACTCAGTATGCTCTGCTGTCGGCTGGAAGCTGCTTTGGC GAGATCAGTATCCTTAACATTAAGGGCAGTAAAATGGGCAATCGACGCACAGCTAATA TCCGCAGCCTGGGCTACTCAGATCTCTTCTGCTTGTCCAAGGATGATCTTATGGAAGC TGTGACTGAGTACCCTGATGCCAAGAAAGTCCTAGAAGAGAGGGGTCGGGAGATCCTC ATGAAGGAGGGACTGCTGGATGAGAACGAAGTGGCAACCAGCATGGAGGTCGACGTGC AGGAGAAGCTAGGGCAGCTGGAGACCAACATGGAAACCTTGTACACTCGCTTTGGCCG CCTGCTGGCTGAGTACACGGGGGCCCAGCAGAAGCTCAAGCAGCGCATCACAGTTCTG GAAACCAAGATGAAACAGAACAATGAAGATGACTACCTGTCTGATGGGATGAACAGCC CTGAGCTGGCTGCTGCTGACGAGCCA**TAA**GACCTGGGGCCCAACTGCCTCTCCAGCAT TGGCCTTGGCCTTGATCCCAGAA		
	ORF Start: ATG at 65	ORF Stop: TAA at 2057	
	SEQ ID NO: 18	664 aa	MW at 76047.3kD
NOV7a, CG88748-01 Protein Sequence	MTEKTNGVKSSPANNHNHHAPPAIKANGKDDHRTSSRPHSAADDDTSSELQRLADVDA PQQGRSGFRRIVRLVGIIREWANKNFREEEPRPDSFLERFRGPELQTVTTQEGDGKGD KDGEDKGTKKKFELFVLDPAGDWYYCWLFVIAMPVLYNWCLLVARACFSDLQKGYYLV WLVLDYVSDVVYIADLFIRLRTGFLEQGLLVKDTKKLRDNYIHTLQFKLDVASIIPTD LIYFAVDIHSPEVRFNRLLHFARMFEFFDRTETRTNYPNIFRISNLVLYILVIIHWNA CIYYAISKSIGFGVDTWVYPNITDPEYGYLAREYIYCLYWSTLTLTTIGETPPPVKDE EYLFVIFDFLIGVLIFATIVGNVGSMISNMNATRAEFQAKIDAVKHYMQFRKVSKGME AKVIRWFDYLWTNKKTVDEREILKNLPAKLRAEIAINVHLSTLKKVRIFHDCEAGLLV ELVLKLRPQVFSPGDYICRKGDIGKEMYIIKEGKLAVVADDGVTQYALLSAGSCFGEI SILNIKGSKMGNRRTANIRSLGYSDLFCLSKDDLMEAVTEYPDAKKVLEERGREILMK EGLLDENEVATSMEVDVQEKLGQLETNMETLYTRFGRLLAEYTGAQQKLKQRITVLET KMKQNNEDDYLSDGMNSPELAAADEP		

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

Table 7B. Protein Sequence Properties NOV7a

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

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**
AAE04894	Human transporter and ion channel-7 (TRICH-7) protein - Homo sapiens, 664 aa. [WO200146258-A2, 28-JUN-2001]	1..664 1..664	664/664 (100%) 664/664 (100%)	0.0
AAM47673	MOL10b protein sequence - Homo sapiens, 575 aa. [WO200181578-A2, 01-NOV-2001]	124..657 18..555	290/540 (53%) 394/540 (72%)	e-170
AAM47672	MOL10a protein sequence - Unidentified, 578 aa. [WO200181578-A2, 01-NOV-2001]	132..657 27..558	290/534 (54%) 391/534 (72%)	e-168
ABG27071	Novel human diagnostic protein #27062 - Homo sapiens, 259 aa. [WO200175067-A2, 11-OCT-2001]	198..399 57..258	151/202 (74%) 176/202 (86%)	3e-88
ABG27071	Novel human diagnostic protein #27062 - Homo sapiens, 259 aa. [WO200175067-A2, 11-OCT-2001]	198..399 57..258	151/202 (74%) 176/202 (86%)	3e-88

In a BLAST search of public sequence datbases, 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**
S35691	cyclic nucleotide-gated channel protein - rabbit, 732 aa.	1..664 69..732	624/664 (93%) 643/664 (95%)	0.0
Q28718	Cyclic-nucleotide-gated olfactory channel (Cyclic-nucleotide-gated cation channel 2) (CNG channel 2) (CNG-2) (CNG2) (Aorta CNG channel) (RACNG) - Oryctolagus cuniculus (Rabbit), 664 aa.	1..664 1..664	624/664 (93%) 643/664 (95%)	0.0

Q03041	Cyclic-nucleotide-gated olfactory channel (Cyclic-nucleotide-gated cation channel 2) (CNG channel 2) (CNG-2) (CNG2) - Bos taurus (Bovine), 663 aa.	1..657 1..657	618/657 (94%) 639/657 (97%)	0.0
Q62398	Cyclic-nucleotide-gated olfactory channel (Cyclic-nucleotide-gated cation channel 2) (CNG channel 2) (CNG-2) (CNG2) - Mus musculus (Mouse), 664 aa.	1..662 2..664	618/663 (93%) 636/663 (95%)	0.0
Q00195	Cyclic-nucleotide-gated olfactory channel (Cyclic-nucleotide-gated cation channel 2) (CNG channel 2) (CNG2) (CNG-2) (OCNC1) - Rattus norvegicus (Rat), 664 aa.	1..662 2..664	615/663 (92%) 636/663 (95%)	0.0

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

Table 7E. Domain Analysis of NOV7a

Pfam Domain	NOV7a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ion_trans: domain 1 of 1	174..371	35/236 (15%) 152/236 (64%)	5.1e-22
cNMP_binding: domain 1 of 1	469..565	34/120 (28%) 81/120 (68%)	1.4e-25

Example 8.

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: 19	2273 bp
NOV8a, CG90021-01 DNA Sequence	TCCTTTCCTTTCCCATTGTCTTTGGATGACTATCGCTGGGCTGGGAC**ATG**AGGCGGGC AGAGGCGCGGGTCACCCTTAGGACCCCCCTCTTGCTGCTGGGGCTCTGGGCGCTCCTG GCTCCGGTCCGGTGTTCTCAAGGCCGTCCCTTGTGGCACTATGCCTCCTCCGAGGTGG TGATTCCCAGGAAGGAGACACACCATAGCAAAGGCCTTCAGTTTCCCGGCTGGCTGTC CTACAGCCTGTGTTTTGGGTTTTGGGGTCAAAGACACGTCATTCACATGCGGAGGAAA CACCTTCTTTGGCCTAGACATCTGCTGGTGACAACTCAGGATGACCAAGGAGTCTTGC AGATGGGTGACCCCTACATCCCTCCAGACTGCTACTACCTCGGCTACCTGGAGGAGGT GCCTCTGTCCATGGTCACCGTCGACACGTGCTATGGGGACCTCAGAGGCATCATGAGG CTGGACGACCTTGCGTACGAAATCAAACCCCTCCAGGATTCCCGCAGGTTTGAACATG TTGTTTTTCAGATAGTGGCCGAGCCCAACGCAACAGGGCCCACATTTAGAGATGATGA CAATGAGACAGACCCCCTGTTCTCTGAAGCAAATGACAGCATGAATCCCAGGATATCT

	AATTCGCTGTATAGTTCTCATAGAGGCAATATAAAAGGCCACGTTCAATGTTCCAATT CATATTATCGCATATATGGCAATATTACAACTTGTTCCAAAGAGGTGGTCCAGATGTT CAGTCTCATTGACAGCATTGCTCAAAATATTGATCTGCGGTACTATATTTATCTTTTG ACCATATATAATAATCGTGACCCAGCCCCTGTGAATGAATATCGAATTCAGAGTGCAA TGTTTACCTATTTTAAAACAACTTTTTTTGATACTTTTCATGTTCATTCATCCACACT ACTTATTAAATACGTGCCACATGAATCTAACTATGAACCTGAAAGGTATAACTTCTGT TCCCGTATAGCCCTGTTACACATTGGTACTCCAGGCAGACATTATTTATTGGTAGCCG TCATAATAACCCAGACACAGATGAGAAGTATTGGTCTGGAGTATGATGATAACTACTG CACATGTCAGAGAAGGGCCTCCTGCATTATGCAGCGATTTCCTGGGATGACAGATGCG TTCAGTAACTGTTCTTATGGACATGCACAAAATTGTTTTATACATTCAGGCCGGTGTG TTTTTGAAACACTTGCTCCTGTGTATAACGAAACCATGACAACGGTTCGCTGTGGAAA CCTCATAGTGGAGGGGAGGGAGGAATGTGACTGTGGCTCCTTCAAGCAGTGTTATGCC AGTTATTGCTGCCAAAGTGACTGTCACTTAACACCGGGGAGCATCTGCCATATAGGAG AGTGCTGTACAAACTGCAGCTTCTCCCCACCAGGGACTCTCTGCAGACCTATCCAAAA TATATGTGACCTTCCAGAGTACTGTCACGGGACCACCGTGACATGTCCCGCAAACGTT TATATGCAAGATGGAACCCCGTGCACTGAAGAAGGCTACTGCTATCGTGGGAACTGCA CTGATCGCAATGTGCTCTGCAAGGCGATCTTTGGTGTCAGTGCTGAGGATGCTCCCGA GGTCTGCTATGACATAAATCTTGAAAGCTACCGATTTGGACATTGTATTAGACAACAA ACATATCTCAGCTACCAGGCTTGTGCAGGAATAGATAAGTTTTGTGGAAGACTGCAGT GTACCAATGTGACCCATCTTCCCCGGCTGCAGGAACGTGTTTCATTCCATCACTCAGT GAGAGGAGGGTTTCAGTGTTTTGGACTGGATGAACACCATGCAACAGACACGACTGAT GTTGGGCGTGTGATAGATGGCACTCCTTGTGTTCATGGAAACTTCTGTAATAACACCC AGTGCAATGTGACTATCACTTCACTGGGCTACAACTGCCACCCTCAGAAGTGCGGTCA TAGAGGAGTCTGCAACAACAGAAGGAACTGCCATTGCCATATAGGCTGGGATCCTCCA CTGTGCCTAAGAAGAGGTGCTGGTGGGAGTGTCAACAGCGGGCCACCTCCAAAAAGAA CACGTTCCGTCAAACAAAGCCAGCAATCAGTGATGTATCTGAGAGTGGTCTTTGGTCG TATTTACGCCTTCATAATTGCACTGCTCTTTGGGACAGCCAAAAATGTGCGAACTATC AGGACCACCACCGTTAAGGAAGGGACAGTTACTAACCCTGAA**TAA**CACTAATTCAGCC TCCCGATCCCT
	ORF Start: ATG at 48 | ORF Stop: TAA at 2247
	SEQ ID NO: 20 | 733 aa | MW at 83206.8kD
NOV8a, CG90021-01 Protein Sequence	MRRAEARVTLRTPLLLLGLWALLAPVRCSQGRPLWHYASSEVVIPRKETHHSKGLQFP GWLSYSLCFGFWGQRHVIHMRRKHLLWPRHLLVTTQDDQGVLQMGDPYIPPDCYYLGY LEEVPLSMVTVDTCYGDLRGIMRLDDLAYEIKPLQDSRRFEHVVFQIVAEPNATGPTF RDDDNETDPLFSEANDSMNPRISNSLYSSHRGNIKGHVQCSNSYYRIYGNITTCSKEV VQMFSLIDSIAQNIDLRYYIYLLTIYNNRDPAPVNEYRIQSAMFTYFKTTFFDTFHVH SSTLLIKYVPHESNYEPERYNFCSRIALLHIGTPGRHYLLVAVIITQTQMRSIGLEYD DNYCTCQRRASCIMQRFPGMTDAFSNCSYGHAQNCFIHSGRCVFETLAPVYNETMTTV RCGNLIVEGREECDCGSFKQCYASYCCQSDCHLTPGSICHIGECCTNCSFSPPGTLCR PIQNICDLPEYCHGTTVTCPANVYMQDGTPCTEEGYCYRGNCTDRNVLCKAIFGVSAE DAPEVCYDINLESYRFGHCIRQQTYLSYQACAGIDKFCGRLQCTNVTHLPRLQERVSF HHSVRGGFQCFGLDEHHATDTTDVGRVIDGTPCVHGNFCNNTQCNVTITSLGYNCHPQ KCGHRGVCNNRRNCHCHIGWDPPLCLRRGAGGSVNSGPPPKRTRSVKQSQQSVMYLRV VFGRIYAFIIALLFGTAKNVRTIRTTTVKEGTVTNPE

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

Table 8B. Protein Sequence Properties NOV8a

PSort analysis:	0.4600 probability located in plasma membrane; 0.3000 probability located in lysosome (membrane); 0.2800 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 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
AAU72892	Human metalloprotease partial protein sequence #4 - Homo sapiens, 731 aa. [WO200183782-A2, 08-NOV-2001]	1..733 1..731	639/734 (87%) 669/734 (91%)	0.0
AAE15652	Human disintegrin-like protein, NOV3 - Homo sapiens, 737 aa. [WO200194416-A2, 13-DEC-2001]	1..733 1..737	609/740 (82%) 645/740 (86%)	0.0
AAE14340	Human protease PRTS-5 protein - Homo sapiens, 576 aa. [WO200183775-A2, 08-NOV-2001]	447..733 294..576	253/287 (88%) 263/287 (91%)	e-157
AAB47561	Protease PRTS-3 - Homo sapiens, 559 aa. [WO200171004-A2, 27-SEP-2001]	1..329 1..332	270/335 (80%) 283/335 (83%)	e-153
AAY28655	Human SVPH1-8 protease - Homo sapiens, 722 aa. [WO9936549-A1, 22-JUL-1999]	8..676 8..675	243/692 (35%) 365/692 (52%)	e-123

In a BLAST search of public sequence datbases, 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
Q28484	TESTICULAR METALLOPROTEASE-LIKE, DISINTEGRIN-LIKE, CYSTEINE-RICH PROTEIN IVA - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 732 aa.	1..733 1..732	622/734 (84%) 666/734 (89%)	0.0

Q28485	TESTICULAR METALLOPROTEASE-LIKE, DISINTEGRIN-LIKE, CYSTEINE-RICH PROTEIN IVB - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 713 aa (fragment).	20..733 1..713	603/715 (84%) 651/715 (90%)	0.0
O19050	CELLULAR DISINTEGRIN ADAM 6D - Oryctolagus cuniculus (Rabbit), 731 aa.	4..733 4..731	424/736 (57%) 526/736 (70%)	0.0
O19051	CELLULAR DISINTEGRIN ADAM 6E - Oryctolagus cuniculus (Rabbit), 730 aa.	14..733 10..730	415/724 (57%) 518/724 (71%)	0.0
P70535	TMDC IV PROTEIN - Rattus norvegicus (Rat), 751 aa.	1..720 8..732	385/728 (52%) 487/728 (66%)	0.0

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

Table 8E. Domain Analysis of NOV8a

Pfam Domain	NOV8a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Pep_M12B_propep: domain 1 of 1	77..192	42/119 (35%) 102/119 (86%)	1.3e-44
Reprolysin: domain 1 of 1	216..395	45/210 (21%) 108/210 (51%)	0.00095
metalthio: domain 1 of 1	395..458	14/67 (21%) 32/67 (48%)	7
disintegrin: domain 1 of 1	414..489	32/76 (42%) 44/76 (58%)	2.2e-18

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: 21	1677 bp

NOV9a, CG90709-01 DNA Sequence	TTAATCTTCTGTGGCAGAAATGCA**ATG**GCACATCGTGATTCTGAGATGAAAGAAGAAT GTCTAAGGGAAGACCTGAAGTTTTACTTCATGAGCCCTTGTGAAAAATACCGAGCCAG ACGCCAGATTCCGTGGAAACTGGGTTTGCAGATTTTGAAGATAGTCATGGTCACCACA CAGCTTGTTCGTTTTGGTTTAAGTAACCAGCTGGTGGTTGCTTTCAAAGAAGATAACA CTGTTGCTTTTAAGCACTTGTTTTTGAAAGGATATTCTGGTACAGATGAAGATGACTA CAGCTGCAGTGTATATACTCAAGAGGATGCCTATGAGAGCATCTTTTTTGCTATTAAT CAGTATCATCAGCTAAAGGACATTACCCTGGGGACCCTTGGTTATGGAGAAAATGAAG ACAATAGAATTGGCTTAAAAGTCTGTAAGCAGCATTACAAGAAAGGGACCATGTTTCC TTCTAATGAGACACTGAATATTGACAACGACGTTGAGCTCAACTGTGGGGTTGTGGCG ATATACATTTTAAAGTGTTATTCCCTAAGAGATATTATGACAATTTATACCTTTCAAT ATATTTTATTCAGGCTCTTACAGGTTGAAATCTCCTTTCATCTTAAAGGCATTGACCT ACAGACAATTCATTCCCGTGAGTTACCAGACTGTTATGTCTTTCAGAATACGATTATC TTTGACAATAAAGCTCACAGTGGCAAAATCAAAATCTATTTTGACAGTGATGCCAAAA TTGAAGAATGTAAAGACTTGAACATATTTGGATCTAGTAAGTATGCTCTGGTGTTTGA TGCATTTGTCATTGTGATTTGCTTGGCATCTCTTATTCTGTGTACAAGATCCATTGTT CTTGCTCTAAGGTTACGGAGATTTCTAAATTTCTTCCTGGAGAAGTACAAGCGGCCTG TGTGTGACACCGACCAGTGGGAGTTCATCAACGGCTGGTATGTCCTGGTGATTATCAG CGACCTAATGACAATCATTGGCTCCATATTAAAAATGGAAATCAAAGCAAAGAATCTC ACAAACTATGATCTCTGCAGCATTTTTCTTGGAACCTCTACGCTCTTGGTTTGGGTTG GAGTCATCAGATACCTGGGTTATTTCCAGGCATATAATGTACTGATTTTAACAATGCA GGCCTCACTGCCAAAAGTTCTTCGGTTTTGTGCTTGTGCTGGTATGATTTATCTGGGT TACACATTCTGTGGCTGGATTGTCTTAGGACCATACCATCTACAGTTTGAAAATCTGA ACACAGTTGCTGAGTGTCTGTTTTCTCTGGTCAACGGTGATGACATGTTTGCAACCTT TGCCCAAATCCAGCAGAAGAGCATCTTGGTGTGGCTGTTCAGTCGTCTGTATTTATAT TCCTTCATCAGCCTTTTTATATATATGATTCTCAGTCTTTTTATTGCACTTATTACAG ATTCTTATGACACCATTAAGAAATTCCAACAGAATGGGTTTCCTGAAACGGATTTGCA GGAATTCCTGAAGGAATGCAGTAGCAAAGAAGAGTATCAGAAAGAGTCCTCAGCCTTC CTGTCCTGCATCTGCTGTCGGAGGAGGTCAGTATCATGTTTATTCTCCATGCTCCTGA GATGGGCTGTTCTGTTGTCT**TAA**GAAAGAGCCCCTCCAAGATTACCATTACAT
	ORF Start: ATG at 25 / ORF Stop: TAA at 1645
	SEQ ID NO: 22 / 540 aa / MW at 62760.5kD
NOV9a, CG90709-01 Protein Sequence	MAHRDSEMKEECLREDLKFYFMSPCEKYRARRQIPWKLGLQILKIVMVTTQLVRFGLS NQLVVAFKEDNTVAFKHLFLKGYSGTDEDDYSCSVYTQEDAYESIFFAINQYHQLKDI TLGTLGYGENEDNRIGLKVCKQHYKKGTMFPSNETLNIDNDVELNCGVVAIYILKCYS LRDIMTIYTFQYILFRLLQVEISFHLKGIDLQTIHSRELPDCYVFQNTIIFDNKAHSG KIKIYFDSDAKIEECKDLNIFGSSKYALVFDAFVIVICLASLILCTRSIVLALRLRRF LNFFLEKYKRPVCDTDQWEFINGWYVLVIISDLMTIIGSILKMEIKAKNLTNYDLCSI FLGTSTLLVWVGVIRYLGYFQAYNVLILTMQASLPKVLRFCACAGMIYLGYTFCGWIV LGPYHLQFENLNTVAECLFSLVNGDDMFATFAQIQQKSILVWLFSRLYLYSFISLFIY MILSLFIALITDSYDTIKKFQQNGFPETDLQEFLKECSSKEEYQKESSAFLSCICCRR RSVSCLFSMLLRWAVLLS
	SEQ ID NO: 23 / 1671 bp
NOV9b, CG90709-02 DNA Sequence	TTAAAATTAATCTTCTGTGGCAGAAATGCA**ATG**GCACATCGTGATTCTGAGATGAAAG AAGAATGTCTAAGGGAAGACCTGAAGTTTTACTTCATGAGCCCTTGTGAAAAATACCG AGCCAGACGCCAGATTCCGTGGAAACTGGGTTTGCAGATTTTGAAGATAGTCATGGTC ACCACACAGCTTGTTCGTTTTGGTTTAAGTAACCAGCTGGTGGTTGCTTTCAAAGAAG ATAACACTGTTGCTTTTAAGCACTTGTTTTTGAAAGGATATTCTGGTACAGATGAAGA TGACTACAGCTGCAGTGTATATACTCAAGAGGATGCCTATGAGAGCATCTTTTTTGCT ATTAATCAGTATCATCAGCTAAAGGACATTACCCTGGGGACCCTTGGTTATGGAGAAA ATGAAGACAATAGAATTGGCTTAAAAGTCTGTAAGCAGCATTACAAGAAAGGGACCAT GTTTCCTTCTAATGAGACACTGAATATTGACAACGACGTTGAGCTAGATTGTGTTCAA TTAGACCTTCAGGACCTCTCCAAGAAGCCTCCGGACTGGAAGAACTCATCATTCTTCA GACTGGAATTTTATCGGCTCTTACAGGTTGAAATCTCCTTTCATCTTAAAGGCATTGA CCTACAGACAATTCATTCCCGTGAGTTACCAGACTGTTATGTCTTTCAGAATACGATT ATCTTTGACAATAAAGCTCACAGTGGCAAAATCAAAATCTATTTTGACAGTGATGCCA AAATTGAAGAATGTAAAGACTTGAACATATTTGGATCAGCTCAGAAAAATGCTCAGTA TGTCCTGGTGTTTGATGCATTTGTCATTGTGATTTGCTTGGCATCTCTTATTCTGTGT ACAAGATCCATTGTTCTTGCTCTAAGGTTACGGAAGAGATTTCTAAATTTCTTCCTGG

	AGAAGTACAAGCGGCCTGTGTGTGACACCGACCAGTGGGAGTTCATCAACGGCTGGTA TGTCCTGGTGATTATCAGCGACCTAATGACAATCATTGGCTCCATATTAAAAATGGAA ATCAAAGCAAAGAATCTCACAAACTATGATCTCTGCAGCATTTTTCTTGGAACCTCTA CGCTCTTGGTTTGGGTTGGAGTCATCAGATACCTGGGTTATTTCCAGGCATATAATGT ACTGATTTTAACAATGCAGGCCTCACTGCCAAAAGTTCTTCGGTTTTGTGCTTGTGCT GGTATGATTTATCTGGGTTACACATTCTGTGGCTGGATTGTCTTAGGACCATACCATG ACAAGTTTGAAAATCTGAACACAGTTGCTGAGTGTCTGTTTTCTCTGGTCAACGGTGA TGACATGTTTGCAACCTTTGCCCAAATCCAGCAGAAGAGCATCTTGGTGTGGCTGTTC AGTCGTCTGTATTTATATTCCTTCATCAGCCTTTTTATATATATGATTCTCAGTCTTT TTATTGCACTTATTACAGATTCTTATGACACCATTAAGAAATTCCAACAGAATGGGTT TCCTGAAACGGATTTGCAGGAATTCCTGAAGGAATGCAGTAGCAAAGAAGAGTATCAG AAAGAGTCCTCAGCCTTCCTGTCCTGCATCTGCTGTCGGAGGAGGTCAGTATCATGTT TATTCTCCATGCTCCTGAGATGGGCTGTTCTGTTGTCT**TAA**GAAAGA		
	ORF Start: ATG at 31	ORF Stop: TAA at 1663	
	SEQ ID NO: 24	544 aa	MW at 63298.8kD
NOV9b, CG90709-02 Protein Sequence	MAHRDSEMKEECLREDLKFYFMSPCEKYRARRQIPWKLGLQILKIVMVTTQLVRFGLS NQLVVAFKEDNTVAFKHLFLKGYSGTDEDDYSCSVYTQEDAYESIFFAINQYHQLKDI TLGTLGYGENEDNRIGLKVCKQHYKKGTMFPSNETLNIDNDVELDCVQLDLQDLSKKP PDWKNSSFFRLEFYRLLQVEISFHLKGIDLQTIHSRELPDCYVFQNTIIFDNKAHSGK IKIYFDSDAKIEECKDLNIFGSAQKNAQYVLVFDAFVIVICLASLILCTRSIVLALRL RKRFLNFFLEKYKRPVCDTDQWEFINGWYVLVIISDLMTIIGSILKMEIKAKNLTNYD LCSIFLGTSTLLVWVGVIRYLGYFQAYNVLILTMQASLPKVLRFCACAGMIYLGYTFC GWIVLGPYHDKFENLNTVAECLFSLVNGDDMFATFAQIQQKSILVWLFSRLYLYSFIS LFIYMILSLFIALITDSYDTIKKFQQNGFPETDLQEFLKECSSKEEYQKESSAFLSCI CCRRRSVSCLFSMLLRWAVLLS		
	SEQ ID NO: 25	2130 bp	
NOV9c, CG90709-03 DNA Sequence	TTAAAATTAATCTTCTGTGGCAGAAATGCA**ATG**GCACATCGTGATTCTGAGATGAAAG AAGAATGTCTAAGGGAAGACCTGAAGTTTTACTTCATGAGCCCTTGTGAAAAATACCG AGCCAGACGCCAGATTCCGTGGAAACTGGGTTTGCAGATTTTGAAGATAGTCATGGTC ACCACACAGCTTGTTCGTTTTGGTTTAAGTAACCAGCTGGTGGTTGCTTTCAAAGAAG ATAACACTGTTGCTTTTAAGCACTTGTTTTTGAAAGGATATTCTGGTACAGATGAAGA TGACTACAGCTGCAGTGTATATACTCAAGAGGATGCCTATGAGAGCATCTTTTTTGCT ATTAATCAGTATCATCAGCTAAAGGACATTACCCTGGGGACCCTTGGTTATGGAGAAA ATGAAGACAATAGAATTGGCTTAAAAGTCTGTAAGCAGCATTACAAGAAAGGGACCAT GTTTCCTTCTAATGAGACACTGAATATTGACAACGACGTTGAGCTAGATTGTGTTCAA TTAGACCTTCAGGACCTCTCCAAGAAGCCTCCGGACTGGAAGAACTCATCATTCTTCA GACTGGAATTTTATCGGCTCTTACAGGTTGAAATCTCCTTTCATCTTAAAGGCATTGA CCTACAGACAATTCATTCCCGTGAGTTACCAGACTGTTATGTCTTTCAGAATACGATT ATCTTTGACAATAAAGCTCACAGTGGCAAAATCAAAATCTATTTTGACAGTGATGCCA AAATTGAAGAATGTAAAGACTTGAACATATTTGGATCAGCTCAGAAAAATGCTCAGTA TGTCCTGGTGTTTGATGCATTTGTCATTGTGATTTGCTTGGCATCTCTTATTCTGTGT ACAAGATCCATTGTTCTTGCTCTAAGGTTACGGAAGAGATTTCTAAATTTCTTCCTGG AGAAGTACAAGCGGCCTGTGTGTGACACCGACCAGTGGGAGTTCATCAACGGCTGGTA TGTCCTGGTGATTATCAGCGACCTAATGACAATCATTGGCTCCATATTAAAAATGGAA ATCAAAGCAAAGAATCTCACAAACTATGATCTCTGCAGCATTTTTCTTGGAACCTCTA CGCTCTTGGTTTGGGTTGGAGTCATCAGATACCTGGGTTATTTCCAGGCATATAATGT ACTGATTTTAACAATGCAGGCCTCACTGCCAAAAGTTCTTCGGTTTTGTGCTTGTGCT GGTATGATTTATCTGGGTTACACATTCTGTGGCTGGATTGTCTTAGGACCATACCATG ACAAGTTTGAAAATCTGAACACAGTTGCTGAGTGTCTGTTTTCTCTGGTCAACGGTGA TGACATGTTTGCAACCTTTGCCCAAATCCAGCAGAAGAGCATCTTGGTGTGGCTGTTC AGTCGTCTGTATTTATATTCCTTCATCAGCCTTTTTATATATATGATTCTCAGTCTTT TTATTGCACTTATTACAGATTCTTATGACACCATTAAGAAATTCCAACAGAATGGGTT TCCTGAAACGGATTTGCAGGAATTCCTGAAGGAATGCAGTAGCAAAGAAGAGTATCAG AAAGAGTCCTCAGCCTTCCTGTCCTGCATCTGCTGTCGGAGGAGGAAAAGAAGTGATG ATCACTTGATACCTATTAGC**TAA**AGTTCTGCTAAAGATGATTAAAGTTCAGGCATCCT TATCCAGCAGCTGAGCAGAGGAACCCCAAATGACTTGGACAAGCAGTTCCAAAATGAC TCTCTTATTTAATGTGGAGTGGGAAAGAGGACTCACAGTTAGCCAGCTGACCATGACT GAAGTTCCAGCTTTACTTGTTATAAAACTTGAATGATAAAGAATAGACCATGGGCTAC		

	TACTGGGCATTAGTGCAATATAACCAGCCGATAATAAAATTTCTCTATTAGTCTGTTA CTTTATGACATGATCTCGGAATGGCAAAGATTCATTTCCAGAAGTGTGCGAAATAATA GTTCTTACCCTGTTAATTACACATTGTGCGTCCTCGGCCCCAAGGGACTGGCACAAAG GGAACTGCGGTGGAAAACATTTGTTAATACCGGGCTCGGTCACAAAAGACCCGGTGGG GCATCCATTTAAGAGTCACGGGCGAACTACACGGGCAAGACC		
	ORF Start: ATG at 31	ORF Stop: TAA at 1645	
	SEQ ID NO: 26	538 aa	MW at 62653.9kD
NOV9c, CG90709-03 Protein Sequence	MAHRDSEMKEECLREDLKFYFMSPCEKYRARRQIPWKLGLQILKIVMVTTQLVRFGLS NQLVVAFKEDNTVAFKHLFLKGYSGTDEDDYSCSVYTQEDAYESIFFAINQYHQLKDI TLGTLGYGENEDNRIGLKVCKQHYKKGTMFPSNETLNIDNDVELDCVQLDLQDLSKKP PDWKNSSFFRLEFYRLLQVEISFHLKGIDLQTIHSRELPDCYVFQNTIIFDNKAHSGK IKIYFDSDAKIEECKDLNIFGSAQKNAQYVLVFDAFVIVICLASLILCTRSIVLALRL RKRFLNFFLEKYKRPVCDTDQWEFINGWYVLVIISDLMTIIGSILKMEIKAKNLTNYD LCSIFLGTSTLLVWVGVIRYLGYFQAYNVLILTMQASLPKVLRFCACAGMIYLGYTFC GWIVLGPYHDKFENLNTVAECLFSLVNGDDMFATFAQIQQKSILVWLFSRLYLYSFIS LFIYMILSLFIALITDSYDTIKKFQQNGFPETDLQEFLKECSSKEEYQKESSAFLSCI CCRRRKRSDDHLIPIS		
	SEQ ID NO: 27	2067 bp	
NOV9d, CG90709-04 DNA Sequence	ACGCGTTACGGGGAGGGGCGAAATGAGTCGGCCGTGAACGGTGTTTCCTGTTCCGAAT CCCGAGACCCCTGGAAAGTTTTGAAGGAGGAGGC**ATG**GCCCGGCAGCCTTATCGTTTT CCCCAGGCAAGGATTCCGGAGAGAGGATCAGGTGTTTTCAGGTTAACCGTCAGAAATG CAATGGCACATCGTGATTCTGAGATGAAAGAAGAATGTCTAAGGGAAGACCTGAAGTT TTACTGCATGAGCCCTTGTGAAAAATACCGAGCCAGACGCCAGATTCCGTGGAAACTG GGTTTGCAGATTTTGAAGATAGTCATGGTCACCACACAGCTTGTTCGTTTTGGTTTAA GTAACCAGCTGGTGGTTGCTTTCAAAGAAGATAACACTGTTGCTTTTAAGCACTTGTT TTTGAAAGGATATTCTGGTACAGATGAAGATGACTACAGCTGCAGTGTATATACTCAA GAGGATGCCTATGAGAGCATCTTTTTTGCTATTAATCAGTATCATCAGCTAAAGGACA TTACCCTGGGGACCCTTGGTTATGGAGAAAATGAAGACAATAGAATTGGCTTAAAAGT CTGTAAGCAGCATTACAAGAAAGGGACCATGTTTCCTTCTAATGAGACACTGAATATT GACAACGACGTTGAGCTAGATTGTGTTCAATTAGACCTTCAGGACCTCTCCAAGAAGC CTCCGGACTGGAAGAACTCATCATTCTTCAGACTGGAATTTTATCGGCTCTTACAGGT TGAAATCTCCTTTCATCTTAAAGGCATTGACCTACAGACAATTCATTCCCGTGAGTTA CCAGACTGTTATGTCTTTCAGAATACGATTATCTTTGACAATAAAGCTCACAGTGGCA AAATCAAAATCTATTTTGACAGTGATGCCAAAATTGAAGAATGTAAAGACTTGAACAT ATTTGGATCAGCTCAGAAAAATGCTCAGTATGTCCTGGTGTTTGATGCATTTGTCATT GTGATTTGCTTGGCATCTCTTATTCTGTGTACAAGATCCATTGTTCTTGCTCTAAGGT TACGGAAGAGATTTCTAAATTTCTTCCTGGAGAAGTACAAGCGGCCTGTGTGTGACAC CGACCAGTGGGAGTTCATCAACGGCTGGTATGTCCTGGTGATTATCAGCGACCTAATG ACAATCATTGGCTCCATATTAAAAATGGAAATCAAAGCAAAGAATCTCACAAACTATG ATCTCTGCAGCATTTTTCTTGGAACCTCTACGCTCTTGGTTTGGGTTGGAGTCATCAG ATACCTGGGTTATTTCCAGGCATATAATGTACTGATTTTAACAATGCAGGCCTCACTG CCAAAAGTTCTTCGGTTTTGTGCTTGTGCTGGTATGATTTATCTGGGTTACACATTCT GTGGCTGGATTGTCTTAGGACCATACCATGACAAGTTTGAAAATCTGAACACAGTTGC TGAGTGTCTGTTTTCTCTGGTCAACGGTGATGACATGTTTGCAACCTTTGCCCAAATC CAGCAGAAGAGCATCTTGGTGTGGCTGTTCAGTCGTCTGTATTTATATTCCTTCATCA GCCTTTTTATATATATGATTCTCAGTCTTTTTATTGCACTTATTACAGATTCTTATGA CACCATTAAGAAATTCCAACAGAATGGGTTTCCTGAAACGGATTTGCAGGAATTCCTG AAGGAATGCAGTAGCAAAGAAGAGTATCAGAAAGAGTCCTCAGCCTTCCTGTCCTGCA TCTGCTGTCGGAGGAGGAAAAGAAGTGATGATCACTTGATACCTATTAGC**TAA**AGTTC TGCTAAAGATGATTAAAGTTCAGGCATCCTTATCCAGCAGCTGAGCAGAGGAACCCCA AATGACTTGGACAAGCAGTTCCAAAATGACTCTCTTATTTAATTGTGGAGTGGGAAAG AGGACTCACAGTTAGCCAGCTGACCATGACTGAAGTTCCAGCTTTACTTTTTATAAAC TTGAATGATAAAGAATAGACCATGGGCTACTACTGGGCATTAGTGCAATATAACAGCG ATAATAAAATTCTCTATTAGTCTGTTAATTTATGAAA		
	ORF Start: ATG at 93	ORF Stop: TAA at 1791	
	SEQ ID NO: 28	566 aa	MW at 65866.6kD

NOV9d, CG90709-04 Protein Sequence	MARQPYRFPQARIPERGSGVFRLTVRNAMAHRDSEMKEECLREDLKFYCMSPCEKYRA RRQIPWKLGLQILKIVMVTTQLVRFGLSNQLVVAFKEDNTVAFKHLFLKGYSGTDEDD YSCSVYTQEDAYESIFFAINQYHQLKDITLGTLGYGENEDNRIGLKVCKQHYKKGTMF PSNETLNIDNDVELDCVQLDLQDLSKKPPDWKNSSFFRLEFYRLLQVEISFHLKGIDL QTIHSRELPDCYVFQNTIIFDNKAHSGKIKIYFDSDAKIEECKDLNIFGSAQKNAQYV LVFDAFVIVICLASLILCTRSIVLALRLRKRFLNFFLEKYKRPVCDTDQWEFINGWYV LVIISDLMTIIGSILKMEIKAKNLTNYDLCSIFLGTSTLLVWVGVIRYLGYFQAYNVL ILTMQASLPKVLRFCACAGMIYLGYTFCGWIVLGPYHDKFENLNTVAECLFSLVNGDD MFATFAQIQQKSILVWLFSRLYLYSFISLFIYMILSLFIALITDSYDTIKKFQQNGFP ETDLQEFLKECSSKEEYQKESSAFLSCICCRRRKRSDDHLIPIS

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

Table 9B. Comparison of NOV9a against NOV9b through NOV9d.

Protein Sequence	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV9b	1..540 1..544	485/545 (88%) 493/545 (89%)
NOV9c	1..526 1..530	481/531 (90%) 489/531 (91%)
NOV9d	1..526 29..558	480/531 (90%) 488/531 (91%)

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

Table 9C. Protein Sequence Properties NOV9a

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:	Cleavage site between residues 65 and 66

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

Table 9D. Geneseq Results for NOV9a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV9a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAM51858	Human TRP-like calcium channel TLCC-2 - Homo sapiens, 580 aa. [WO200177331-A1, 18-OCT-2001]	9..533 37..579	262/550 (47%) 374/550 (67%)	e-142
AAB74707	Human membrane associated protein MEMAP-13 - Homo sapiens, 580 aa. [WO200112662-A2, 22-FEB-2001]	9..533 37..579	262/550 (47%) 374/550 (67%)	e-142
AAB93412	Human protein sequence SEQ ID NO:12616 - Homo sapiens, 497 aa. [EP1074617-A2, 07-FEB-2001]	109..523 76..497	241/426 (56%) 318/426 (74%)	e-139
AAB08906	Human secreted protein sequence encoded by gene 16 SEQ ID NO:63 - Homo sapiens, 511 aa. [WO200017222-A1, 30-MAR-2000]	42..533 1..510	244/517 (47%) 349/517 (67%)	e-131
ABB11279	Human secreted protein homologue, SEQ ID NO:1649 - Homo sapiens, 164 aa. [WO200157188-A2, 09-AUG-2001]	334..497 1..164	161/164 (98%) 163/164 (99%)	3e-90

In a BLAST search of public sequence datbases, 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**
Q9CQD3	3300002C04RIK PROTEIN - Mus musculus (Mouse), 538 aa.	1..523 1..527	426/528 (80%) 466/528 (87%)	0.0

AAL84622	MUCOLIPIN-3 - Homo sapiens (Human), 553 aa.	10..523 35..553	304/525 (57%) 396/525 (74%)	e-177
AAL84623	MUCOLIPIN-3 - Mus musculus (Mouse), 553 aa.	3..523 23..553	306/534 (57%) 398/534 (74%)	e-176
Q9H4B3	MUCOLIPIDIN - Homo sapiens (Human), 580 aa.	9..533 37..579	262/550 (47%) 374/550 (67%)	e-142
Q9GZU1	CDNA: FLJ22449 FIS, CLONE HRC09609 (MUCOLIPIN) (MUCOLIPIDOSIS TYPE IV PROTEIN) (MUCOLIPIN 1) - Homo sapiens (Human), 580 aa.	9..533 37..579	262/550 (47%) 374/550 (67%)	e-142

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

Table 9F. Domain Analysis of NOV9a

Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
DUF214: domain 1 of 1	231..407	30/267 (11%) 117/267 (44%)	9.2
ion_trans: domain 1 of 1	314..474	31/236 (13%) 117/236 (50%)	0.01

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: 29	642 bp	
NOV10a, CG90739-01 DNA Sequence	CACCAAGGGCCTGGCACCAGGTGCCCAGTCTTCCAGTTGCGAGGGCAAGCAAACCCGT CATGAGCAACTCCCTTCCCCATCTCTGCTCACC**ATG**TGGACGCTGAAATCGTCCCTGG TCCTGCTTCTGTGCCTCACCTGCAGCTATGCCTTTATGTTCTCTTCTCTGAGACAGAA AACTAGCGAACCCCAGGGGAAGGTGCCGTGTGGAGAGCACTTTCGGATTCGGCAGAAC CTACCAGAGCACACCCAAGGCTGGCTTGGGAGCAAATGGCTCTGGCTTTTGTTTGCTG TTGTGCCGTTTGTGATACTGCAGTGTCAAAGAGACAGTGAGAAGAATAAGGAGCAGAG TCCTCCTGGCCTTCGAGGCTTCCCATTTCGCACTCCACTAAAGAAAAATCAAAATGCT TCTCTTTACAAAGACTGTGTATTCAATACCTTAAACGAACTTGAAGTGGAGCTTTTGA AATTTGTGTCCGAAGTGCAGAATCTTAAAGGTGCCATGGCAACAGGCAGTGGCAGTAA CCTCAAGCTTCGAAGGTCAGAGATGCCTGCAGATCCATACCATGTCACAATCTGTAAA ATATGGGGAGAAGAAAGCTCTAGC**TGA**ATGGATTTGTGTGTCAGGAGAGAAAAAAGTT GAGT		
	ORF Start: ATG at 92	ORF Stop: TGA at 605	
	SEQ ID NO: 30	171 aa	MW at 19498.4kD

NOV10a, CG90739-01 Protein Sequence	MWTLKSSLVLLLCLTCSYAFMFSSLRQKTSEPQGKVPCGEHFRIRQNLPEHTQGWLGS KWLWLLFAVVPFVILQCQRDSEKNKEQSPPGLRGFPFRTPLKKNQNASLYKDCVFNTL NELEVELLKFVSEVQNLKGAMATGSGSNLKLRRSEMPADPYHVTICKIWGEESSS		
	SEQ ID NO: 31	141 bp	
NOV10b, 172390256 DNA Sequence	**GGA**TCCTTTATGTTCTCTTCTCTGAGACAGAAAACTAGCGAACCCCAGGGGAAGGTGC CGTGTGGAGAGCACTTTCGGATTCGGCAGAACCTACCAGAGCACACCCAAGGCTGGCT TGGGAGCAAATGGCTCTGGCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 32	47 aa	MW at 5515.2kD
NOV10b, 172390256 Protein Sequence	GSFMFSSLRQKTSEPQGKVPCGEHFRIRQNLPEHTQGWLGSKWLWLE		
	SEQ ID NO: 33	141 bp	
NOV10c, 172390440 DNA Sequence	**GGA**TCCTTTATGTTCTCTTCTCTGAGACAGAAAACTAGCGAACCCCAGGGGAAGGTGC AATACGGAGAGCACTTTCGGATTCGGCAGAATCTACCAGAGCACACCCAAGGCTGGCT TGGGAGCAAATGGCTCTGGCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 34	47 aa	MW at 5606.3kD
NOV10c, 172390440 Protein Sequence	GSFMFSSLRQKTSEPQGKVQYGEHFRIRQNLPEHTQGWLGSKWLWLE		
	SEQ ID NO: 35	468 bp	
NOV10d, 172390569 DNA Sequence	**GGA**TCCTTTATGTTCTCTTCTCTGAGACAGAAAACTAGCGAACCCCAGGGGAAGGTGC CGTGTGGAGAGCACTTTCGGATTCGGCAGAACCTACCAGAGCACACCCAAGGCTGGCT TGGGAGCAAATGGCTCTGGCTTTTGTTTGCTGTTGTGCCGTTTGTGATACTGAAGTGT CAAAGAGACAGTGAGAAGAATAAGGAGCAGAGTCCTCCTGGCCTTCGAGGCTTCCCAT TTCGCACTCCACTAAAGAAAAATCAAAATGCTTCTCTTTACAAAGACTGTGTATTCAA TACCTTAAACGAACTTGAAGTGGAGCTTTTGAAATTTGTGTCCGAAGTGCAGAATCTT AAAGGTGCCATGGCAACAGGCAGTGGCAGTAACCTCAAGCTTCGAAGGTCAGAGATGC CTGCAGATCCATACCATGTCACAATCTGTAAAATATGGGGAGAAGAAAGCTCTAGCCT CGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 36	156 aa	MW at 17757.3kD
NOV10d, 172390569 Protein Sequence	GSFMFSSLRQKTSEPQGKVPCGEHFRIRQNLPEHTQGWLGSKWLWLLFAVVPFVILKC QRDSEKNKEQSPPGLRGFPFRTPLKKNQNASLYKDCVFNTLNELEVELLKFVSEVQNL KGAMATGSGSNLKLRRSEMPADPYHVTICKIWGEESSSLE		
	SEQ ID NO: 37	468 bp	
NOV10e, 172390587 DNA Sequence	**GGA**TCCTTTATGTTCTCTTCTCTGAGACAGAAAACTAGCGAACCCCAGGGGAAGGTGC CGTGTGGAGAGCACTTTCGGATTCGGCAGAACCTACCAGAGCACACCCAAGGCTGGCT TGGGAGCAAATGGCTCTGGCTTTTGTTTGCTGTTGTGCCGTTTGTGATACTGCAGTGT CAAAGAGACAGTGAGAAGAATAAGGAGCAGAGTCCTCCTGGCCTTCGAGGCTTCCCAT TTCGCACTCCACTAAAGAAAAATCAAAATGCTTCTCTTTACAAAGACTGTGTATTCAA TACCTTAAACGAACTTGAAGTGGAGCTTTTGAAATTTGTGTCCGAAGTGCAGAATCTT AAAGGTGCCATGGCAACAGGCAGTGGCAGTAACCTCAAGCTTCGAAGGTCAGAGATGC CTGCAGATCCATACCATGTCACAATCTGTAAAATATGGGGAGAAGAAAGCTCTAGCCT CGAG		
	ORF Start: at 1	ORF Stop: end of sequence	

	SEQ ID NO: 38	156 aa	MW at 17757.2kD
NOV10e, 172390587 Protein Sequence	GSFMFSSLRQKTSEPQGKVPCGEHFRIRQNLPEHTQGWLGSKWLWLLFAVVPFVILQC QRDSEKNKEQSPPGLRGFPFRTPLKKNQNASLYKDCVFNTLNELEVELLKFVSEVQNL KGAMATGSGSNLKLRRSEMPADPYHVTICKIWGEESSSLE		
	SEQ ID NO: 39	468 bp	
NOV10f, 172390603 DNA Sequence	**GGA**TCCTTTATGTTCTCTTCTCTGAGACAGAAAACTAGCGAACCCCAGGGGAAGGTGC CGTGTGGAGAGCACTTTCGGATTCGGCAGAACCTACCAGAGCACACCCAAGGCTGGCT TGGGAGCAAATGGCTCTGGCTTTTGTTTGCTGTTGTGCCGTTTGTGATACTGAAGTGT CAAAGAGACAGTGAGAAGAATAAGGAGCAGAGTCCTCCTGGCCTTCGAGGCTTCCCAT TTCGCACTCCACTAAAGAAAAATCAAAATGCTTCTCTTTACAAAGACTGTGTATTCAA TACCTTAAACGAACTTGAAGTGGAGCTTTTGAAATTTGTGTCCGAAGTGCAGAACCTT AAAGGTGCCATGGCAACAGGCAGTGGCAGTAACCTCAAGCTTCGAAGGTCAGAGATGC CTGCAGATCCATACCATGTCACAATCTGTAAAATATGGGGAGAAGAAAGCTCTAGCCT CGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 40	156 aa	MW at 17757.3kD
NOV10f, 172390603 Protein Sequence	GSFMFSSLRQKTSEPQGKVPCGEHFRIRQNLPEHTQGWLGSKWLWLLFAVVPFVILKC QRDSEKNKEQSPPGLRGFPFRTPLKKNQNASLYKDCVFNTLNELEVELLKFVSEVQNL KGAMATGSGSNLKLRRSEMPADPYHVTICKIWGEESSSLE		
	SEQ ID NO: 41	468 bp	
NOV10g, 172390624 DNA Sequence	**GGA**TCCTTTATGTTCTCTTCTCTGAGACAGAAAACTAGCGAACCCCAGGGGAAGGTGC CGTGTGGAGAGCACTTTCGGATTCGGCAGAACCTACCAGAGCACACCCAAGGCTGGCT TGGGAGCAAATGGCTCTGGCTTTTGTTTGCTGTTGTGCCGTTTGTGATACTGAAGTGT CAAAGAGACAGTGAGAAGAATAAGGAGCAGAGTCCTCCTGGCCTTCGAGGCTTCCCAT TTCGCATTCCACTAAAGAAAAATCAAAATGCTTCTCTTTACAAAGACTGTGTATTCAA TACCTTAAACGAACTTGAAGTGGAGCTTTTGAAATTTGTGTCCGAAGTGCAAAATCTT AAAGGTGCCATGGCAACAGGCAGTGGCAGTAACCTCAAGCTTCGAAGGTCAGAGATGC CTGCAGATCCATACCATGTCACAATCTGTAAAATATGGGGAGAAGAAAGCTCTAGCCT CGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 42	156 aa	MW at 17769.3kD
NOV10g, 172390624 Protein Sequence	GSFMFSSLRQKTSEPQGKVPCGEHFRIRQNLPEHTQGWLGSKWLWLLFAVVPFVILKC QRDSEKNKEQSPPGLRGFPFRIPLKKNQNASLYKDCVFNTLNELEVELLKFVSEVQNL KGAMATGSGSNLKLRRSEMPADPYHVTICKIWGEESSSLE		
	SEQ ID NO: 43	468 bp	
NOV10h, 172390644 DNA Sequence	**GGA**TCCTTTATGTTCTCTTCTCTGAGACAGAAAACTAGCGAACCCCAGGGGAAGGTGC AATACGGAGAGCACTTTCGGATTCGGCAGAACCTACCAGAGCACACCCAAGGCTGGCT TGGGAGCAAATGGCTCTGGCTTTTGTTTGCTGTTGTGCCGTTTGTGATACTGAAGTGT CAAAGAGACAGTGAGAAGAATAAGGAGCAGAGTCCTCCTGGCCTTCGAGGCTTCCCAT TTCGCACTCCACTAAAGAAAAATCAAAATGCTTCTCTTTACAAAGACTGTGTATTCAA TACCTTAAACGAACTTGAAGTGGAGCTTTTGAAATTTGTGTCCGAAGTGCAAAATCTT AAAGGTGCCATGGCAACAGGCAGTGGCAGTAACCTCAAGCTTCGAAGGTCAGAGATGC CTGCAGATCCATACCATGTCACAATCTGTAAAATATGGGGAGAAGAAAGCTCTAGCCT CGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 44	156 aa	MW at 17848.3kD
NOV10h, 172390644 Protein Sequence	GSFMFSSLRQKTSEPQGKVQYGEHFRIRQNLPEHTQGWLGSKWLWLLFAVVPFVILKC QRDSEKNKEQSPPGLRGFPFRTPLKKNQNASLYKDCVFNTLNELEVELLKFVSEVQNL KGAMATGSGSNLKLRRSEMPADPYHVTICKIWGEESSSLE		

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

Table 10B. Comparison of NOV10a against NOV10b through NOV10h.

Protein Sequence	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV10b	19..63 2..46	44/45 (97%) 45/45 (99%)
NOV10c	19..63 2..46	42/45 (93%) 43/45 (95%)
NOV10d	19..171 2..154	151/153 (98%) 153/153 (99%)
NOV10e	19..171 2..154	152/153 (99%) 153/153 (99%)
NOV10f	19..171 2..154	151/153 (98%) 153/153 (99%)
NOV10g	19..171 2..154	150/153 (98%) 152/153 (99%)
NOV10h	19..171 2..154	149/153 (97%) 151/153 (98%)

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

Table 10C. Protein Sequence Properties NOV10a

PSort analysis:	0.4600 probability located in plasma membrane; 0.1031 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 20 and 21

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

Table 10D. Geneseq Results for NOV10a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV10a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU68550	Human novel cytokine encoded by cDNA 790CIP2D_11 #1 - Homo sapiens, 239 aa. [WO200175093-A1, 11-OCT-2001]	1..171 69..239	150/171 (87%) 158/171 (91%)	5e-85
AAY53032	Human secreted protein clone di393_2 protein sequence SEQ ID NO:70 - Homo sapiens, 171 aa. [WO9957132-A1, 11-NOV-1999]	1..171 1..171	150/171 (87%) 158/171 (91%)	5e-85
AAG00463	Human secreted protein, SEQ ID NO: 4544 - Homo sapiens, 101 aa. [EP1033401-A2, 06-SEP-2000]	1..101 1..101	92/101 (91%) 93/101 (91%)	2e-49
AAY12683	Human 5' EST secreted protein SEQ ID NO:273 - Homo sapiens, 101 aa. [WO9906549-A2, 11-FEB-1999]	1..101 1..101	92/101 (91%) 93/101 (91%)	2e-49
AAM87953	Human immune/haematopoietic antigen SEQ ID NO:15546 - Homo sapiens, 89 aa. [WO200157182-A2, 09-AUG-2001]	83..171 1..89	70/89 (78%) 79/89 (88%)	1e-34

In a BLAST search of public sequence datbases, the NOV10a protein was found to have homology to the proteins shown in the BLASTP data in Table 10E.

Table 10E. Public BLASTP Results for NOV10a				
Protein Accession Number	**Protein/Organism/Length**	**NOV10a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9HCV6	DJ1153D9.4 (NOVEL PROTEIN) - Homo sapiens (Human), 138 aa (fragment).	34..171 1..138	138/138 (100%) 138/138 (100%)	7e-79

Q9D9T2	1700029J11RIK PROTEIN - Mus musculus (Mouse), 170 aa.	4..170 5..169	99/168 (58%) 122/168 (71%)	2e-46
Q96C09	SIMILAR TO NEURONAL THREAD PROTEIN - Homo sapiens (Human), 106 aa.	1..88 1..88	83/88 (94%) 85/88 (96%)	9e-45
Q9HCV7	DJ1153D9.3 (NOVEL PROTEIN) - Homo sapiens (Human), 94 aa.	1..86 1..86	81/86 (94%) 81/86 (94%)	4e-42
Q9CRL6	2810426N06RIK PROTEIN - Mus musculus (Mouse), 300 aa (fragment).	13..51 188..232	17/45 (37%) 24/45 (52%)	4.4

No significant matches were found in a PFam analysis of the NOV10a protein.

Example 11.

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: 45	1152 bp	
NOV11a, CG91667-01 DNA Sequence	**ATG**ACCGCGACCGAAGCCCTCCTGCGCGTCCTCTTGCTCCTGCTGGCTTTCGGCCACA GCACCTATGGGGCTGAATGCTTCCCGGCCTGCAACCCCCAAAATGGATTCTGCGAGGA TGACAATGTTTGCAGGTGCCATGTCGGCTGGCAGGGTCCCCTTTGTGACCAGTGCGTG ACCTCTCCCGGCTGCCTTCACGGACTCTGTGGAGAACCCGGGCAGTGCATTTGCACCG ACGGCTGGGACGGGGAGCTCTGTGATAGAGATGTTCGGGCCTGCTCCTCGGCCCCCTG TGCCAACAACGGGACCTGCGTGAGCCTGGACGGTGGCCTCTATGAATGCTCCTGTGCC CCCGGGTACTCGGGAAAGGACTGCCAGAAAAAGGACGGGCCCTGTGTGATCAACGGCT CCCCCTGCCAGCACGGAGGCACCTGCGTGGATGATGAGGGCCGGGCCTCCCATGCCTC CTGCCTGTGCCCCCCTGGCTTCTCAGGCAATTTCTGCGAGATCGTGGCCAACAGCTGC ACCCCCAACCCATGCGAGAACGACGGCGTCTGCACTGACATTGGGGGCGACTTCCGCT GCCGGTGCCCAGCCGGCTTCATCGACAAGACCTGCAGCCGCCCGGTGACCAACTGCGC CAGCAGCCCGTGCCAGAACGGGGGCACCTGCCTGCAGCACACCCAGGTGAGCTACGAG TGTCTGTGCAAGCCCGAGTTCACAGGTCTCACCTGTGTCAAGAAGCGCGCGCTGAGCC CCCAGCAGGTCACCCGTCTGCCCAGCGGCTATGGGCTGGCCTACCGCCTGACCCCTGG GGTGCACGAGCTGCCGGTGCAGCAGCCGGAGCACCGCATCCTGAAGGTGTCCATGAAA GAGCTCAACAAGAAAACCCCTCTCCTCACCGAGGGCCAGGCCATCTGCTTCACCATCC TGGGCGTGCTCACCAGCCTGGTGGTGCTGGGCACTGTGGGTATCGTCTTCCTCAACAA GTGCGAGACCTGGGTGTCCAACCTGCGCTACAACCACATGCTGCGGAAGAAGAAGAAC CTGCTGCTTCAGTACAACAGCGGGGAGGACCTGGCCGTCAACATCATCTTCCCCGAGA AGATCGACATGACCACCTTCAGCAAGGAGGCCGGCGACGAGGAGATC**TAA**		
	ORF Start: ATG at 1	ORF Stop: TAA at 1150	
	SEQ ID NO: 46	383 aa	MW at 41153.6kD
NOV11a, CG91667-01 Protein Sequence	MTATEALLRVLLLLLAFGHSTYGAECFPACNPQNGFCEDDNVCRCHVGWQGPLCDQCV TSPGCLHGLCGEPGQCICTDGWDGELCDRDVRACSSAPCANNGTCVSLDGGLYECSCA PGYSGKDCQKKDGPCVINGSPCQHGGTCVDDEGRASHASCLCPPGFSGNFCEIVANSC TPNPCENDGVCTDIGGDFRCRCPAGFIDKTCSRPVTNCASSPCQNGGTCLQHTQVSYE CLCKPEFTGLTCVKKRALSPQQVTRLPSGYGLAYRLTPGVHELPVQQPEHRILKVSMK ELNKKTPLLTEGQAICFTILGVLTSLVVLGTVGIVFLNKCETWVSNLRYNHMLRKKKN LLLQYNSGEDLAVNIIFPEKIDMTTFSKEAGDEEI		

	SEQ ID NO: 47	1299 bp
NOV11b, CG91667-02 DNA Sequence	TCCGCAACCAGAAGCCCAGTGCGGCGCCAGGAGCCGGACCCGCGCCCGCACCGCTCCC GGGACCGCGACCCCGGCCGCCCAGAG**ATG**ACCGCGACCGAAGCCCTCCTGCGCGTCCT CTTGCTCCTGCTGGCTTTCGGCCACAGCACCTATGGGGCTGAATGCTTCCCGGCCTGC AACCCCCAAAATGGATTCTGCGAGGATGACAATGTTTGCAGGTGCCAGCCTGGCTGGC AGGGTCCCCTTTGTGACCAGTGCGTGACCTCTCCCGGCTGCCTTCACGGACTCTGTGG AGAACCCGGGCAGTGCATTTGCACCGACGGCTGGGACGGGGAGCTCTGTGATAGAGAT GTTCGGGCCTGCTCCTCGGCCCCCTGTGCCAACAACGGGACCTGCGTGAGCCTGGACG ATGGCCTCTATGAATGCTCCTGTGCCCCCGGGTACTCGGGAAAGGACTGCCAGAAAAA GGACGGGCCCTGTGTGATCAACGGCTCCCCCTGCCAGCACGGAGGCACCTGCGTGGAT GATGAGGGCCGGGCCTCCCATGCCTCCTGCCTGTGCCCCCCTGGCTTCTCAGGCAATT TCTGCGAGATCGTGGCCAACAGCTGCACCCCCAACCCATGCGAGAACGACGGCGTCTG CACTGACATCGGGGGCGACTTCCGCTGCCGGTGCCCAGCCGGCTTCATCGACAAGACC TGCAGCCGCCCGGTGACCAACTGCGCCAGCAGCCCGTGCCAGAACGGGGGCACCTGCC TGCAGCACACCCAGGTGAGCTACGAGTGTCTGTGCAAGCCCGAGTTCACAGGTCTCAC CTGTGTCAAGAAGCGCGCGCTGAGCCCCCAGCAGGTCACCCGTCTGCCCAGCGGCTAT GGGCTGGCCTACCGCCTGACCCCTGGGGTGCACGAGCTGCCGGTGCAGCAGCCGGAGC ACCGCATCCTGAAGGTGTCCATGAAAGAGCTCAACAAGAAAACCCCTCTCCTCACCGA GGGCCAGGCCATCTGCTTCACCATCCTGGGCGTGCTCACCAGCCTGGTGGTGCTGGGC ACTGTGGGTATCGTCTTCCTCAACAAGTGCGAGACCTGGGTGTCCAACCTGCGCTACA ACCACATGCTGCGGAAGAAGAAGAACCTGCTGCTTCAGTACAACAGCGGGGAGGACCT GGCCGTCAACATCATCTTCCCCGAGAAGATCGACATGACCACCTTCAGCAAGGAGGCC GGCGACGAGGAGATC**TAA**GCAGCGTTCCCACAGCCCCCTCTAGATTCTTGGAGTTCCT CAGAGCTTACTATACGCGGTCTG	
	ORF Start: ATG at 85	ORF Stop: TAA at 1234
	SEQ ID NO: 48	383 aa	MW at 41200.6kD
NOV11b, CG91667-02 Protein Sequence	MTATEALLRVLLLLLAFGHSTYGAECFPACNPQNGFCEDDNVCRCQPGWQGPLCDQCV TSPGCLHGLCGEPGQCICTDGWDGELCDRDVRACSSAPCANNGTCVSLDDGLYECSCA PGYSGKDCQKKDGPCVINGSPCQHGGTCVDDEGRASHASCLCPPGFSGNFCEIVANSC TPNPCENDGVCTDIGGDFRCRCPAGFIDKTCSRPVTNCASSPCQNGGTCLQHTQVSYE CLCKPEFTGLTCVKKRALSPQQVTRLPSGYGLAYRLTPGVHELPVQQPEHRILKVSMK ELNKKTPLLTEGQAICFTILGVLTSLVVLGTVGIVFLNKCETWVSNLRYNHMLRKKKN LLLQYNSGEDLAVNIIFPEKIDMTTFSKEAGDEEI	

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

Table 11B. Comparison of NOV11a against NOV11b.

Protein Sequence	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV11b	1..383 1..383	353/383 (92%) 353/383 (92%)

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

Table 11C. Protein Sequence Properties NOV11a

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 24 and 25

A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 11D.

Table 11D. Geneseq Results for NOV11a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR56166	Neuroendocrine tumor dlk - Homo sapiens, 383 aa. [WO9413701-A, 23-JUN-1994]	1..383 1..383	380/383 (99%) 380/383 (99%)	0.0
AAR56167	Neuroendocrine tumor dlk - Mus sp, 385 aa. [WO9413701-A, 23-JUN-1994]	1..383 1..385	330/385 (85%) 348/385 (89%)	0.0
AAY77124	Human neurotransmission-associated protein (NTAP) 1296451 - Homo sapiens, 272 aa. [WO200001821-A2, 13-JAN-2000]	1..185 1..163	157/185 (84%) 159/185 (85%)	2e-95
AAE13632	Human preadipocyte factor-1-like protein - Homo sapiens, 383 aa. [WO200157233-A2, 09-AUG-2001]	7..322 10..325	120/319 (37%) 168/319 (52%)	1e-63
AAG67516	Amino acid sequence of a human secreted polypeptide - Homo sapiens, 383 aa. [WO200166690-A2, 13-SEP-2001]	7..322 10..325	120/319 (37%) 167/319 (51%)	2e-63

In a BLAST search of public sequence datbases, the NOV11a protein was found to have homology to the proteins shown in the BLASTP data in Table 11E.

Table 11E. Public BLASTP Results for NOV11a				
Protein Accession Number	**Protein/Organism/Length**	**NOV11a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
P80370	Delta-like protein precursor (DLK) (pG2) [Contains: Fetal antigen 1 (FA1)] - Homo sapiens (Human), 383 aa.	1..383 1..383	381/383 (99%) 381/383 (99%)	0.0
Q96DW5	UNKNOWN (PROTEIN FOR MGC:17291) - Homo sapiens (Human), 383 aa.	1..383 1..383	380/383 (99%) 380/383 (99%)	0.0
Q969Y6	HYPOTHETICAL 41.2 KDA PROTEIN (SIMILAR TO DELTA-LIKE HOMOLOG) (DROSOPHILA) - Homo sapiens (Human), 383 aa.	1..383 1..383	379/383 (98%) 380/383 (98%)	0.0
Q925U3	DLK (DELTA LIKE) - Mus musculus (Mouse), 385 aa.	1..383 1..385	332/385 (86%) 350/385 (90%)	0.0
A54785	preadipocyte factor 1 precursor, long form - mouse, 385 aa.	1..383 1..385	331/385 (85%) 349/385 (89%)	0.0

PFam analysis indicates that the NOV11a protein contains the domains shown in Table 11F.

Table 11F. Domain Analysis of NOV11a			
Pfam Domain	**NOV11a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
EGF: domain 1 of 6	26..54	10/47 (21%) 20/47 (43%)	2.6
Bowman-Birk_leg: domain 1 of 1	70..85	8/22 (36%) 13/22 (59%)	3.3
EGF: domain 2 of 6	57..85	9/47 (19%) 21/47 (45%)	0.1
metalthio: domain 1 of 1	61..117	14/67 (21%) 33/67 (49%)	5.9
EGF: domain 3 of 6	92..124	19/47 (40%) 28/47 (60%)	8.2e-09
EGF: domain 4 of 6	131..167	17/47 (36%) 28/47 (60%)	5.4e-08

EGF: domain 5 of 6	174..205	15/47 (32%) 28/47 (60%)	7.4e-09
EGF: domain 6 of 6	212..244	16/47 (34%) 25/47 (53%)	1.2e-07
Keratin_B2: domain 1 of 1	134..253	29/183 (16%) 58/183 (32%)	2.4

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: 49	3121 bp
NOV12a, CG92293-01 DNA Sequence	AATTCGC**ATG**GTCAACATGAAAAGTAAGGAACCTGCCGTGGGATCTAGATTCTTCTCT AGAATTAGTAGTTGGAGAAATTCAACAGTGACTGGACATCCATGGCAGGTCTCCCTAA AATCAGATGAGCACCACTTCTGTGGAGGAAGCTTGATTCAAGAAGATCGGGTTGTTAC AGCAGCACACTGCCTGGACAGCCTCAGTGAGAAGCAGCTGAAGAATATAACTGTGACT TCTGGGGAGTACAGCCTCTTTCAGAAGGATAAGCAAGAACAGAATATTCCTGTCTCAA AAATTATTACCCATCCTGAATACAACAGCCGTGAATATATGAGTCCTGATATTGCACT GCTGTATCTAAAACACAAAGTCAAGTTTGGAAATGCTGTTCAGCCAATCTGTCTTCCT GACAGCGATGATAAAGTTGAACCAGGAATTCTTTGCTTATCCAGTGGATGGGGCAAGA TTTCCAAAACATCAGAATATTCAAATGTCCTACAAGAAATGGAACTTCCCATCATGGA TGACAGAGCGTGTAATACTGTGCTCAAGAGCATGAACCTCCCTCCCCTGGGAAGGACC ATGCTGTGTGCTGGCTTCCCTGATGGGGGAATGGACGCCTGCCAGGGGGACTCTGGAG GACCACTGGTTTGTAGAAGAGGTGGTGGAATCTGGATTCTTGCTGGGATAACTTCCTG GGTAGCTGGTTGTGCTGGAGGTTCAGTTCCCGTAAGAAACAACCATGTGAAGGCATCA CTTGGCATTTTCTCCAAAGTGTCTGAGTTGATGGATTTTATCACTCAAAACCTGTTCA CAGGTTCCATTTATTACATTTTCTTCACCTTCCCCTACCCCAGCTTATATGTTTGGAA AATAATGGTACCAGAAGATAAAATAATCCTGATAAAATTTACAAGTTTAGACATGGAA AAGCAAGTTGGATGTGATCATGACTATGTATCTTTACGATCAAGCAGTGGAGTGCTTT TTAGTAAGGTCTGTGGAAAAATATTGCCTTCACCATTGCTGGCAGAGACCAGTGAGGC CATGGTTCCATTTGTTTCTGATACAGAAGACAGTGGCAGTGGCTTTGAGCTTACCGTT ACTGCTGTACAGAAGTCAGAAGCAGGGTCAGGTTGTGGGAGTCTGGCTATATTGGTAG AAGAAGGGACAAATCACTCTGCCAAGTATCCTGATTTGTATCCCAGTAACACAAGGTG TCATTGGTTCATTTGTGCTCCAGAGAAGCACATTATAAAGTTGACATTTGAGGACTTT GCTGTCAAATTTAGTCCAAACTGTATTTATGATGCTGTTGTGATTTACGGTGATTCTG AAGAAAAGCACAAGTTAGCTAAACTTTGTGGAATGCTGACCATCACTTCAATATTCAG TTCTAGTAACATGACGGTGATATACTTTAAAAGTGATGGTAAAAATCGTTTACAAGGC TTCAAGGCCAGATTTACCATTTTGCCCTCAGAGTCTTTAAACAAATTTGAACCAAAGT TACCTCCCCAAAACAATCCTGTATCTACCGTAAAAGCTATTCTGCATGATGTCTGTGG CATCCCTCCATTTAGTCCCCAGTGGCTTTCCAGAAGAATCGCAGGAGGGGAAGAAGCC TGCCCCCACTGTTGGCCATGGCAGGTGGGTCTGAGGTTTCTAGGCGATTACCAATGTG GAGGTGCCATCATCAACCCAGTGTGGATTCTGACCGCAGCCCACTGTGTGCAATTGAA GAATAATCCACTCTCCTGGACTATTATTGCTGGGGACCATGACAGAAACCTGAAGGAA TCAACAGAGCAGGTGAGAAGGGCCAAACACATAATAGTGCATGAAGACTTTAACACAC TAAGTTATGACTCTGACATTGCCCTAATACAACTAAGCTCTCCTCTGGAGTACAACTC GGTGGTGAGGCCAGTATGTCTCCCACACAGCGCAGAGCCTCTATTTTCCTCGGAGATC TGTGCTGTGACCGGATGGGGAAGCATCAGTGCAGATGGTGGCCTAGCAAGTCGCCTAC AGCAGATTCAAGTGCATGTGTTAGAAAGAGAGGTCTGTGAACACACTTACTATTCTGC CCATCCAGGAGGGATCACAGAGAAGATGATCTGTGCTGGCTTTGCAGCATCTGGAGAG AAAGATTTCTGCCAGGGAGACTCTGGTGGGCCACTAGTATGTAGACATGAAAATGGTC

	CCTTTGTCCTCTATGGCATTGTCAGCTGGGGAGCTGGCTGTGTCCAGCCATGGAAGCC GGGTGTATTTGCCAGAGTGATGATCTTCTTGGACTGGATCCAATCAAAAATCAATGGT CCTGCTTCACTTCAGACAAATAATAAATGCAAAACCTTAAAACAACAATTGCCACCAC CCACACCTTCACCAGACAGTGCATCTTGGCCAGGTTGTTGCTCTGAAGCAGAGCTAGA AAAGCCTAGAGGCTTTTTTCCCACACCACGGTATCTACTGGATTATAGAGGAAGACTG GAATGTTCTTGGGTGCTCAGAGTTTCACCAAGCAGTATGGCAAAATTTACCATTGAGT ATCTGTCACTCCTGGGGTCTCCTGTGTGTCAAGACTCAGTTCTAATTATTTATGAAGA AAGACACAGTAAGAGAAAGACGGCAGGTGGATTACATGGAAGAAGACTTTACTCAATG ACTTTCATGAGTCCTGGACCGCTGGTGAGGGTGACATTCCATGCCCTTGTACGAGGTG CATTTGGTATAAGCTATATTGTCTTGAAAGTCCTAGGTCCAAAGGACAGTAAAATAAC CAGACTTTCCCAAAGTTCAAACAGAGAGCACTTGGTCCCTTGTGAGGATGTTCTTCTG ACCAAGCCAGAAGGGATCATGCGGATCCCAAGAAATTCTCACAGAACTACTATGGGCT CATTTACATGGCTCCAAGAAAGAGTTTATCTTGATATCCAGTGCTGCTTACCTGACTG TGCATTT**TAA**GACTGATGAGTCTGAGAGAAAGAGGTTTTAAGCTTATTTTAGAAGAGA TGATTCAGGAGCAATCACAGAAGAGCAATATTGAGACCCAATTTCCTATCAGTGGAGA GTTTTCACTACTAATCTGGTGCCAGACTCCCACAACCTGACCCTGCT
	ORF Start: ATG at 8 — ORF Stop: TAA at 2966
	SEQ ID NO: 50 — 986 aa — MW at 109103.2kD
NOV12a, CG92293-01 Protein Sequence	MVNMKSKEPAVGSRFFSRISSWRNSTVTGHPWQVSLKSDEHHFCGGSLIQEDRVVTAA HCLDSLSEKQLKNITVTSGEYSLFQKDKQEQNIPVSKIITHPEYNSREYMSPDIALLY LKHKVKFGNAVQPICLPDSDDKVEPGILCLSSGWGKISKTSEYSNVLQEMELPIMDDR ACNTVLKSMNLPPLGRTMLCAGFPDGGMDACQGDSGGPLVCRRGGGIWILAGITSWVA GCAGGSVPVRNNHVKASLGIFSKVSELMDFITQNLFTGSIYYIFFTFPYPSLYVWKIM VPEDKIILIKFTSLDMEKQVGCDHDYVSLRSSSGVLFSKVCGKILPSPLLAETSEAMV PFVSDTEDSGSGFELTVTAVQKSEAGSGCGSLAILVEEGTNHSAKYPDLYPSNTRCHW FICAPEKHIIKLTFEDFAVKFSPNCIYDAVVIYGDSEEKHKLAKLCGMLTITSIFSSS NMTVIYFKSDGKNRLQGFKARFTILPSESLNKFEPKLPPQNNPVSTVKAILHDVCGIP PFSPQWLSRRIAGGEEACPHCWPWQVGLRFLGDYQCGGAIINPVWILTAAHCVQLKNN PLSWTIIAGDHDRNLKESTEQVRRAKHIIVHEDFNTLSYDSDIALIQLSSPLEYNSVV RPVCLPHSAEPLFSSEICAVTGWGSISADGGLASRLQQIQVHVLEREVCEHTYYSAHP GGITEKMICAGFAASGEKDFCQGDSGGPLVCRHENGPFVLYGIVSWGAGCVQPWKPGV FARVMIFLDWIQSKINGPASLQTNNKCKTLKQQLPPPTPSPDSASWPGCCSEAELEKP RGFFPTPRYLLDYRGRLECSWVLRVSPSSMAKFTIEYLSLLGSPVCQDSVLIIYEERH SKRKTAGGLHGRRLYSMTFMSPGPLVRVTFHALVRGAFGISYIVLKVLGPKDSKITRL SQSSNREHLVPCEDVLLTKPEGIMRIPRNSHRTTMGSFTWLQERVYLDIQCCLPDCAF
	SEQ ID NO: 51 — 2929 bp
NOV12b, CG92293-02 DNA Sequence	AATTCGC**ATG**GTCAACATGAAAAGTAAGGAACCTGCCGTGGGATCTAGATTCTTCTCT AGAATTAGTAGTTGGAGAAATTCAACAGTGACTGGACATCCATGGCAGGTCTCCCTAA AATCAGATGAGCACCACTTCTGTGGAGGAAGCTTGATTCAAGAAGATCGGGTTGTTAC AGCAGCACACTGCCTGGACAGCCTCAGTGAGAAGCAGCTGAAGAATATAACTGTGACT TCTGGGGAGTACAGCCTCTTTCAGAAGGATAAGCAAGAACAGAATATTCCTGTCTCAA AAATTATTACCCATCCTGAATACAACAGCCGTGAATATATGAGTCCTGATATTGCACT GCTGTATCTAAAACACAAAGTCAAGTTTGGAAATGCTGTTCAGCCAATCTGTCTTCCT GACAGCGATGATAAAGTTGAACCAGGAATTCTTTGCTTATCCAGTGGATGGGGCAAGA TTTCCAAAACATCAGAATATTCAAATGTCCTACAAGAAATGGAACTTCCCATCATGGA TGACAGAGCGTGTAATACTGTGCTCAAGAGCATGAACCTCCCTCCCCTGGGAAGGACC ATGCTGTGTGCTGGCTTCCCTGATGGGGGAATGGACGCCTGCCAGGGGGACTCTGGAG GACCACTGGTTTGTAGAAGAGGTGGTGGAATCTGGATTCTTGCTGGGATAACTTCCTG GGTAGCTGGTTGTGCTGGAGGTTCAGTTCCCGTAAGAAACAACCATGTGAAGGCATCA CTTGGCATTTTCTCCAAAGTGTCTGAGTTGATGGATTTTATCACTCAAAACCTGTTCA CAGGTTCCATTTATTACATTTTCTTCACCTTCCCCTACCCCAGCTTATATGTTTGGAA AATAATGGTACCAGAAGATAAAATAATCCTGATAAAATTTACAAGTTTAGACATGGAA AAGCAAGTTGGATGTGATCATGACTATGTATCTTTACGATCAAGCAGTGGAGTGCTTT TTAGTAAGGTCTGTGGAAAAATATTGCCTTCACCATTGCTGGCAGAGACCAGTGAGGC CATGGTTCCATTTGTTTCTGATACAGAAGACAGTGGCAGTGGCTTTGAGCTTACCGTT ACTGCTGTACAGAAGTCAGAAGCAGGGTCAGGTTGTGGGAGTCTGGCTATATTGGTAG AAGAAGGGACAAATCACTCTGCCAAGTATCCTGATTTGTATCCCAGTAACACAAGGTG TCATTGGTTCATTTGTGCTCCAGAGAAGCACATTATAAAGTTGACATTTGAGGACTTT

	GCTGTCAAATTTAGTCCAAACTGTATTTATGATGCTGTTGTGATTTACGGTGATTCTG AAGAAAAGCACAAGTTAGCTAAACTTTGTGGAATGCTGACCATCACTTCAATATTCAG TTCTAGTAACATGACGGTGATATACTTTAAAAGTGATGGTAAAAATCGTTTACAAGGC TTCAAGGCCAGATTTACCATTTTGCCCTCAGAGTCTTTAAACAAATTTGAACCAAAGT TACCTCCCCAAAACAATCCTGTATCTACCGTAAAAGCTATTCTGCATGATGTCTGTGG CATCCCTCCATTTAGTCCCCAGTGGCTTTCCAGAAGAATCGCAGGAGGGGAAGAAGCC TGCCCCCACTGTTGGCCATGGCAGGTGGGTCTGAGGTTTCTAGGCGATTACCAATGTG GAGGTGCCATCATCAACCCAGTGTGGATTCTGACCGCAGCCCACTGTGTGCAATTGAA GAATAATCCACTCTCCTGGACTATTATTGCTGGGGACCATGACAGAAACCTGAAGGAA TCAACAGAGCAGGCAGATGGTGGCCTAGCAAGTCGCCTACAGCAGATTCAAGTGCATG TGTTAGAAAGAGAGGTCTGTGAACACACTTACTATTCTGCCCATCCAGGAGGGATCAC AGAGAAGATGATCTGTGCTGGCTTTGCAGCATCTGGAGAGAAAGATTTCTGCCAGGGA GACTCTGGTGGGCCACTAGTATGTAGACATGAAAATGGTCCCTTTGTCCTCTATGGCA TTGTCAGCTGGGGAGCTGGCTGTGTCCAGCCATGGAAGCCGGGTGTATTTGCCAGAGT GATGATCTTCTTGGACTGGATCCAATCAAAAATCAATGGTCCTGCTTCACTTCAGACA AATAATAAATGCAAAACCTTAAAACAACAATTGCCACCACCCACACCTTCACCAGACA GTGCATCTTGGCCAGGTTGTTGCTCTGAAGCAGAGCTAGAAAAGCCTAGAGGCTTTTT TCCCACACCACGGTATCTACTGGATTATAGAGGAAGACTGGAATGTTCTTGGGTGCTC AGAGTTTCACCAAGCAGTATGGCAAAATTTACCATTGAGTATCTGTCACTCCTGGGGT CTCCTGTGTGTCAAGACTCAGTTCTAATTATTTATGAAGAAAGACACAGTAAGAGAAA GACGGCAGGTGGATTACATGGAAGAAGACTTTACTCAATGACTTTCATGAGTCCTGGA CCGCTGGTGAGGGTGACATTCCATGCCCTTGTACGAGGTGCATTTGGTATAAGCTATA TTGTCTTGAAAGTCCTAGGTCCAAAGGACAGTAAAATAACCAGACTTTCCCAAAGTTC AAACAGAGAGCACTTGGTCCCTTGTGAGGATGTTCTTCTGACCAAGCCAGAAGGGATC ATGCGGATCCCAAGAAATTCTCACAGAACTACTATGGGCTCATTTACATGGCTCCAAG AAAGAGTTTATCTTGATATCCAGTGCTGCTTACCTGACTGTGCATTT**TAA**GACTGATG AGTCTGAGAGAAAGAGGTTTTAAGCTTATTTTAGAAGAGATGATTCAGGAGCAATCAC AGAAGAGCAATATTGAGACCCAATTTCCTATCAGTGGAGAGTTTTCACTACTAATCTG GTGCCAGACTCCCACAACCTGACCCTGCT
	ORF Start: ATG at 8 — ORF Stop: TAA at 2774
	SEQ ID NO: 52 — 922 aa — MW at 102051.3kD
NOV12b, CG92293-02 Protein Sequence	MVNMKSKEPAVGSRFFSRISSWRNSTVTGHPWQVSLKSDEHHFCGGSLIQEDRVVTAA HCLDSLSEKQLKNITVTSGEYSLFQKDKQEQNIPVSKIITHPEYNSREYMSPDIALLY LKHKVKFGNAVQPICLPDSDDKVEPGILCLSSGWGKISKTSEYSNVLQEMELPIMDDR ACNTVLKSMNLPPLGRTMLCAGFPDGGMDACQGDSGGPLVCRRGGGIWILAGITSWVA GCAGGSVPVRNNHVKASLGIFSKVSELMDFITQNLFTGSIYYIFFTFPYPSLYVWKIM VPEDKIILIKFTSLDMEKQVGCDHDYVSLRSSSGVLFSKVCGKILPSPLLAETSEAMV PFVSDTEDSGSGFELTVTAVQKSEAGSGCGSLAILVEEGTNHSAKYPDLYPSNTRCHW FICAPEKHIIKLTFEDFAVKFSPNCIYDAVVIYGDSEEKHKLAKLCGMLTITSIFSSS NMTVIYFKSDGKNRLQGFKARFTILPSESLNKFEPKLPPQNNPVSTVKAILHDVCGIP PFSPQWLSRRIAGGEEACPHCWPWQVGLRFLGDYQCGGAIINPVWILTAAHCVQLKNN PLSWTIIAGDHDRNLKESTEQADGGLASRLQQIQVHVLEREVCEHTYYSAHPGGITEK MICAGFAASGEKDFCQGDSGGPLVCRHENGPFVLYGIVSWGAGCVQPWKPGVFARVMI FLDWIQSKINGPASLQTNNKCKTLKQQLPPPTPSPDSASWPGCCSEAELEKPRGFFPT PRYLLDYRGRLECSWVLRVSPSSMAKFTIEYLSLLGSPVCQDSVLIIYEERHSKRKTA GGLHGRRLYSMTFMSPGPLVRVTFHALVRGAFGISYIVLKVLGPKDSKITRLSQSSNR EHLVPCEDVLLTKPEGIMRIPRNSHRTTMGSFTWLQERVYLDIQCCLPDCAF

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..986 1..922	894/988 (90%) 903/988 (90%)

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

Table 12C. Protein Sequence Properties NOV12a	
PSort analysis:	0.4820 probability located in mitochondrial matrix space; 0.4298 probability located in microbody (peroxisome); 0.1907 probability located in mitochondrial inner membrane; 0.1907 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence

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**
ABG24246	Novel human diagnostic protein #24237 - Homo sapiens, 913 aa. [WO200175067-A2, 11-OCT-2001]	1..771 13..738	660/771 (85%) 670/771 (86%)	0.0
ABG24246	Novel human diagnostic protein #24237 - Homo sapiens, 913 aa. [WO200175067-A2, 11-OCT-2001]	1..771 13..738	660/771 (85%) 670/771 (86%)	0.0
ABG19887	Novel human diagnostic protein #19878 - Homo sapiens, 1576 aa. [WO200175067-A2, 11-OCT-2001]	1..770 852..1576	659/770 (85%) 669/770 (86%)	0.0

ABG14588	Novel human diagnostic protein #14579 - Homo sapiens, 1576 aa. [WO200175067-A2, 11-OCT-2001]	1..770 852..1576	659/770 (85%) 669/770 (86%)	0.0
ABG10218	Novel human diagnostic protein #10209 - Homo sapiens, 1576 aa. [WO200175067-A2, 11-OCT-2001]	1..770 852..1576	659/770 (85%) 669/770 (86%)	0.0

In a BLAST search of public sequence datbases, 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
Q91674	POLYPROTEIN - Xenopus laevis (African clawed frog), 1524 aa.	15..955 53..995	413/969 (42%) 567/969 (57%)	0.0
P79953	OVIDUCTIN - Xenopus laevis (African clawed frog), 1004 aa.	15..778 42..828	291/798 (36%) 430/798 (53%)	e-146
Q90WD8	OVIDUCTIN - Bufo japonicus (Japanese toad), 974 aa.	10..801 41..849	284/829 (34%) 436/829 (52%)	e-141
Q9BK47	SEA STAR REGENERATION-ASSOCIATED PROTEASE SRAP - Luidia foliolata, 267 aa.	513..769 12..264	111/264 (42%) 156/264 (59%)	2e-51
O96899	PLASMINOGEN ACTIVATOR SPA - Scolopendra subspinipes, 277 aa.	532..767 33..264	104/241 (43%) 148/241 (61%)	8e-49

PFam analysis indicates that the NOV12a protein contains the domains shown in Table 12F.

Table 12F. Domain Analysis of NOV12a

Pfam Domain	NOV12a Match Region	Identities/ Similarities for the Matched Region	Expect Value
trypsin: domain 1 of 2	19..263	100/275 (36%) 186/275 (68%)	2.5e-76
CUB: domain 1 of 3	266..365	31/116 (27%) 64/116 (55%)	7.5e-06

CUB: domain 2 of 3	377..486	40/116 (34%) 73/116 (63%)	1.8e-22
trypsin: domain 2 of 2	533..765	109/264 (41%) 182/264 (69%)	2.5e-82
CUB: domain 3 of 3	804..912	23/118 (19%) 70/118 (59%)	7.3e-05

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: 53	2544 bp
NOV13a, CG92384-01 DNA Sequence	TCCTGGGCCCCAGCCCCGCGCAGGCCAAGG**ATG**AGGCCGAGGCCCGAAGGTAGGGGGC TCCGGGCGGGAGTCGCGCTGTCCCCCGCGCTACTGCTGCTGCTGCTGCTGCCGCCGCC GCCGACGCTGCTGGGGCGCCTGTGGGCAGCGGGCACACCCTCGCCGTCGGCGCCCGGA GCTCGGCAGGACGGCGCGCTGGGAGCCGGCCGCGTCAAACGCGGCTGGGTGTGGAACC AGTTCTTCGTGGTAGAGGAGTACACGGGCACGGAGCCCCTGTATGTGGGCAAGATCCA CTCCGACTCAGACGAGGGTGACGGGGCCATCAAGTACACCATCTCAGGCGAGGGTGCT GGGACCATCTTCCTGATCGACGAGCTGACAGGCGACATTCATGCCATGGAGCGCCTGG ACCGCGAGCAGAAAACCTTCTACACGCTGCGGGCCCAGGCTCGGGATCGCGCCACCAA CCGCCTACTGGAGCCCGAGTCGGAGTTCATCATCAAGGTGCAGGACATCAATGACAGT GAGCCCCGCTTCCTGCACGGCCCCTATATTGGCAGCGTGGCCGAGCTCTCACCTACAG GTACGTCGGTGATGCAGGTGATGGCCTCGGATGCGGATGACCCCACGTACGGCAGCAG CGCTCGGCTGGTGTACAGCGTGCTGGACGGCGAGCACCACTTCACCGTGGACCCCAAG ACCGGTGTAATCCGGACGGCTGTGCCTGACCTTGACCGCGAGAGCCAGGAGCGCTACG AGGTGGTGATCCAGGCCACAGACATGGCGGGTCAGCTGGGTGGCCTCTCGGGCTCCAC TACCGTCACCATCGTAGTCACCGACGTCAATGACAACCCGCCCCGTTTCCCGCAGGAG ATGTACCAGTTCAGCATCCAGGAGTCAGCCCCCATTGGAACGGCTGTGGGACGTGTGA AGGCTGAGGACTCAGACGTGGGAGAGAACACAGACATGACTTACCACCTTAAGGACGA GAGCAGCAGCGGCGGCGATGTGTTCAAGGTCACCACAGACAGCGACACTCAGGAGGCC ATCATCGTAGTGCAGAAGCGCCTGGACTTCGAATCCCAGCCCGTGCACACCGTGATCC TGGAGGCCCTCAACAAGTTCGTGGACCCCCGCTTCGCCGACCTGGGCACGTTCCGCGA CCAGGCGATCGTGCGCGTGGCCGTGACCGACGTGGACGAGCCCCCCGAGTTCCGGCCG CCCTCCGGCCTCCTGGAGGTGCAGGAGGACGCGCAGGTGGGCTCCCTGGTCGGCGTGG TGACGGCGCGGGACCCCGACGCCGCCAACCGGCCCGTCCGGTACGCCATTGACCGCGA ATCAGATTTGGACCAGATCTTCGATATCGATGCGGACACAGGCGCCATCGTGACTGGC AAGGGGCTGGACCGCGAGACGGCCGGCTGGCACAACATCACAGTGCTGGCCATGGAGG CGGACAATCATGCACAGCTATCCCGGGCATCCCTAAGGATCCGAATCCTGGATGTGAA CGACAATCCCCCAGAACTGGCCACACCCTACGAGGCAGCTGTATGCGAGGATGCCAAG CCAGGCCAGCTCATCCAGACCATCAGCGTGGTGGACAGAGACGAGCCCCAAGGCGGGC ACCGCTTCTATTTCCGCCTGGTGCCTGAAGCTCCCAGCAACCCTCATTTCTCTCTGCT TGACATCCAAGACAACACCGCTGCAGTGCACACGCAGCACGTGGGCTTCAACCGGCAG GAGCAGGACGTGTTCTTCCTGCCCATCCTGGTGGTAGACAGTGGGCCGCCCACACTGA GCAGCACAGGCACGCTCACCATCCGCATCTGTGGCTGCGACAGCTCCGGCACCATCCA GTCCTGCAACACCACGGCCTTTGTCATGGCCGCCTCCCTCAGCCCCGGCGCCCTCTTG GTCTGCGTTCTCATCCTGGTTGTGCTGGTGCTGCTGATCCTCACCCTCAGGCGCCACC ACAAGAGCCACCTGAGCTCGGACGAGGATGAAGACATGCGGGACAACGTCATCATATA CAACGACGAAGGCGGCGGCGAGCAGGACACCGAAGCCTACGACATGTCGGCGCTGCGG AGCCTCTACGACTTCGGCGAGCTCAAGGGCGGCGACGGGGGCGGCAGCGCGGGCAGCC CCCCGCAGGCCCACCTGCCCTCCGAGCGCCACTCGCTGCCGCAGGGGCCGCCGAGCCC

	CGAGCCAGACTTCTCAGTGTTCAGGGACTTCATCAGCCGCAAGGTGGCACTGGCGGAC GGGGACCTGTCGGTGCCGCCCTACGACGCCTTCCAGACCTACGCCCTCGAGGGCGCGG ACTCGCCGGCCGCCTCGCTCAGCTCCCTGCACAGCGGCTCGTCGGGCTCCGAGCAGGA CTTCGCCTATTTCAGCATGTGGGGTCATTCGGGCTCCGAGCAGGACTTCGCCTATCTC AGCAGCTGGGGTCCGCGCTTCCGGCCCCTGGCCGCGCTCTACGCCGGCCACCGCGGGG ACGACGAGGCCCAGGCCTCC**TAG**CCCCTCGCCCTGCCGTCGGGGCGCGGC		
	ORF Start: ATG at 31	ORF Stop: TAG at 2515	
	SEQ ID NO: 54	828 aa	MW at 89732.6kD
NOV13a, CG92384-01 Protein Sequence	MRPRPEGRGLRAGVALSPALLLLLLLPPPPTLLGRLWAAGTPSPSAPGARQDGALGAG RVKRGWVWNQFFVVEEYTGTEPLYVGKIHSDSDEGDGAIKYTISGEGAGTIFLIDELT GDIHAMERLDREQKTFYTLRAQARDRATNRLLEPESEFIIKVQDINDSEPRFLHGPYI GSVAELSPTGTSVMQVMASDADDPTYGSSARLVYSVLDGEHHFTVDPKTGVIRTAVPD LDRESQERYEVVIQATDMAGQLGGLSGSTTVTIVVTDVNDNPPRFPQEMYQFSIQESA PIGTAVGRVKAEDSDVGENTDMTYHLKDESSSGGDVFKVTTDSDTQEAIIVVQKRLDF ESQPVHTVILEALNKFVDPRFADLGTFRDQAIVRVAVTDVDEPPEFRPPSGLLEVQED AQVGSLVGVVTARDPDAANRPVRYAIDRESDLDQIFDIDADTGAIVTGKGLDRETAGW HNITVLAMEADNHAQLSRASLRIRILDVNDNPPELATPYEAAVCEDAKPGQLIQTISV VDRDEPQGGHRFYFRLVPEAPSNPHFSLLDIQDNTAAVHTQHVGFNRQEQDVFFLPIL VVDSGPPTLSSTGTLTIRICGCDSSGTIQSCNTTAFVMAASLSPGALLVCVLILVVLV LLILTLRRHHKSHLSSDEDEDMRDNVIIYNDEGGGEQDTEAYDMSALRSLYDFGELKG GDGGGSAGSPPQAHLPSERHSLPQGPPSPEPDFSVFRDFISRKVALADGDLSVPPYDA FQTYALEGADSPAASLSSLHSGSSGSEQDFAYFSMWGHSGSEQDFAYLSSWGPRFRPL AALYAGHRGDDEAQAS		

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

Table 13B. Protein Sequence Properties NOV13a

PSort analysis:	0.4600 probability located in plasma membrane; 0.1561 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 34 and 35

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 13C.

Table 13C. Geneseq Results for NOV13a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU09864	Novel human secreted protein #5 - Homo sapiens, 801 aa. [WO200179454-A1, 25-OCT-2001]	41..821 38..796	412/784 (52%) 540/784 (68%)	0.0

AAM78375	Human protein SEQ ID NO 1037 - Homo sapiens, 788 aa. [WO200157190-A2, 09-AUG-2001]	50..825 43..788	393/779 (50%) 546/779 (69%)	0.0
AAW13132	Full length human cadherin-8 - Homo sapiens, 793 aa. [US5597725-A, 28-JAN-1997]	17..824 7..792	394/816 (48%) 560/816 (68%)	0.0
AAW25635	Human cadherin-8 - Homo sapiens, 793 aa. [US5646250-A, 08-JUL-1997]	17..824 7..792	394/816 (48%) 560/816 (68%)	0.0
AAW13126	Full length rat cadherin-8 - Rattus rattus, 799 aa. [US5597725-A, 28-JAN-1997]	17..824 14..798	390/813 (47%) 557/813 (67%)	0.0

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

5

Table 13D. Public BLASTP Results for NOV13a

Protein Accession Number	Protein/Organism/Length	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UJ99	DJ998H6.1 (ORTHOLOG OF RAT PB-CADHERIN) - Homo sapiens (Human), 828 aa.	1..828 1..828	810/843 (96%) 811/843 (96%)	0.0
Q9WTP5	PB-CADHERIN - Mus musculus (Mouse), 813 aa.	1..828 1..813	762/833 (91%) 777/833 (92%)	0.0
Q63315	LONG TYPE PB-CADHERIN - Rattus norvegicus (Rat), 813 aa.	1..828 1..813	761/833 (91%) 775/833 (92%)	0.0
Q63561	SHORT TYPE PB-CADHERIN - Rattus norvegicus (Rat), 694 aa.	1..688 1..690	637/695 (91%) 649/695 (92%)	0.0
Q9ULB5	Cadherin-7 precursor - Homo sapiens (Human), 785 aa.	56..816 41..776	420/764 (54%) 552/764 (71%)	0.0

PFam analysis indicates that the NOV13a protein contains the domains shown in Table 13E.

Table 13E. Domain Analysis of NOV13a			
Pfam Domain	**NOV13a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
cadherin: domain 1 of 5	68..159	32/107 (30%) 60/107 (56%)	2.4e-12
FBPase: domain 1 of 1	200..225	12/28 (43%) 20/28 (71%)	2.9
cadherin: domain 2 of 5	173..268	46/109 (42%) 80/109 (73%)	1.3e-29
cadherin: domain 3 of 5	282..386	30/111 (27%) 75/111 (68%)	6.5e-14
cadherin: domain 4 of 5	399..490	37/108 (34%) 69/108 (64%)	1.2e-17
cadherin: domain 5 of 5	503..600	27/113 (24%) 71/113 (63%)	1.3e-10
Cadherin_C_term: domain 1 of 1	646..819	75/179 (42%) 147/179 (82%)	1.3e-65

Example 14.

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: 55	1170 bp
NOV14a, CG92455-01 DNA Sequence	TAAGCTTCTCTGAAC**ATG**CAGAGCAGTGGTAACTATGAGGAGGCACAGTCTCTCCTGA GCATGACTCTGGCTGTTGGTGTGGCTTTGCTGCAGCTGCCGTGTGCCATTGATGATCC TGCTCCTCTTCCTTCGGGAAAGCTGCCTGATGGAGTCTGGCTCCTGGAGCTGAGCCAC AATAATCTCAGCCATCTGCCGGCTGGCGCCTTCCAGGGCTTTTGGGGACTGCGGGTGT TGCTGCTTTCTCTCAATATCCTGCGGGATCTGTCTGATGGGGCCCTAGGGGGCCTCAG TTTCCTGGAGCAGCTGAACCTCAGCCATAACCAGCTGGCCCATCTGCCCACAGACTTC TCGGCTACCCTGGGCTCTCTGCTCTGCCTGGACCTCTCTCACAACCTACTCACTTCCC TGGACCCCACCAGCCTGTGGCGCCTGGGGGGCCTGGAGCAGCTCAACCTGAGCCACAA CCAGCTGGCTGAACTGGCCGCAGGGGTCTTTGGGGGCCTCTTCCACCTACACTGGCTC TCGCTGGCTGGGAACCAGCTGCAGCGGGTGAAGGGTGCTGCCCTGACCACTGTGCCGG GCTTGGAAGTCCTCTCTGTAGCTGGGAATGACATCAGTGCCTTCGGAAAGTTGGGTCA CCTGCGGCACTTGAGTGTCGTAGACCTGGGCATCCTGACTTGCGCTGGGCCCGAAAGG CTGTCAGGGGCAGTGCTGAGTGGTGTGGAGGCCCAGCTTTGCCTGGCTGAGACTGCCA CTGTGCTGGGCATCACAGGCACCGTGCTGCTCACAGTGGCTGTGGCTGTGCTGATGGC TGAGCGCAAGCGAAGACAGGGCCCGCAAGAAGCCGGGGAGCTGGGGAGCTTTCTGGAG AGGCTATTTAATCAGCAGGCAGATCAACAGGCCAAAATTTCCACCATAACTGCAAACT TGAATGCGGATGAATGGATTACAACATGGTCAAACCTTGCAATGAGAGGCATTATAAT GTATGGTGCTATAATTGATAGTGATTACCGGGGAGAGTTAAAGGTCATTTTATACAAT	

	ACCACTCCAGATTCTTTTGCTATAAAACCGCAGATGCAGGTTGCTCAATTGTTAGTGG TATCTTGTCAACAACCCCCGAGGAAATTTCCACCCCAA**TAG**AAACAACATATAGAACG GAACATTCAG		
	ORF Start: ATG at 16	ORF Stop: TAG at 1141	
	SEQ ID NO: 56	375 aa	MW at 40138.8kD
NOV14a, CG92455-01 Protein Sequence	MQSSGNYEEAQSLLSMTLAVGVALLQLPCAIDDPAPLPSGKLPDGVWLLELSHNNLSH LPAGAFQGFWGLRVLLLSLNILRDLSDGALGGLSFLEQLNLSHNQLAHLPTDFSATLG SLLCLDLSHNLLTSLDPTSLWRLGGLEQLNLSHNQLAELAAGVFGGLFHLHWLSLAGN QLQRVKGAALTTVPGLEVLSVAGNDISAFGKLGHLRHLSVVDLGILTCAGPERLSGAV LSGVEAQLCLAETATVLGITGTVLLTVAVAVLMAERKRRQGPQEAGELGSFLERLFNQ QADQQAKISTITANLNADEWITTWSNLAMRGIIMYGAIIDSDYRGELKVILYNTTPDS FAIKPQMQVAQLLVVSCQQPPRKFPPQ		

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

Table 14B. Protein Sequence Properties NOV14a

PSort analysis:	0.4600 probability located in plasma membrane; 0.1285 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 31 and 32

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
ABG16717	Novel human diagnostic protein #16708 - Homo sapiens, 550 aa. [WO200175067-A2, 11-OCT-2001]	287..375 182..268	73/89 (82%) 78/89 (87%)	5e-33
ABG16717	Novel human diagnostic protein #16708 - Homo sapiens, 550 aa. [WO200175067-A2, 11-OCT-2001]	287..375 182..268	73/89 (82%) 78/89 (87%)	5e-33
ABG05979	Novel human diagnostic protein #5970 - Homo sapiens, 258 aa. [WO200175067-A2, 11-OCT-2001]	272..367 55..152	69/98 (70%) 78/98 (79%)	2e-30

ABG05979	Novel human diagnostic protein #5970 - Homo sapiens, 258 aa. [WO200175067-A2, 11-OCT-2001]	272..367 55..152	69/98 (70%) 78/98 (79%)	2e-30
AAB82352	Protein sequence SEQ ID NO.2 - Homo sapiens, 794 aa. [WO200138357-A2, 31-MAY-2001]	42..250 93..303	79/213 (37%) 103/213 (48%)	8e-20

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

5

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
Q9N4G6	Y71F9B.8 PROTEIN (1D304) - Caenorhabditis elegans, 542 aa.	45..271 190..436	80/248 (32%) 127/248 (50%)	6e-20
CAC42683	SEQUENCE 1 FROM PATENT WO0142286 - Homo sapiens (Human), 794 aa.	42..250 93..303	79/213 (37%) 103/213 (48%)	2e-19
Q9UGS3	DJ756G23.1 (NOVEL LEUCINE RICH PROTEIN) - Homo sapiens (Human), 797 aa (fragment).	42..250 101..311	79/213 (37%) 103/213 (48%)	2e-19
O70211	INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID-LABILE SUBUNIT - Rattus norvegicus (Rat), 603 aa.	44..236 386..595	74/215 (34%) 99/215 (45%)	1e-18
P70193	MEMBRANE GLYCOPROTEIN - Mus musculus (Mouse), 1091 aa.	44..214 237..415	67/179 (37%) 92/179 (50%)	3e-18

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

Table 14E. Domain Analysis of NOV14a

Pfam Domain	NOV14a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LRR: domain 1 of 7	45..68	10/25 (40%) 17/25 (68%)	1
LRR: domain 2 of 7	69..92	9/25 (36%) 20/25 (80%)	4.2

LRR: domain 3 of 7	93..115	10/25 (40%) 16/25 (64%)	0.39
LRR: domain 4 of 7	117..140	12/25 (48%) 16/25 (64%)	0.057
LRR: domain 5 of 7	141..164	11/25 (44%) 19/25 (76%)	0.0059
LRR: domain 6 of 7	165..188	6/25 (24%) 15/25 (60%)	39
LRR: domain 7 of 7	189..210	7/25 (28%) 16/25 (64%)	44
dUTPase: domain 1 of 1	269..375	28/139 (20%) 75/139 (54%)	0.00014

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: 57	2328 bp
NOV15a, CG92531-01 DNA Sequence	**ATG**CCCGGCTACAGTGCTCTTTCTAATAAACCCATGCTGGAAACAACCCAAATGTCTA TCACTAGAGGAATGGGTAAGCTACTTGTGGTACGGTGTGGTACCGAGAAGGCTGGACC AGCAGTTCCAGGCGGCATGGAGGGGCCCCGGAGCTCCACCCATGTCCCCTTGGTGCTG CCGCTTCTTGTACTTCTGCTGCTGGCCCCGGCTAGGCAGGCCGCCGCCCAGCGCTGCC CACAGGCCTGCATCTGTGACAACTCCAGGCGACACGTTGCCTGCCGGTACCAGAACCT CACTGAGGTGCCAGACGCCATCCCTGAGCTGACCCAGCGGCTGGACCTGCAGGGCAAT TTGCTGAAGGTGATCCCCGCAGCCGCCTTCCAGGGCGTGCCTCACCTCACACACCTGG ACCTGCGCCACTGCGAGGTGGAGCTGGTGGCCGAGGGCGCCTTCCGTGGCCTGGGCCG CCTGCTCCTGCTCAACCTGGCCTCCAACCACCTGCGTGAGCTGCCCCAGGAGGCGCTG GACGGGCTGGGCTCGTTGCGGCGGCTGGAGCTGGAGGGGAACGCACTGGAGGAGCTGC GGCCGGGGACGTTCGGGGCACTGGGTGCGCTGGCCACGCTAAACCTGGCCCACAACGC CCTGGTTTACCTGCCCGCCATGGCCTTCCAGGGGCTACTGCGCGTCCGCTGGCTGCGG CTGTCGCACAACGCGCTCAGCGTGCTGGCCCCCGAGGCCCTGGCTGGCCTGCCCGCCC TGAGACGGCTCAGCCTACACCACAACGAGCTCCAGGCTCTGCCCGGGCCTGTCTTGTC CCAGGCCCGCGGCCTGGCCCGTCTGGAGCTGGGCCACAACCCGCTCACCTACGCGGGC GAGGAGGACGGGCTGGCGCTGCCCGGCCTGCGGGAGCTGCTGCTGGACGGCGGGGCCC TGCAGGCCCTGGGTCCCAGGGCCTTCGCACACTGTCCGCGCCTGCACACCCTCGACCT CCGCGGGAACCAGCTAGACACCCTGCCCCCGCTGCAGGGCCCGGGCCAGCTGCGCCGG CTGCGGCTGCAGGGGAATCCGCTGTGGTGCGGCTGCCAGGCGCGGCCCCTACTCGAGT GGCTGGCGCGGGCGCGCGTGCGCTCGGACGGCGCGTGCCAGGGGCCGCGGCGCCTGCG GGGCGAGGCTCTGGACGCCCTGCGGCCCTGGGACCTGCGCTGCCCTGGGGACGCGGCG CAGGAAGAGGAAGAGCTGGAAGAGCGGGCTGTGGCCGGGCCCCGCGCCCCTCCGCGCG GCCCTCCGCGCGGCCCCGGGGAGGAGCGGGCAGTCGCGCCTTGCCCTCGCGCCTGCGT GTGCGTCCCCGAGTCCCGGCACAGCAGCTGCGAGGGCTGCGGCCTGCAGGCGGTGCCC CGCGGCTTCCCCAGCGACACCCAGCTCCTGGACCTGAGGCGGAACCACTTCCCCTCGG TGCCCCGAGCGGCCTTCCCCGGCCTGGGCCACCTGGTGTCGCTGCACCTGCAGCACTG CGGCATCGCGGAGCTGGAAGCGGGCGCCCTGGCCGGGCTGGGCCGCCTGATCTACCTG TACCTCTCCGACAACCAGCTCGCAGGCCTCAGCGCTGCTGCCCTTGAAGGGGCTCCCC GCCTCGGCTACCTGTACCTAGAACGCAACCGTTTCCTGCAGGTGCCAGGGGCTGCCCT GCGCGCCCTGCCCAGCCTCTTCTCCCTGCACCTGCAGGACAACGCTGTGGACCGCCTG

	GCACCTGGGGACCTGGGGAGAACACGGGCCTTGCGCTGGGTCTACCTGAGTGGAAACC GCATCACCGAAGTGTCCCTTGGGGCGCTGGGCCCAGCTCGGGAGCTGGAGAAGCTGCA CCTGGACAGGAATCAGCTGCGAGAGGTGCCCACTGGGGCCTTGGAGGGGCTGCCTGCC CTCCTGGAGCTGCAGCTCTCGGGCAACCCACTCAGGGCCTTGCGTGACGGAGCCTTCC AGCCTGTGGGCAGGTCGCTGCAGCACCTCTTCCTGAACAGCAGTGGCCTGGAGCAGGT GGGCACTGGGCATCTGGCGGGGTTGGTGCAGGAGGCGGCACAAGGCCACAGGCAGCGT GCATTCACTCAACAAGCATTTGCCAGCCCCTTGGTGCCAGGCCTGGGGCCCGGGCTCC AGAGCCTGCACCTGCAGAAGAACCAGCTTCGGGCCCTGCCTGCCCTGCCCAGTCTCAG CCAGCTGGAGCTCATCGACCTCAGCAGCAATCCCTTCCACTGTGACTGCCAGCTGCTT CCGCTGCACAGGCACACCATGTCCATGCCCATCCGAGCAGCTTGTGGGGAGGGGCGGG TCCTG**TGA**		
	ORF Start: ATG at 1	ORF Stop: TGA at 2326	
	SEQ ID NO: 58	775 aa	MW at 83600.4kD
NOV15a, CG92531-01 Protein Sequence	MPGYSALSNKPMLETTQMSITRGMGKLLVVRCGTEKAGPAVPGGMEGPRSSTHVPLVL PLLVLLLLAPARQAAAQRCPQACICDNSRRHVACRYQNLTEVPDAIPELTQRLDLQGN LLKVIPAAAFQGVPHLTHLDLRHCEVELVAEGAFRGLGRLLLLNLASNHLRELPQEAL DGLGSLRRLELEGNALEELRPGTFGALGALATLNLAHNALVYLPAMAFQGLLRVRWLR LSHNALSVLAPEALAGLPALRRLSLHHNELQALPGPVLSQARGLARLELGHNPLTYAG EEDGLALPGLRELLLDGGALQALGPRAFAHCPRLHTLDLRGNQLDTLPPLQGPGQLRR LRLQGNPLWCGCQARPLLEWLARARVRSDGACQGPRRLRGEALDALRPWDLRCPGDAA QEEEELEERAVAGPRAPPRGPPRGPGEERAVAPCPRACVCVPESRHSSCEGCGLQAVP RGFPSDTQLLDLRRNHFPSVPRAAFPGLGHLVSLHLQHCGIAELEAGALAGLGRLIYL YLSDNQLAGLSAAALEGAPRLGYLYLERNRFLQVPGAALRALPSLFSLHLQDNAVDRL APGDLGRTRALRWVYLSGNRITEVSLGALGPARELEKLHLDRNQLREVPTGALEGLPA LLELQLSGNPLRALRDGAFQPVGRSLQHLFLNSSGLEQVGTGHLAGLVQEAAQGHRQR AFTQQAFASPLVPGLGPGLQSLHLQKNQLRALPALPSLSQLELIDLSSNPFHCDCQLL PLHRHTMSMPIRAACGEGRVL		

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

Table 15B. Protein Sequence Properties NOV15a

PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 70 and 71

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 15C.

Table 15C. Geneseq Results for NOV15a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV15a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB82352	Protein sequence SEQ ID NO.2 - Homo sapiens, 794 aa. [WO200138357-A2, 31-MAY-2001]	12..766 1..728	696/756 (92%) 703/756 (92%)	0.0
AAE03600	Human leucine-rich repeat-containing protein, AZAD - Homo sapiens, 794 aa. [WO200142286-A2, 14-JUN-2001]	12..766 1..728	696/756 (92%) 703/756 (92%)	0.0
AAB99488	Human chondroadherin protein sequence - Homo sapiens, 381 aa. [WO200137861-A1, 31-MAY-2001]	436..758 41..335	124/323 (38%) 178/323 (54%)	2e-56
AAR85888	WD-40 domain-contg. insulin-like growth factor binding protein - Synthetic, 605 aa. [WO9521252-A2, 10-AUG-1995]	71..541 35..537	166/556 (29%) 209/556 (36%)	8e-37
AAB38400	Fragment of human secreted protein encoded by gene 3 clone HSYAV50 - Homo sapiens, 723 aa. [WO200061623-A1, 19-OCT-2000]	39..450 36..464	137/441 (31%) 180/441 (40%)	3e-31

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

5

Table 15D. Public BLASTP Results for NOV15a				
Protein Accession Number	**Protein/Organism/Length**	**NOV15a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9UGS3	DJ756G23.1 (NOVEL LEUCINE RICH PROTEIN) - Homo sapiens (Human), 797 aa (fragment).	43..766 40..737	684/724 (94%) 689/724 (94%)	0.0
CAC42683	SEQUENCE 1 FROM PATENT WO0142286 - Homo sapiens (Human), 794 aa.	12..766 1..728	696/756 (92%) 703/756 (92%)	0.0

O70210	CHONDROADHERIN PRECURSOR - Rattus norvegicus (Rat), 358 aa.	436..758 18..312	125/323 (38%) 178/323 (54%)	3e-56
Q96RJ5	CHONDROADHERIN - Homo sapiens (Human), 359 aa.	436..758 19..313	124/323 (38%) 178/323 (54%)	6e-56
A53860	chondroadherin precursor - bovine, 361 aa.	436..758 21..315	124/323 (38%) 178/323 (54%)	1e-55

PFam analysis indicates that the NOV15a protein contains the domains shown in Table 15E.

Table 15E. Domain Analysis of NOV15a

Pfam Domain	NOV15a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Trypan_glycop: domain 1 of 1	56..76	10/21 (48%) 20/21 (95%)	2.7
GASA: domain 1 of 1	51..100	16/109 (15%) 33/109 (30%)	9.8
LRRNT: domain 1 of 2	76..105	12/31 (39%) 21/31 (68%)	1.9e-05
LRR: domain 1 of 19	107..130	8/25 (32%) 18/25 (72%)	1.8
LRR: domain 2 of 19	131..154	6/25 (24%) 19/25 (76%)	2.8
LRR: domain 3 of 19	155..178	9/25 (36%) 19/25 (76%)	0.087
LRR: domain 4 of 19	179..202	10/25 (40%) 18/25 (72%)	0.082
LRR: domain 5 of 19	203..226	9/25 (36%) 16/25 (64%)	0.71
LRR: domain 6 of 19	227..250	11/25 (44%) 19/25 (76%)	0.017
LRR: domain 7 of 19	251..274	7/25 (28%) 18/25 (72%)	4.5
LRR: domain 8 of 19	299..322	9/25 (36%) 17/25 (68%)	68
LRR: domain 9 of 19	323..344	9/25 (36%) 18/25 (72%)	0.25

LRRCT: domain 1 of 1	354..402	20/55 (36%) 34/55 (62%)	0.0078
LRRNT: domain 2 of 2	439..468	14/31 (45%) 20/31 (65%)	0.047
LRR: domain 10 of 19	470..493	8/25 (32%) 16/25 (64%)	41
LRR: domain 11 of 19	494..517	5/25 (20%) 20/25 (80%)	0.35
LRR: domain 12 of 19	518..541	8/25 (32%) 20/25 (80%)	0.22
LRR: domain 13 of 19	542..565	7/25 (28%) 18/25 (72%)	11
LRR: domain 14 of 19	566..589	7/25 (28%) 15/25 (60%)	1.4e+02
LRR: domain 15 of 19	590..613	5/25 (20%) 19/25 (76%)	1.8
LRR: domain 16 of 19	614..637	9/25 (36%) 21/25 (84%)	0.0028
LRR: domain 17 of 19	638..661	9/25 (36%) 15/25 (60%)	38
LRR: domain 18 of 19	663..686	5/25 (20%) 17/25 (68%)	84
LRR: domain 19 of 19	714..735	8/25 (32%) 18/25 (72%)	0.91

Example 16.

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

Table 16A. NOV16 Sequence Analysis

	SEQ ID NO: 59	2800 bp
NOV16a, CG92715-01 DNA Sequence	TAGACGCGGAGCCCAAGGAGGTAAA**ATG**CACACTTGCTGCCCCCCAGTAACTTTGGAA CAGGACCTTCACAGAAAAATGCATAGCTGGATGCTGCAGACTCTAGCGTTTGCTGTAA CATCTCTCGTCCTTTCGTGTGCAGAAACCATCGATTATTACGGGGAAATCTGTGACAA TGCATGTCCTTGTGAGGAAAAGGACGGCATTTTAACTGTGAGCTGTGAAAACCGGGGG ATCATCAGTCTCTCTGAAATTAGCCCTCCCCGTTTCCCAATCTACCACCTCTTGTTGT CCGGAAACCTTTTGAACCGTCTCTATCCCAATGAGTTTGTCAATTACACTGGGGCTTC AATTTTGCATCTAGGTAGCAATGTTATCCAGGACATTGAGACCGGGGCTTTCCATGGG CTACGGGGTTTGAGGAGATTGCATCTAAACAATAATAAACTGGAACTTCTGCGAGATG ATACCTTCCTTGGCTTGGAGAACCTGGAGTACCTACAGGTCGATTACAACTACATCAG CGTCATTGAACCCAATGCTTTTGGGAAACTGCATTTGTTGCAGGTGCTTATCCTCAAT GACAATCTTTTGTCCAGTTTACCCAACAATCTTTTCCGTTTTGTGCCCTTAACGCACT TGGACCTCCGGGGGAACCGGCTGAAACTTCTGCCCTACGTGGGGCTCTTGCAGCACAT

	GGATAAAGTTGTGGAGCTACAGCTGGAGGAAAACCCTTGGAATTGTTCTTGTGAGCTG ATCTCTCTAAAGGATTGGTTGGACAGCATCTCCTATTCAGCCCTGGTGGGGGATGTAG TTTGTGAGACCCCCTTCCGCTTACACGGAAGGGACTTGGACGAGGTATCCAAGCAGGA ACTTTGCCCAAGGAGACTTATTTCTGACTACGAGATGAGGCCGCAGACGCCTTTGAGC ACCACGGGGTATTTACACACCACCCCGGCGTCAGTGAATTCTGTGGCCACTTCTTCCT CTGCTGTTTACAAACCCCCTTTGAAGCCCCCTAAGGGGACTCGCCAACCCAACAAGCC CAGGGTGCGCCCCACCTCTCGGCAGCCCTCTAAGGACTTGGGCTACAGCAACTATGGC CCCAGCATCGCCTATCAGACCAAATCCCCGGTGCCTTTGGAGTGTCCCACCGCGTGCT CTTGCAACCTGCAGATCTCTGATCTGGGCCTCAACGTAAACTGCCAGGAGCGAAAGAT CGAGAGCATCGCTGAACTGCAGCCCAAGCCCTACAATCCCAAGAAAATGTATCTGACA GAGAACTACATCGCTGTCGTGCGCAGGACAGACTTCCTGGAGGCCACGGGGCTGGACC TCCTGCACCTGGGGAATAACCGCATCTCGATGATCCAGGACCGCGCTTTCGGGGATCT CACCAACCTGAGGCGCCTCTACCTGAATGGCAACAGGATCGAGAGGCTGAGCCCGGAG TTATTCTATGGCCTGCAGAGCCTGCAGTATCTCTTCCTCCAGTACAATCTCATCCGCG AGATTCAGTCTGGAACTTTTGACCCGGTCCCAAACCTCCAGCTGCTATTCTTGAATAA CAACCTCCTGCAGGCCATGCCCTCAGGCGTCTTCTCTGGCTTGACCCTCCTCAGGCTA AACCTGAGGAGTAACCACTTCACCTCCTTGCCAGTGAGTGGAGTTTTGGACCAGCTGA AGTCACTCATCCAAATCGACCTGCATGACAATCCTTGGGATTGTACCTGTGACATTGT GGGCATGAAGCTGTGGGTGGAGCAGCTCAAAGTGGGCGTCCTAGTGGACGAGGTGATC TGTAAGGCGCCCAAAAAATTCGCTGAGACCGACATGCGCTCCATTAAGTCGGAGCTGC TGTGCCCTGACTATTCAGATGTAGTAGTTTCCACGCCCACACCCTCCTCTATCCAGGT CCCTGCGAGGACCAGCGCCGTGACTCCTGCGGTCCGGTTGAATAGCACCGGGGCCCCC GCGAGCTTGGGCGCAGGCGGAGGGGCGTCGTCGGTGCCCTTGTCTGTGTTAATTCTCA GCCTCCTGCTGGTTTTCATCATGTCCGTCTTCGTGGCCGCCGGGCTCTTCGTGCTGGT CATGAAGCGCAGGAAGAAGAACCAGAGCGACCACACCAGCACCAACAACTCCGACGTG AGCTCCTTTAACATGCAGTACAGCGTGTACGGCGGCGGCGGCGGCACGGGCGGCCACC CACACGCGCACGTGCATCACCGCGGGCCCGCGCTGCCCAAGGTGAAGACGCCCGCGGG CCACGTGTATGAATACATCCCCCACCCACTGGGCCACATGTGCAAAAACCCCATCTAC CGCTCCCGAGAGGGCAACTCCGTAGAGGATTACAAAGACCTGCACGAGCTCAAGGTCA CCTACAGCAGCAACCACCACCTGCAGCAGCAGCAGCAGCCGCCGCCGCCACCGCAGCA GCCACAGCAGCAGCCCCCGCCGCAGCTGCAGCTGCAGCCTGGGGAGGAGGAGAGGCGG GAAAGCCACCACTTGCGGAGCCCCGCCTACAGCGTCAGCACCATCGAGCCCCGGGAGG ACCTGCTGTCGCCGGTGCAGGACGCCGACCGCTTTTACAGGGGCATTTTAGAACCAGA CAAACACTGCTCCACCACCCCCGCCGGCAATAGCCTCCCGGAATATCCCAAATTCCCG TGCAGCCCCGCTGCTTACACTTTCTCCCCCAACTATGACCTGAGACGCCCCCATCAGT ATTTGCACCCGGGGGCAGGGGACAGCAGGCTACGGGAACCGGTGCTCTACAGCCCCCC GAGTGCTGTCTTTGTA		
	ORF Start: ATG at 26	ORF Stop: end of sequence	
	SEQ ID NO: 60	925 aa	MW at 103516.1kD
NOV16a, CG92715-01 Protein Sequence	MHTCCPPVTLEQDLHRKMHSWMLQTLAFAVTSLVLSCAETIDYYGEICDNACPCEEKD GILTVSCENRGIISLSEISPPRFPIYHLLLSGNLLNRLYPNEFVNYTGASILHLGSNV IQDIETGAFHGLRGLRRLHLNNNKLELLRDDTFLGLENLEYLQVDYNYISVIEPNAFG KLHLLQVLILNDNLLSSLPNNLFRFVPLTHLDLRGNRLKLLPYVGLLQHMDKVVELQL EENPWNCSCELISLKDWLDSISYSALVGDVVCETPFRLHGRDLDEVSKQELCPRRLIS DYEMRPQTPLSTTGYLHTTPASVNSVATSSSAVYKPPLKPPKGTRQPNKPRVRPTSRQ PSKDLGYSNYGPSIAYQTKSPVPLECPTACSCNLQISDLGLNVNCQERKIESIAELQP KPYNPKKMYLTENYIAVVRRTDFLEATGLDLLHLGNNRISMIQDRAFGDLTNLRRLYL NGNRIERLSPELFYGLQSLQYLFLQYNLIREIQSGTFDPVPNLQLLFLNNNLLQAMPS GVFSGLTLLRLNLRSNHFTSLPVSGVLDQLKSLIQIDLHDNPWDCTCDIVGMKLWVEQ LKVGVLVDEVICKAPKKFAETDMRSIKSELLCPDYSDVVVSTPTPSSIQVPARTSAVT PAVRLNSTGAPASLGAGGGASSVPLSVLILSLLLVFIMSVFVAAGLFVLVMKRRKKNQ SDHTSTNNSDVSSFNMQYSVYGGGGGTGGHPHAHVHHRGPALPKVKTPAGHVYEYIPH PLGHMCKNPIYRSREGNSVEDYKDLHELKVTYSSNHHLQQQQQPPPPPQQPQQQPPPQ LQLQPGEEERRESHHLRSPAYSVSTIEPREDLLSPVQDADRFYRGILEPDKHCSTTPA GNSLPEYPKFPCSPAAYTFSPNYDLRRPHQYLHPGAGDSRLREPVLYSPPSAVFV		
	SEQ ID NO: 61	4500 bp	
NOV16b, CG92715-02 DNA	CGGAACCCGCGGTCGCCACCGCGGCGGCGGCCCCAGGCTGGAGGCGTCCGGGCGCCTC TTTCCTCCAGCCTCTGGGACTGCGCTGCTCGCAGTCTCCTCGCCCTGCCTGGGCTTGA		

Sequence	GAAACCTAGTGCATACCCCAAAGAGGGTTTTTGTGTATGTGTGTGTTTTTAAAGGGTG GCT**ATG**ATGACTGGGCCTTGGAGACGCGGAGACCAAGGAGGTAAAATGCACACTTGCT GCCCCCCAGTAACTTTGGAACAGGACCTTCACAGAAAAATGCATAGCTGGATGCTGCA GACTCTAGCGTTTGCTGTAACATCTCTCGTCCTTTCGTGTGCAGAAACCATCGATTAT TACGGGGAAATCTGTGACAATGCATGTCCTTGTGAGGAAAAGGACGGCATTTTAACTG TGAGCTGTGAAAACCGGGGGATCATCAGTCTCTCTGAAATTAGCCCTCCCCGTTTCCC AATCTACCACCTCTTGTTGTCCGGAAACCTTTTGAACCGTCTCTATCCCAATGAGTTT GTCAATTACACTGGGGCTTCAATTTTGCATCTAGGTAGCAATGTTATCCAGGACATTG AGACCGGGGCTTTCCATGGGCTACGGGGTTTGAGGAGATTGCATCTAAACAATAATAA ACTGGAACTTCTGCGAGATGATACCTTCCTTGGCTTGGAGAACCTGGAGTACCTACAG GTCGATTACAACTACATCAGCGTCATTGAACCCAATGCTTTTGGGAAACTGCATTTGT TGCAGGTGCTTATCCTCAATGACAATCTTTTGTCCAGTTTACCCAACAATCTTTTCCG TTTTGTGCCCTTAACGCACTTGGACCTCCGGGGGAACCGGCTGAAACTTCTGCCCTAC GTGGGGCTCTTGCAGCACATGGATAAAGTTGTGGAGCTACAGCTGGAGGAAAACCCTT GGAATTGTTCTTGTGAGCTGATCTCTCTAAAGGATTGGTTGGACAGCATCTCCTATTC AGCCCTGGTGGGGGATGTAGTTTGTGAGACCCCCTTCCGCTTACACGGAAGGGACTTG GACGAGGTATCCAAGCAGGAACTTTGCCCAAGGAGACTTATTTCTGACTACGAGATGA GGCCGCAGACGCCTTTGAGCACCACGGGGTATTTACACACCACCCCGGCGTCAGTGAA TTCTGTGGCCACTTCTTCCTCTGCTGTTTACAAACCCCCTTTGAAGCCCCCTAAGGGG ACTCGCCAACCCAACAAGCCCAGGGTGCGCCCCACCTCTCGGCAGCCCTCTAAGGACT TGGGCTACAGCAACTATGGCCCCAGCATCGCCTATCAGACCAAATCCCCGGTGCCTTT GGAGTGTCCCACCGCGTGCTCTTGCAACCTGCAGATCTCTGATCTGGGCCTCAACGTA AACTGCCAGGAGCGAAAGATCGAGAGCATCGCTGAACTGCAGCCCAAGCCCTACAATC CCAAGAAAATGTATCTGACAGAGAACTACATCGCTGTCGTGCGCAGGACAGACTTCCT GGAGGCCACGGGGCTGGACCTCCTGCACCTGGGGAATAACCGCATCTCGATGATCCAG GACCGCGCTTTCGGGGATCTCACCAACCTGAGGCGCCTCTACCTGAATGGCAACAGGA TCGAGAGGCTGAGCCCGGAGTTATTCTATGGCCTGCAGAGCCTGCAGTATCTCTTCCT CCAGTACAATCTCATCCGCGAGATTCAGTCTGGAACTTTTGACCCGGTCCCAAACCTC CAGCTGCTATTCTTGAATAACAACCTCCTGCAGGCCATGCCCTCAGGCGTCTTCTCTG GCTTGACCCTCCTCAGGCTAAACCTGAGGAGTAACCACTTCACCTCCTTGCCAGTGAG TGGAGTTTTGGACCAGCTGAAGTCACTCATCCAAATCGACCTGCATGACAATCCTTGG GATTGTACCTGTGACATTGTGGGCATGAAGCTGTGGGTGGAGCAGCTCAAAGTGGGCG TCCTAGTGGACGAGGTGATCTGTAAGGCGCCCAAAAAATTCGCTGAGACCGACATGCG CTCCATTAAGTCGGAGCTGCTGTGCCCTGACTATTCAGATGTAGTAGTTTCCACGCCC ACACCCTCCTCTATCCAGGTCCCTGCGAGGACCAGCGCCGTGACTCCTGCGGTCCGGT TGAATAGCACCGGGGCCCCCGCGAGCTTGGGCGCAGGCGGAGGGGCGTCGTCGGTGCC CTTGTCTGTGTTAATTCTCAGCCTCCTGCTGGTTTTCATCATGTCCGTCTTCGTGGCC GCCGGGCTCTTCGTGCTGGTCATGAAGCGCAGGAAGAAGAACCAGAGCGACCACACCA GCACCAACAACTCCGACGTGAGCTCCTTTAACATGCAGTACAGCGTGTACGGCGGCGG CGGCGGCACGGGCGGCCACCCACACGCGCACGTGCATCACCGCGGGCCCGCGCTGCCC AAGGTGAAGACGCCCGCGGGCCACGTGTATGAATACATCCCCCACCCACTGGGCCACA TGTGCAAAAACCCCATCTACCGCTCCCGAGAGGGCAACTCCGTAGAGGATTACAAAGA CCTGCACGAGCTCAAGGTCACCTACAGCAGCAACCACCACCTGCAGCAGCAGCAGCAG CCGCCGCCGCCACCGCAGCAGCCACAGCAGCAGCCCCCGCCGCAGCTGCAGCTGCAGC CCGGGGAGGAGGAGAGGCGGGAAAGCCACCACTTGCGGAGCCCCGCCTACAGCGTCAG CACCATCGAGCCCCGGGAGGACCTGCTGTCGCCGGTGCAGGACGCCGACCGCTTTTAC AGGGGCATTTTAGAACCAGACAAACACTGCTCCACCACCCCCGCCGGCAATAGCCTCC CGGAATATCCCAAATTCCCGTGCAGCCCCGCTGCTTACACTTTCTCCCCCAACTATGA CCTGAGACGCCCCCATCAGTATTTGCACCCGGGGGCAGGGGACAGCAGGCTACGGGAA CCGGTGCTCTACAGCCCCCCGAGTGCTGTCTTTGTAGAACCCAACCGGAACGAATATC TGGAGTTAAAAGCAAAACTAAACGTTGAGCCGGACTACCTCGAAGTGCTGGAAAAACA GACCACGTTTAGCCAGTTC**TAA**AAGCAAAGAAACTCTCTTGGAGCTTTTGCATTTAAA ACAAACAAGCAAGCAGACACACACAGTGAACACATTTGATTAATTGTGTTGTTTCAAC GTTTAGGGTGAAGTGCCTTGGCACGGGATTTCTCAGCTTCGGTGGAAGATACGAAAAG GGTGTGCAATTTCCTTTAAAATTTACACGTGGGAAACATTTGTGTAAACTGGGCACAT CACTTTCTCTTCTTGCGTGTGGGGCAGGTGTGGAGAAGGGCTTTAAGGAGGCCAATTT GCTGCGCGGGTGACCTGTGAAAGGTCACAGTCATTTTTGTAGTGGTTGGAAGTGCTAA GAATGGTGGATGATGGCAGAGCATAGATTCTACTCTTCCTCTTTAGCTTCCTCCCCAT CCAACGAACCCTGCCCAACACTCTAAATATCCACCAGATAAGACATGGAATGAGGTCT AAATGACACAAAGTGAAGAAATCAACACAACACAAACTTTACAGCTAACAACAAATGA TCAACAAAAACCGAACCAACAAGACAACCATCGAACCTCACCACTCCACACTCACAAC

	AACTCATATCAAGACAACAACACAATGACGTTAAAGGAAACGAAATCAATGCAAAAAT AGACATTTGACAATACAAAAAAACAAGAACCGTGATCACACTACAACCGAAGCAACCA TAGATGTGAGAAAAAACAACAAACAAAACACCGAGCTATATGATCCATAATTGATTAG TCAAAATAACTTATTGATGAAATATACAAATATTTTATTGTAGCACCTATTTTTATAT GCACATTTAGCATTCCTCTTTCCTTCACTATTTAGCCTATGATTTTGCAGAGGTGTCA CACTGTATTAGGATCTGCATTTCTAAAACTGACGTGGTATCAGGAAGGCATTTTCAAT CATTCAAAATGTGGAGAATTTAATGGCTAAATCTTTAAAAGCCAATGCAACCCACCCA ATTGAATCTGCATTTTCTTTTAAGAAAACAGAGCTGATTGTATCCCAATGTATTTTAA AAAATAGGGCAATTGATTGGGCCATTCCGAGAGAATTGTTTGCAAGTTTTGGGTTTTA TTAGAAAATATTTGAAAGTATTTTTATTAATGAACCAAAATGACATGTTCATTTGACT ACTATTGTAGCCGATTTTCGATTGTTTAACCAAACCCAGTTGCATTTGTACAGATCCA CGTGTACTGGCACCTCAGAAGACCAAATCATGGACTGTACAAGTCTCTATACAATGTC TTTATCCCTGTGGGCAGCAAGCAATGATGATAATGACAAACAGGATATCTGTAAGATG GGGCTACTGTTGTTACAGTCTCATATGTATCCCAGCACATGTAATTTTTTAAATAGTT TCTGAATAAACACTTGATAACTATGTCAAAAAAA		
	ORF Start: ATG at 178	ORF Stop: TAA at 3094	
	SEQ ID NO: 62	972 aa	MW at 109043.3kD
NOV16b, CG92715-02 Protein Sequence	MMTGPWRRGDQGGKMHTCCPPVTLEQDLHRKMHSWMLQTLAFAVTSLVLSCAETIDYY GEICDNACPCEEKDGILTVSCENRGIISLSEISPPRFPIYHLLLSGNLLNRLYPNEFV NYTGASILHLGSNVIQDIETGAFHGLRGLRRLHLNNNKLELLRDDTFLGLENLEYLQV DYNYISVIEPNAFGKLHLLQVLILNDNLLSSLPNNLFRFVPLTHLDLRGNRLKLLPYV GLLQHMDKVVELQLEENPWNCSCELISLKDWLDSISYSALVGDVVCETPFRLHGRDLD EVSKQELCPRRLISDYEMRPQTPLSTTGYLHTTPASVNSVATSSSAVYKPPLKPPKGT RQPNKPRVRPTSRQPSKDLGYSNYGPSIAYQTKSPVPLECPTACSCNLQISDLGLNVN CQERKIESIAELQPKPYNPKKMYLTENYIAVVRRTDFLEATGLDLLHLGNNRISMIQD RAFGDLTNLRRLYLNGNRIERLSPELFYGLQSLQYLFLQYNLIREIQSGTFDPVPNLQ LLFLNNNLLQAMPSGVFSGLTLLRLNLRSNHFTSLPVSGVLDQLKSLIQIDLHDNPWD CTCDIVGMKLWVEQLKVGVLVDEVICKAPKKFAETDMRSIKSELLCPDYSDVVVSTPT PSSIQVPARTSAVTPAVRLNSTGAPASLGAGGGASSVPLSVLILSLLLVFIMSVFVAA GLFVLVMKRRKKNQSDHTSTNNSDVSSFNMQYSVYGGGGGTGGHPHAHVHHRGPALPK VKTPAGHVYEYIPHPLGHMCKNPIYRSREGNSVEDYKDLHELKVTYSSNHHLQQQQQP PPPPQQPQQQPPPQLQLQPGEEERRESHHLRSPAYSVSTIEPREDLLSPVQDADRFYR GILEPDKHCSTTPAGNSLPEYPKFPCSPAAYTFSPNYDLRRPHQYLHPGAGDSRLREP VLYSPPSAVFVEPNRNEYLELKAKLNVEPDYLEVLEKQTTFSQF		

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

Table 16B. Comparison of NOV16a against NOV16b.

Protein Sequence	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV16b	1..925 15..939	752/925 (81%) 752/925 (81%)

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

Table 16C. Protein Sequence Properties NOV16a

PSort analysis:	0.8500 probability located in endoplasmic reticulum (membrane); 0.4400

	microbody (peroxisome); 0.3000 probability located in nucleus
SignalP analysis:	Cleavage site between residues 41 and 42

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

Table 16D. Geneseq Results for NOV16a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB95753	Human protein sequence SEQ ID NO:18665 - Homo sapiens, 958 aa. [EP1074617-A2, 07-FEB-2001]	1..925 1..925	925/925 (100%) 925/925 (100%)	0.0
ABB12025	Human IGFALS homologue, SEQ ID NO:2395 - Homo sapiens, 977 aa. [WO200157188-A2, 09-AUG-2001]	1..925 20..944	924/925 (99%) 924/925 (99%)	0.0
AAG67524	Amino acid sequence of a human secreted polypeptide - Homo sapiens, 845 aa. [WO200166690-A2, 13-SEP-2001]	46..925 27..812	423/886 (47%) 563/886 (62%)	0.0
AAE01232	Human gene 1 encoded secreted protein HMIAJ30, SEQ ID NO:94 - Homo sapiens, 845 aa. [WO200134769-A2, 17-MAY-2001]	46..925 27..812	422/886 (47%) 562/886 (62%)	0.0
AAE01312	Human gene 1 encoded secreted protein fragment, SEQ ID NO:177 - Homo sapiens, 596 aa. [WO200134769-A2, 17-MAY-2001]	46..630 6..583	336/594 (56%) 436/594 (72%)	0.0

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

Table 16E. Public BLASTP Results for NOV16a				
Protein Accession Number	**Protein/Organism/Length**	**NOV16a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
O94991	Hypothetical protein KIAA0918 - Homo sapiens (Human), 966 aa (fragment).	1..925 9..933	925/925 (100%) 925/925 (100%)	0.0
Q9H156	BG115M3.1 (NOVEL PROTEIN) - Homo sapiens (Human), 845 aa.	46..925 27..812	423/886 (47%) 563/886 (62%)	0.0
O94933	Hypothetical protein KIAA0848 - Homo sapiens (Human), 977 aa.	9..773 4..764	370/787 (47%) 511/787 (64%)	0.0
Q96JH3	KIAA1854 PROTEIN - Homo sapiens (Human), 572 aa (fragment).	46..599 33..571	325/555 (58%) 419/555 (74%)	0.0
CAB65788	BG256O22.1 (SIMILAR TO IGFALS (INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN, ACID LABILE SUBUNIT)) - Homo sapiens (Human), 853 aa (fragment).	16..713 15..683	348/703 (49%) 468/703 (66%)	0.0

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

Table 16F. Domain Analysis of NOV16a			
Pfam Domain	**NOV16a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
LRRNT: domain 1 of 2	47..82	12/37 (32%) 23/37 (62%)	23
LRR: domain 1 of 10	82..105	9/25 (36%) 13/25 (52%)	4.8e+02
LRR: domain 2 of 10	106..129	5/25 (20%) 16/25 (64%)	80
LRR: domain 3 of 10	130..153	9/25 (36%) 19/25 (76%)	1.6
LRR: domain 4 of 10	154..177	10/25 (40%) 21/25 (84%)	0.0061
LRR: domain 5 of 10	178..200	9/25 (36%) 16/25 (64%)	43
LRR: domain 6 of 10	201..222	10/25 (40%) 17/25 (68%)	48

LRRCT: domain 1 of 2	235..285	18/54 (33%) 34/54 (63%)	3e-08
LRRNT: domain 2 of 2	373..406	10/35 (29%) 19/35 (54%)	0.049
LRR: domain 7 of 10	434..457	9/25 (36%) 17/25 (68%)	0.17
LRR: domain 8 of 10	458..481	10/25 (40%) 22/25 (88%)	0.00064
LRR: domain 9 of 10	482..505	6/25 (24%) 18/25 (72%)	0.096
LRR: domain 10 of 10	506..529	10/25 (40%) 19/25 (76%)	0.0085
LRRCT: domain 2 of 2	563..613	12/54 (22%) 38/54 (70%)	5.6e-05

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: 63	15603 bp
NOV17a, CG92813-01 DNA Sequence	CTAATAGAATTCAGCGGCCGCTTTCCCCGGTGCGCAGTTGTGCTTGGACGTTTGTTCC TCCCTCTTCACGCTCTTCGCTGCGGGTAAGTTCTAAAGTTTCTGAAGGCCGTTCTTTG CAATGATTCCTCATATACCTTAGATACAGGCAACTTCTCCCAACTCTCATCCACCCGG GTGAAAACGCTCAGACTATCTGGATTCAAAAACAAAGTAAAAGGGGGCATATATAAGA GGCTTGAGAAACTTTTCTGGGAACTCAGCTCACAGGAGTGTCCCGCGGAATGCCCTGC CGCTTTTCGCCACAGCATCTCTCTTGCACTCCGCGTTCAACTGGCTACCTAGAGTCTT TTGCTGATGCTACTTGCTTTTGCCGGACTGGAGGTTCTTTGAAATAGCAGAGGTCTCA GACCAAGCCGTCAGCTGAATCTTTGCTGGCGCTCCTTAATCCCTGTAAATATCATTGC GTTTGCTTCACCCCTTCCTTCTCTTTATCACATCGTTTTAGGGAGCCAGGACC**ATG**GA CTTAGCACCAGACAGGGCTACTGGCCGCCCGTGGCTCCCGTTGCACACTCTATCAGTA TCTCAGCTCCTTCGAGTGTTTTGGCTACTGTCATTGCTTCCGGGGCAGGCCTGGGTCC ACGGGGCCGAGCCGCGCCAGGTGTTCCAAGTGCTGGAAGAGCAACCTCCAGGCACTCT GGTAGGCACCATCCAGACGCGCCCCGGCTTCACCTACAGGCTCAGCGAAAGCCACGCC CTGTTTGCCATAAACAGTAGCACCGGAGCCCTGTACACCACCTCCACCATCGACCGCG AGAGCCTGCCCAGCGACGTGATCAACCTGGTGGTCCTTTCCAGCGCGCCCACCTACCC CACCGAAGTGCGAGTGCTGGTGCGGGACCTCAATGACAACGCCCCCGTTTTCCCGGAC CCCTCTATCGTGGTCACTTTCAAGGAAGACAGTAGCAGCGGACGCCAAGTCATCTTAG ACACCGCCACCGACTCGGACATCGGCTCAAACGGTGTGGACCACCGCTCCTACCGCAT CATCCGCGGCAATGAGGCGGGGCGCTTCCGTCTGGACATCAACCTGAACCCGAGCGGC GAGGGAGCGTTCCTGCATCTGGTGTCCAAGGGCGGACTGGACCGTGAGGTCACTCCGC AGTACCAGCTCCTGGTTGAGGTGGAGGACAAGGGTGAGCCTAAGCGGCGGGGCTACCT TCAGGTAAACGTGACTGTGCAAGACATTAATGACAACCCCCCGGTTTTTGGCAGTTCT CACTACCAGGCGGGGGTGCCTGAGGACGCGGTTGTGGGTTCCAGCGTCCTCCAGGTGG CGGCGGCGGACGCGGACGAGGGCACCAACGCGGACATCCGCTATCGCCTGCAGGACGA GGGGACCCCCTTCCAAATGGACCCTGAGACGGGACTTATCACGGTGCGGGAGCCCCTG GACTTCGAAGCTCGGCGCCAATACTCGCTTACGGTGCAGGCGATGGACAGAGGCGTGC CTTCCCTCACTGGGCGCGCCGAGGCGCTGATTCAGCTGCTGGACGTGAATGACAATGA

CCCGGTAGTGAAGTTCCGCTACTTCCCGGCCACCTCGCGCTACGCCTCGGTAGATGAG
AATGCTCAAGTGGGCACCGTGGTGGCTCTGCTCACCGTGACGGACGCAGATTCTCCCG
CGGCCAACGGGAACATCTCCGTGCAAATTCTCGGGGGCAATGAGCAGCGCCACTTTGA
AGTGCAAAGCAGCAAAGTGCCGAACCTGAGCCTAATCAAGGTGGCCAGCGCCTTGGAC
CGCGAGCGCATCCCTTCCTACAACCTCACAGTTTCCGTCTCTGATAACTACGGGGCGC
CCCCTGGCGCAGCAGTCCAGGCGCGCTCTTCTGTGGCAAGCCTGGTGATTTTTGTTAA
TGACATCAATGACCATCCTCCTGTCTTTTCACAGCAAGTGTACAGAGTGAACCTGAGC
GAGGAGGCGCCTCCGGGAAGCTATGTGAGTGGGATATCTGCCACTGATGGCGACTCTG
GTCTCAATGCTAATCTGCGTTACAGCATTGTCTCTGGCAATGGACTGGGATGGTTCCA
TATCAGTGAACATAGCGGCCTCGTGACCACTGGGTCCTCTGGGGGCCTGGACCGTGAA
CTTGCTTCCCAGATTGTTCTGAATATAAGTGCCCGGGACCAGGGAGTTCACCCCAAGG
TGTCCTATGCCCAGCTTGTAGTAACTCTCCTAGATGTGAATGATGAAAAGCCAGTATT
TAGCCAGCCAGAAGGGTATGATGTGTCTGTGGTTGAGAATGCCCCAACAGGGACAGAA
CTGTTGATGCTCAGGGCAACTGACGGGGACCTGGGTGACAACGGAACAGTGCGCTTCT
CCTTACAAGAGGCAGAGACTGACCGGAGGTCCTTCCGTCTGGATCCTGTGTCTGGGAG
GTTGAGTACTATTTCCTCCTTGGACAGAGAAGAGCAAGCCTTCTACTCCCTGTTGGTT
CTGGCCACAGATCTGGGCTCCCCTCCCCAGTCATCAATGGCTCGCATAAATGTGAGTC
TTCTGGATATAAATGATAACAGCCCTGTCTTCTACCCGGTCCAATACTTTGCTCACAT
TAAGGAGAATGAGCCTGGAGGTAGCTACATCACCACTGTGTCTGCCACTGACCCAGAC
TTGGGTACCAATGGTACTGTCAAATATAGCATATCTGCTGGGGACAGGTCTCGGTTTC
AGGTCAATGCTCAGAGTGGGGTTATTTCTACAAGAATGGCCCTAGACAGAGAAGAAAA
AACAGCTTATCAGTTGCAAATAGTAGCTACTGATGGTGGCAATTTACAATCTCCCAAC
CAGGCAATAGTAACCATCACTGTATTGGACACTCAAGACAACCCACCTGTATTCAGTC
AGGTTGCCTACAGCTTTGTGGTTTTTGAGAACGTGGCGCTGGGATATCATGTGGGTAG
TGTGTCTGCATCCACCATGGATCTCAATTCCAACATCAGTTATCTCATTACTACTGGG
GATCAGAAAGGTATGTTTGCTATCAACCAGGTCACTGGGCAGCTTACCACAGCAAATG
TGATTGATAGAGAAGAGCAATCCTTTTATCAGCTGAAGGTAGTGGCCAGTGGGGGCAC
AGTGACTGGAGACACTATGGTTAACATAACAGTTAAGGATTTGAATGACAACTCTCCC
CATTTCCTTCAGGCAATAGAGAGTGTAAATGTGGTGGAGAATTGGCAGGCAGGTCACA
GCATTTTCCAGGCCAAAGCTGTGGACCCTGATGAAGGTGTCAATGGCATGGTACTCTA
TAGTCTGAAGCAAAACCCCAAGAACCTGTTTGCTATCAATGAAAAGAATGGCACTATT
AGTCTGCTTGGGCCCCTGGATGTTCATGCTGGCTCCTACCAAATAGAGATCTTGGCAT
CTGACATGGGTGTCCCACAGCTCTCCTCTAGTGTCATCCTAACAGTTTATGTCCATGA
TGTAAATGACAATTCACCAGTGTTTGACCAACTCTCTTATGAAGTCACCCTTTCTGAG
TCAGAACCTGTGAATTCTCGATTCTTTAAAGTACAAGCTTCTGATAAGGATTCAGGAG
CAAATGATGGTCAATTGTATATAAAAAGTGAACTGGACCGTGAACTTCAAGACAGATA
TGTTTTAATGGTTGTTGCTTCTGACAGAGCAGTGGAACCCCTTAGTGCTACTGTGAAT
GTTACTGTAATTTTAGAAGATGTAAATGATAACAGACCTCTTTTTAACAGTACCAATT
ACACATTTTACTTCGAAGAAGAGCAGAGGGCTGGGTCGTTTGTGGGCAAAGTAAGTGC
TGTAGATAAAGACTTTGGGCCAAATGGAGAAGTAAGGTATTCTTTTGAAATGGTGCAG
CCAGATTTTGAGTTGCATGCCATCAGTGGGGAAATTACAAATACTCATCAGTTTGACA
GGGAGTCTCTTATGAGGCGGAGAGGGACTGCTGTGTTTAGCTTTACAGTCATAGCAAC
AGATCAGGGGATCCCTCAGCCTCTCAAGGATCAGGCCACTGTACATGTTTACATGAAG
GATATAAATGATAATGCTCCCAAATTTTTAAAAGACTTTTACCAAGCTACAATATCAG
AATCAGCAGCCAATCTGACACAAGTGTTAAGAGTATCTGCCTCAGATGTTGATGAAGG
TAATAATGGACTTATTCACTATTCTATAATAAAAGGAAATGAAGAAAGACAGTTTGCT
ATAGACAGTACCTCTGGTCAGGTAACACTAATTGGCAAATTAGACTATGAAGCAACAC
CTGCCTATTCCCTTGTAATTCAAGCAGTGGATTCAGGGACAATCCCCCTCAATTCAAC
GTGTACTTTAAATATTGATATTTTAGATGAAAATGACAATACCCCTTCTTTCCTTAAA
TCAACACTGTTTGTTGATGTTTTGGAAAACATGAGAATTGGTGAACTCGTGTCCTCTG
TTACTGCAACTGATTCCGATTCAGGTGACAATGTTGATTTATATTACAGTATTACTGG
GACTAACAACCACGGAACTTTTAGCATTAGCCCAAACACTGGGAGTATTTTTCTTGCC
AAAAAACTGGACTTTGAAACACAGTCTTTGTATAAATTAAATATAACAGCAAAAGACC
AAGGAAGACCTCCTCGTTCATCTACAATGTCAGTGGTTATTCACGTGAGGGACTTTAA
TGACAATCCTCCTAGCTTTCCTCCTGGAGATATTTTCAAGTCTATTGTTGAGAACATT
CCCATTGGTACATCTGTCATTTCAGTGACTGCACATGACCCTGATGCAGACATTAATG
GTCAACTATCCTACACAATCATTCAACAGATGCCAAGAGGCAACCACTTTACCATAGA
TGAAGTCAAAGGGACTATATATACTAATGCTGAAATAGATCGGGAATTTGCTAATCTC
TTTGAGTTGACTGTAAAAGCCAATGATCAAGCTGTGCCAATAGAAACTAGACGGTATG
CTTTGAAGAACGTGACCATTTTGGTTACAGACCTCAATGACAATGTCCCAATGTTTAT
ATCACAAAACGCCCTTGCTGCAGACCCATCAGCTGTGATTGGTTCCGTTCTGACAACA

ATTATGGCTGCTGACCCAGATGAAGGTGCTAATGGAGAAATAGAGTATGAGATCATCA
ATGGGGACACAGACACCTTCATTGTTGATCGTTATAGTGGAGACCTGAGAGTGGCTTC
AGCGTTGGTGCCTTCACAGTTGATCTACAATCTCATAGTTTCAGCAACAGACCTTGGG
CCTGAAAGGAGGAAATCGACCACTGAATTGACCATCATTCTTCAGGGCCTTGATGGAC
CTGTTTTTACTCAACCCAAATATATAACTATTTTGAAGGAAGGAGAACCCATTGGCAC
AAACGTGATATCAATAGAAGCAGCTAGCCCCAGAGGATCTGAGGCCCCAGTGGAGTAT
TATATTGTTTCAGTTCGTTGTGAAGAAAAAACTGTTGGACGCCTCTTTACTATTGGAC
GACATACTGGTATAATTCAGACCGCAGCCATTCTGGACCGGGAGCAAGGAGCATGTCT
TTACCTGGTGGATGTTTATGCCATAGAAAAATCAACTGCTTTTCCCAGAACACAGAGA
GCAGAGGTAGAAACAACACTTCAGGATATCAATGACAATCCACCAGTATTTCCAACGG
ACATGCTGGATCTCACGGTAGAGGAGAACATTGGAGATGGCTCTAAGATTATGCAGCT
GACAGCCATGGATGCTGACGAGGTGCAAATGCTCTCGTCACATACACTATCATTAGTG
GGTTCTTTGGTAGCAGCCATTTTAGCCACGGATGATGACTCTGGTGTGAATGGAGAAA
TTACATATATTGTGAATGAAGATGATGAAGATGGCATCTTTTTCCTGAATCCTATTAC
TGGGGTCTTTAATTTGACTCGATTATTAGATTATGAAGTACAGCAATATTATATCCTC
ACTGTTCGAGCAGAAGATGGTGGGGGACAATTTACTACCATCAGAGTTTATTTCAATA
TTCTAGATGTAAATGATAATCCACCTATTTTCAGCTTGAATTCATACAGCACATCTTT
AATGGAGAATCTACCTGTGGGATCTACTGTTCTTGTGTTTAATGTTACTGATGCAGAT
ATGATGAAGGCAGAAATAAAGATGTTCTTTGAAACCAGTGAGAACAAAGACACAACAT
ACCAGAATCTCTGGGACACATTCAAAGCAGTGTGTAGAGGGAAATTTATAGCACTAAA
TGCCCACAAGAGAAAGCAGGAAAGATCCAAAATTGACACCCTAACATCACAATTAAAA
GAACTAGAAAAGCAAGAGCAAACACATTCAAAAGCTAGCAGAAGGCAAGAAATAACTA
AAATCAGAGCAGAACTGAAGGATATAGAGACACAAAAAACCCTTCAAAAAATTAATGA
ATCCAGGAGCTGGTTTTTTGAAAGGATCAACAAAATTGATAGACCGCTAGCAAGACTA
ATAAAGAAGAAAACAGAGAAGAATCAAATAGACGCAATAAAAAATGATAAAGGGGATA
TCACCATCGATCCCACAGAAATACAAACTACCATCAGAGAATACTGCAAACACCTCTA
TGCAAATAAACTAGAAAATCTAGAAGAAATGGATAAATTCCTCGACACATACACCCTC
CCAAGACTAAACCAGGAAGAAGTTGAATCTCTGAATAGACCAATAACAGACTCTGAAA
CTGTGGCAATAATCAATAGCTTACCAACCAAAAAGAGTCCAGGACCAGATGGATTCAC
AGCCGAATTCTACCAGATGATAACAACCCCAGTCTTTGCACAAGCTTTGTATAAAGTG
GAGATTAATGAAAACACACTTACTGGAACAGATATAATACAAGTGTTCGCAGCAGATG
GAGATGAAGGCACAAATGGACAGGTTCGCTATGGCATTGTTAATGGTAATACCAATCA
GGAATTTCGGATAGACTCTGTCACAGGTGCCATCACTGTCGCTAAACCTTTGGATAGA
GAAAAGACCCCTACCTACCATTTAACTGTTCAGGCAACAGATCGAGGCAGCACACCCA
GAACTGATACCTCCACGGTCAGCATTGTTCTACTGGATATTAATGACTTTGTTCCTGT
ATTTGAGCTATCTCCATATTCTGTAAATGTCCCTGAGAATTTAGGGACACTACCCAGA
ACAATTCTTCAGACTGCTTCGCCTTGCGTGAGGTTTGCCAGCGCCAGTAAAGCGTATT
TCACAACAATTCCTGAGGATGCACCAACTGGAACAGATGTTTTATTGGTAAATGCCTC
AGATGCTGATGCTTCAAAGAATGCAGTTATAAGTTATAGGATCATCGGTGGAAACTCT
CAGTTCACGATCAACCCATCGACAGGACAAATCATCACCAGCGCATTGTTAGATAGGG
AAACAAAAGATAATTATACTTTGGTAGTGGTCTGCAGTGATGCGGGATCCCCAGAGCC
TCTTTCCAGTTCCACCAGTGTGCTTGTCACTGTGACTGATGTCCATGACAATCCACCA
AGATTTCAGCATCACCCATATGTCACTCACATCCCATCTCCTACTCTTCCAGGTTCCT
TTGTCTTTGCGGTTACAGTCACAGATGCTGATATTGGACCAAATTCTGAACTGCATTA
TTCTCTTTCGGGTAGAAATTCTGAAAAATTTCACATTGACCCACTGAGGGGAGCCATT
ATGGCCGCCGGACCACTAAACGGAGCTTCAGAAGTGACATTTTCTGTGCATGTAAAAG
ATGGTGGCTCATTTCCAAAGACAGATTCTACAACAGTGACTGTTAGATTCGTGAATAA
GGCCGATTTCCCTAAAGTCAGAGCCAAAGAACAAACGTTCATGTTTCCTGAAAACCAA
CCAGTCAGCTCTCTTGTCACCACCATCACAGGATCCTCTTTAAGAGGAGAACCTATGT
CATATTATATCGCAAGTGGGAATCTTGGCAATACTTTCCAGATTGATCAGTTAACAGG
GCAGGTGTCTATTAGTCAACCTCTGGATTTTGAAAAGATACAAAAATATGTTGTATGG
ATAGAGGCCAGAGACGGTGGTTTCCCTCCTTTCTCCTCTTACGAGAAACTTGATATAA
CAGTATTAGATGTCAATGATAATGCCCCAATTTTTAAGGAAGACCCATTTATATCTGA
AATATTGGAAAACCTTTCCCCTCGAAAAATACTTACTGTTTCGGCAATGGACAAGGAC
AGTGGACCCAATGGACAGTTAGATTATGAAATTGTTAATGGCAACATGGAAAATAGTT
TCAGTATCAATCATGCTACTGGTGAAATTAGAAGCGTTAGACCTTTGGACAGGGAAAA
AGTATCTCATTATGTCCTAACCATAAAATCATCAGACAAAGGGTCCCCGTCTCAGAGT
ACTTCAGTAAAAGTCATGATTAACATTTTAGATGAAAATGATAATGCCCCTAGGTTTT
CTCAGATATTTAGTGCCCATGTTCCTGAAAATTCCCCCTTAGGATACACAGTTACCCG
TGTCACAACTTCTGATGAAGACATTGGGATCAATGCAATTAGTAGATATTCTATAATG
GATGCAAGTCTTCCATTTACAATTAATCCCAGCACAGGGGATATTGTCATAAGCAGAC

CTTTAAATAGGGAAGATACAGACCGTTACAGAATTCGAGTTTCCGCACATGATTCTGG
GTGGACTGTAAGTACAGATGTCACAATATTTGTGACAGACATCAATGACAATGCTCCA
AGATTTAGCAGAACTTCCTATTATTTAGATTGCCCTGAACTTACTGAGATTGGCTCCA
AAGTAACTCAGGTATTTGCAACAGATCCTGATGAGGGATCAAATGGACAAGTGTTTTA
TTTCATAAAATCCCAATCAGAATATTTCAGGATTAATGCCACCACTGGAGAGATTTTC
AATAAACAGATCTTAAAATACCAAAATGTCACTGGCTTCAGTAATGTGAATATCAACA
GGCATAGTTTTATAGTGACATCTTCAGATCGAGGTAAACCTTCCTTAATTAGTGAGAC
AACAGTTACCATCAATATAGTGGACAGTAATGACAATGCACCTCAATTTCTTAAAAGT
AAATATTTCACTCCAGTCACCAAAAATGTTAAGGTTGGTACGAAGTTAATCAGAGTTA
CAGCAATAGATGACAAAGATTTTGGACTGAATTCAGAAGTGGAGTATTTCATTTCTAA
TGATAACCATTTAGGAAAATTTAAGTTGGACAATGATACGGGGTGGATTTCAGTAGCA
TCCTCCCTGATTTCTGACTTGAACCAAAACTTTTTTATCACAGTCACTGCAAAGGATA
AGGGAAACCCTCCACTTTCTTCCCAAGCAACTGTTCACATAACTGTCACTGAGGAAAA
CTACCATACACCTGAATTCTCTCAAAGCCACATGAGTGCAACCATCCCTGAGAGCCAT
AGCATTGGGTCCATTGTCAGAACTGTTTCTGCAAGAGATAGAGATGCAGCGATGAATG
GCTTGATTAAGTACAGCATTTCTTCAGGAAATGAAGAAGGCATTTTTGCAATCAATTC
TTCTACAGGTATATTAACACTAGCCAAAGCTCTTGATTATGAGCTATGCCAGAAACAC
GAAATGACGATTAGTGCTATAGATGGAGGATGGGTTGCAAGAACTGGTTACTGCAGTG
TGACCGTAAATGTGATTGATGTGAATGATAATTCTCCAGTATTCCTCTCTGATGACTA
TTTCCCTACTGTTTTGGAAAATGCCCCAAGTGGAACAACAGTTATCCACCTAAATGCA
ACAGATGCTGACTCTGGAACAAATGCTGTGATTGCGTATACTGTACAGTCATCTGACA
GTGACCTCTTTGTCATTGACCCTAACACAGGAGTCATAACCACTCAAGGCTTCTTGGA
TTTTGAAACCAAGCAGAGCTACCATCTTACTGTGAAAGCCTTCAATGTCCCCGATGAG
GAAAGGTGTAGCTTTGCCACTGTTAATATACAATTAAAAGGGACAAATGAATATGTGC
CCCGTTTTGTTTCCAAACTTTACTATTTTGAAATCTCAGAAGCAGCTCCTAAAGGTAC
TATTGTTGGAGAAGTGTTTGCTAGCGACCGTGATTTGGGCACTGATGGGGAGGTACAC
TATTTGATTTTTGGTAATAGTCGAAAGAAGGGTTTCCAGATCAATAAGAAGACTGGAC
AGATTTATGTTTCTGGAATTCTTGATCGAAAAAAAGAAGAAAGGGTGTCTTTGAAGGT
ATTGGCCAAGAACTTTGGCAGCATTAGAGGTGCAGATATAGATGAGGTCACTGTAAAT
GTCACCGTGCTTGATGCAAATGACCCACCCATTTTTACTCTAAACATCTACAGTGTGC
AGATCAGTGAAGGGGTCCCAATAGGAACTCATGTGACCTTTGTCAGTGCCTTTGACTC
AGACTCCATCCCCAGCTGGAGCAGGTTTTCTTACTTCATCGGATCAGGGAATGAAAAT
GGTGCCTTTTCTATTAATCCGCAGACAGGACAGATCACCGTTACTGCAGAATTAGATC
GAGAAACCCTTCCCATCTATAATCTCTCAGTTTTGGCTGTTGATTCAGGGACCCCCTC
AGCTACAGGTAGTGCCTCTTTATTAGTCACCCTGGAAGATATAAATGATAACGGGCCC
ATGCTGACTGTCAGTGAAGGAGAAGTCATGGAAAACAAACGGCCAGGCACTTTGGTGA
TGACCCTTCAGTCCACTGACCCTGATCTCCCTCCAAATCAAGGTCCCTTTACTTATTA
CTTGCTGAGCACAGGTCCTGCCACCAGTTATTTCAGTCTGAGCACTGCTGGAGTTCTG
AGCACAACCAGAGAGATTGACAGAGAGCAGATTGCAGACTTCTATCTGTCTGTGGTTA
CCAAGGATTCTGGTGTTCCTCAAATGTCTTCCACAGGAACTGTGCATATCACAGTTAT
AGACCAAAATGACAATCCTTCACAGTCTCGGACGGTGGAGATATTTGTTAATTATTAT
GGTAACTTGTTTCCCGGTGGGATTTTAGGCTCTGTGAAGCCACAGGATCCAGATGTGT
TAGACAGCTTCCACTGCTCCCTTACTTCAGGAGTTACCAGCCTCTTCAGTATTCCAGG
GGGTACTTGTGATCTGAATTCCCAGCCAAGGTCCACAGATGGCACGTTTGATCTGACT
GTCCTTAGCAATGATGGAGTTCACAGCACAGTCACGAGCAACATCCGAGTTTTCTTTG
CTGGATTTTCCAATGCCACAGTGGATAACAGCATCTTACTTCGTCTCGGCGTACCAAC
AGTAAAGGACTTCTTGACCAACCACTATCTTCATTTTTTACGCATTGCCAGCTCACAG
CTGACAGGCTTAGGGACTGCTGTGCAACTGTACAGTGCATATGAAGAGAACAATAGAA
CGTTTCTTTTGGCAGCTGTGAAGCGAAATCATAATCAGTATGTGAATCCCAGTGGCGT
AGCCACCTTCTTTGAAAGCATCAAAGAGATCCTTCTCCGGCAGAGTGGAGTAAAGGTG
GAATCTGTGGATCATGACTCCTGTGTGCATGGCCCATGTCAGAATGGAGGGAGCTGTC
TACGAAGATTGGCTGTGAGCTCCGTATTAAAAAGCCGTGAGAGTCTTCCAGTCATCAT
CGTGGCAAATGAACCTCTGCAGCCTTTCTTATGCAAGTGTCTGCCAGGATATGCGGGT
AGCTGGTGTGAAATAGATATAGATGAATGTCTTCCATCACCTTGCCACAGTGGTGGAA
CCTGTCACAATTTAGTGGGAGGATTTTCATGCAGCTGCCCAGATGGCTTCACTGGTAG
GGCGTGTGAGAGAGATATCAATGAGTGCCTGCAGAGTCCTTGCAAGAATGGTGCCATC
TGCCAGAATTTTCCAGGAAGCTTCAACTGTGTTTGCAAAACTGGATACACAGGTATGA
CAACGTTTGTACTTTTCTCACTAAGACTTGGAAAATGTGTGAATCTTCAGTCAATTAC
TGTGAATGCAACCCCTGCTTTAATGGTGGTTCCTGCCAAAGTGGTGTGGATTCTTATT
ATTGTCATTGTCCATTTGGTGTCTTTGGAACACTGCGAGTTGAACAGTTATGGATTTG
AGGAGTTATCATACATGGAATTTCCAAGCTTGGACCCCAATAACAACTATATTTATGT

CAAATTTGCCACGATTAAAAGTCATGCCTTATTGCTTTACAACTATGACAACCAGACA
GGCGACCGGGCTGAGTTTTTGGCCCTTGAAATTGCCGAAGAAAGACTAAGATTCTCTT
ATAATTTAGGCAGTGGTACATATAAGCTCACCACCATGAAGAAGGTGTCAGATGGACA
TTTTCACACTGTGATTGCCAGGAGAGCAGGAATGGCAGCCTCCTTAACTGTGGACTCC
TGTTCTGAGAACCAAGAGCCAGGATATTGTACTGTCAGTAATGTGGCAGTTTCAGATG
ACTGGACTCTTGATGTTCAGCCAAATAGAGTTACAGTTGGAGGTATCAGATCTCTAGA
ACCAATCCTTCAGAGAAGAGGACACGTGGAAAGCCATGATTTTGTTGGGTGTATAATG
GAGTTTGCAGTCAATGGAAGGCCTCTGGAACCCAGCCAAGCTTTGGCAGCACAAGGCA
TCCTAGATCAGTATGGCGATTTTATTTCTTACTGTTTTAAAGAAAAAAAATGCAAAAA
AGTATGCTTCACTGTTACTCCTGACACTGCCTTATCATTAGAAGGCAAAGGGCGCTTG
GACTACCACATGAGTCAGAATGAGAAGCGGGAATATTTGTTAAGGCAAAGCTTACGAG
GTGCCATGTTGGAGCCTTTTGGTGTGAACAGTCTGGAAGTAAAATTTAGGACCAGAAG
CGAGAATGGCGTTTTAATCCATATCCAAGAAAGCAGCAATTACACTACTGTGAAGGGA
ATGTGTGAATCTTCAGTCAATTACTGTGAATGCAACCCCTGCTTTAATGGTGGTTCCT
GCCAAAGTGGTGTGGATTCTTATTATTGTCATTGTCCATTTGGTGTCTTTGGAAAACA
CTGCGAGTTGAACAGTTATGGATTTGAGGAGTTATCATACATGGAATTTCCAAGCTTG
GACCCCAATAACAACTATATTTATGTCAAATTTGCCACGATTAAAAGTCATGCCTTAT
TGCTTTACAACTATGACAACCAGACAGGCGACCGGGCTGAGTTTTTGGCCCTTGAAAT
TGCCGAAGAAAGACTAAGATTCTCTTATAATTTAGGCAGTGGTACATATAAGCTCACC
ACCATGAAGAAGGTGTCAGATGGACATTTTCACACTGTGATTGCCAGGAGAGCAGGAA
TGACTCTTGATGTTCAGCCAAATAGAGTTACAGTTGGAGGTATCAGATCTCTAGAACC
AATCCTTCAGAGAAGAGGACACGTGGAAAGCCATGATTTTGTTGGGTGTATAATGGAG
TTTGCAGTCAATGGAAGGCCTCTGGAACCCAGCCAAGCTTTGGCAGCACAAGGCATCC
TAGATCAGTATGGCGATTTTATTTCTTACTGTTTTAAAGAAAAAAAATGCAAAAAGTA
TGCTTCACTTGGCCTCCATCTCGGGAAGCATAGCTTGGCCTCCATCTCAAAAACAGAT
CCCTCAGTGAAGATTGGCTGCCGTGGCCCGAACATTTGTGCCAGCAACCCCTGCTGGG
GTGATTTGCTGTGCATTAATCAGTGGTATGCCTACAGGTGTGTCCCTCCTGGGGACTG
TGCCTCCCACCCGTGCCAGAATGGTGGCAGCTGTGAGCCAGGCCTGCACTCCGGCTTC
ACCTGTAGCTGCCCAGACTCGCACACGGGAAGGACCTGTGAGATGGTGGTGGCCTGTC
TTGGCGTCCTCTGTCCTCAGGGGAAGGTGTGCAAAGCTGGAAGTCCTGCGGGGCATGT
CTGTGTTCTGAGTCAGGGCCCTGAAGAGATCTCTCTGCCTTTGTGGGCTGTGCCTGCC
ATCGTGGGCAGCTGCGCAACCGTCTTGGCCCTCCTGGTCCTTAGCCTGATCCTGTGTA
ACCAGTGCAGGGGGAAGAAGGCCAAAAATCCCAAAGAGGAGAAGAAACCGAAGGAGAA
GAAGAAAAAGGGAAGTGAGAACGTTGCTTTTGATGACCCTGACAATATCCCTCCCTAT
GGGGATGACATGACTGTGAGGAAGCAGCCTGAAGGGAACCCAAAACCAGATATCATTG
AAAGGGAAAACCCCTACCTTATCTATGATGAAACTGATATTCCTCACAACTCAGAAAC
CATCCCCAGCGCCCCTTTGGCATCTCCAGAGCAGGAGATAGAGCACTATGACATTGAC
AACGCCAGCAGCATCGCCCCTTCGGATGCAGACATCATTCAACACTACAAGCAGTTCC
GCAGCCACACACCAAAATTTTCAATCCAGAGGCACAGTCCCCTAGGCTTTGCAAGGCA
ATCCCCCATGCCCTTAGGAGCAAGCAGTTTGACTTACCAGCCTTCATATGGTCAAGGT
TTGAGAACCAGCTCCCTAAGCCACTCAGCATGCCCAACTCCCAACCCTCTGTCTCGAC
ACAGTCCAGCCCCTTTCTCCAAATCTTCTACGTTCTATAGAAACAGCCCAGCAAGGGA
ATTGCATCTTCCTATAAGGGATGGTAATACTTTGGAAATGCATGGTGACACCTGCCAA
CCTGGCATTTTCAACTATGCCACAAGGCTGGGAAGGAGAAGCAAGAGTCCTCAGGCCA
TGGCATCACATGGTTCTAGACCAGGGAGTCGCCTAAAGCAGCCGATTGGGCAGATTCC
ACTGGAATCTTCTCCTCCAGTCGGACTTTCTATTGAAGAAGTGGAGAGGCTCAACACA
CCTCGCCCTAGAAACCCAAGTATCTGCAGTGCAGACCATGGGAGGTCTTCTTCAGAGG
AGGACTGCAGAAGGCCACTGTCTAGAACAAGGAATCCAGCGGATGGCATTCCAGCTCC
AGAATCCTCTTCTGATAGTGACTCCCATGAATCTTTCACTTGCTCAGAAATGGAATAT
GACAGGGAGAAGCCAATGGTATATACTTCCAGAATGCCCAAATTATCTCAAGTCAATG
AATCTGATGCAGATGATGAAGATAATTATGGAGCCAGACTGAAGCCTCGAAGGTACCA
CGGTCGCAGGGCCGAGGGAGGACCTGTGGGCACCCAGGCAGCAGCACCAGGCACTGCT
GACAACACACTGCCCATGAAGCTAGGGCAGCAAGCAGGGACTTTCAACTGGGACAACC
TTTTGAACTGGGGCCCTGGCTTTGGCCATTATGTAGATGTTTTTAAAGATTTGGCATC
TCTTCCAGAAAAAGCAGCAGCAAATGAAGAAGGCAAAGCTGGGACAACTAAACCAGTC
CCCAAAGATGGGGAAGCAGAACAGTATGTG**TGA**AGTTTATGTACTGGCACTATAAAAT
ATAAAAACAAGAAATAATACTCAAACCATTGTAAAGTTGCTGACTAGGTTGGGTCACA
TTTGAAAAACAGGCCAGTATGGACTAGTGGTGGAGGGAAAACTTTAAAAATAATAACC
ACAATGCTGCTGAAACAGACTCACAACAACTCTTAATTTAAACATGTGTGGTTGAATT
C

	ORF Start: ATG at 518	ORF Stop: TGA at 15401	
	SEQ ID NO: 64	4961 aa	MW at 543673.9kD
NOV17a, CG92813-01 Protein Sequence	MDLAPDRATGRPWLPLHTLSVSQLLRVFWLLSLLPGQAWVHGAEPRQVFQVLEEQPPG TLVGTIQTRPGFTYRLSESHALFAINSSTGALYTTSTIDRESLPSDVINLVVLSSAPT YPTEVRVLVRDLNDNAPVFPDPSIVVTFKEDSSSGRQVILDTATDSDIGSNGVDHRSY RIIRGNEAGRFRLDINLNPSGEGAFLHLVSKGGLDREVTPQYQLLVEVEDKGEPKRRG YLQVNVTVQDINDNPPVFGSSHYQAGVPEDAVVGSSVLQVAAADADEGTNADIRYRLQ DEGTPFQMDPETGLITVREPLDFEARRQYSLTVQAMDRGVPSLTGRAEALIQLLDVND NDPVVKFRYFPATSRYASVDENAQVGTVVALLTVTDADSPAANGNISVQILGGNEQRH FEVQSSKVPNLSLIKVASALDRERIPSYNLTVSVSDNYGAPPGAAVQARSSVASLVIF VNDINDHPPVFSQQVYRVNLSEEAPPGSYVSGISATDGDSGLNANLRYSIVSGNGLGW FHISEHSGLVTTGSSGGLDRELASQIVLNISARDQGVHPKVSYAQLVVTLLDVNDEKP VFSQPEGYDVSVVENAPTGTELLMLRATDGDLGDNGTVRFSLQEAETDRRSFRLDPVS GRLSTISSLDREEQAFYSLLVLATDLGSPPQSSMARINVSLLDINDNSPVFYPVQYFA HIKENEPGGSYITTVSATDPDLGTNGTVKYSISAGDRSRFQVNAQSGVISTRMALDRE EKTAYQLQIVATDGGNLQSPNQAIVTITVLDTQDNPPVFSQVAYSFVVFENVALGYHV GSVSASTMDLNSNISYLITTGDQKGMFAINQVTGQLTTANVIDREEQSFYQLKVVASG GTVTGDTMVNITVKDLNDNSPHFLQAIESVNVVENWQAGHSIFQAKAVDPDEGVNGMV LYSLKQNPKNLFAINEKNGTISLLGPLDVHAGSYQIEILASDMGVPQLSSSVILTVYV HDVNDNSPVFDQLSYEVTLSESEPVNSRFFKVQASDKDSGANDGQLYIKSELDRELQD RYVLMVVASDRAVEPLSATVNVTVILEDVNDNRPLFNSTNYTFYFEEEQRAGSFVGKV SAVDKDFGPNGEVRYSFEMVQPDFELHAISGEITNTHQFDRESLMRRRGTAVFSFTVI ATDQGIPQPLKDQATVHVYMKDINDNAPKFLKDFYQATISESAANLTQVLRVSASDVD EGNNGLIHYSIIKGNEERQFAIDSTSGQVTLIGKLDYEATPAYSLVIQAVDSGTIPLN STCTLNIDILDENDNTPSFLKSTLFVDVLENMRIGELVSSVTATDSDSGDNVDLYYSI TGTNNHGTFSISPNTGSIFLAKKLDFETQSLYKLNITAKDQGRPPRSSTMSVVIHVRD FNDNPPSFPPGDIFKSIVENIPIGTSVISVTAHDPDADINGQLSYTIIQQMPRGNHFT IDEVKGTIYTNAEIDREFANLFELTVKANDQAVPIETRRYALKNVTILVTDLNDNVPM FISQNALAADPSAVIGSVLTTIMAADPDEGANGEIEYEIINGDTDTFIVDRYSGDLRV ASALVPSQLIYNLIVSATDLGPERRKSTTELTIILQGLDGPVFTQPKYITILKEGEPI GTNVISIEAASPRGSEAPVEYYIVSVRCEEKTVGRLFTIGRHTGIIQTAAILDREQGA CLYLVDVYAIEKSTAFPRTQRAEVETTLQDINDNPPVFPTDMLDLTVEENIGDGSKIM QLTAMDADEVQMLSSHTLSLVGSLVAAILATDDDSGVNGEITYIVNEDDEDGIFFLNP ITGVFNLTRLLDYEVQQYYILTVRAEDGGGQFTTIRVYFNILDVNDNPPIFSLNSYST SLMENLPVGSTVLVFNVTDADMMKAEIKMFFETSENKDTTYQNLWDTFKAVCRGKFIA LNAHKRKQERSKIDTLTSQLKELEKQEQTHSKASRRQEITKIRAELKDIETQKTLQKI NESRSWFFERINKIDRPLARLIKKKTEKNQIDAIKNDKGDITIDPTEIQTTIREYCKH LYANKLENLEEMDKFLDTYTLPRLNQEEVESLNRPITDSETVAIINSLPTKKSPGPDG FTAEFYQMITTPVFAQALYKVEINENTLTGTDIIQVFAADGDEGTNGQVRYGIVNGNT NQEFRIDSVTGAITVAKPLDREKTPTYHLTVQATDRGSTPRTDTSTVSIVLLDINDFV PVFELSPYSVNVPENLGTLPRTILQTASPCVRFASASKAYFTTIPEDAPTGTDVLLVN ASDADASKNAVISYRIIGGNSQFTINPSTGQIITSALLDRETKDNYTLVVVCSDAGSP EPLSSSTSVLVTVTDVHDNPPRFQHHPYVTHIPSPTLPGSFVFAVTVTDADIGPNSEL HYSLSGRNSEKFHIDPLRGAIMAAGPLNGASEVTFSVHVKDGGSFPKTDSTTVTVRFV NKADFPKVRAKEQTFMFPENQPVSSLVTTITGSSLRGEPMSYYIASGNLGNTFQIDQL TGQVSISQPLDFEKIQKYVVWIEARDGGFPPFSSYEKLDITVLDVNDNAPIFKEDPFI SEILENLSPRKILTVSAMDKDSGPNGQLDYEIVNGNMENSFSINHATGEIRSVRPLDR EKVSHYVLTIKSSDKGSPSQSTSVKVMINILDENDNAPRFSQIFSAHVPENSPLGYTV TRVTTSDEDIGINAISRYSIMDASLPFTINPSTGDIVISRPLNREDTDRYRIRVSAHD SGWTVSTDVTIFVTDINDNAPRFSRTSYYLDCPELTEIGSKVTQVFATDPDEGSNGQV FYFIKSQSEYFRINATTGEIFNKQILKYQNVTGFSNVNINRHSFIVTSSDRGKPSLIS ETTVTINIVDSNDNAPQFLKSKYFTPVTKNVKVGTKLIRVTAIDDKDFGLNSEVEYFI SNDNHLGKFKLDNDTGWISVASSLISDLNQNFFITVTAKDKGNPPLSSQATVHITVTE ENYHTPEFSQSHMSATIPESHSIGSIVRTVSARDRDAAMNGLIKYSISSGNEEGIFAI NSSTGILTLAKALDYELCQKHEMTISAIDGGWVARTGYCSVTVNVIDVNDNSPVFLSD DYFPTVLENAPSGTTVIHLNATDADSGTNAVIAYTVQSSDSDLFVIDPNTGVITTQGF LDFETKQSYHLTVKAFNVPDEERCSFATVNIQLKGTNEYVPRFVSKLYYFEISEAAPK GTIVGEVFASDRDLGTDGEVHYLIFGNSRKKGFQINKKTGQIYVSGILDRKKEERVSL KVLAKNFGSIRGADIDEVTVNVTVLDANDPPIFTLNIYSVQISEGVPIGTHVTFVSAF		

	DSDSIPSWSRFSYFIGSGNENGAFSINPQTGQITVTAELDRETLPIYNLSVLAVDSGT PSATGSASLLVTLEDINDNGPMLTVSEGEVMENKRPGTLVMTLQSTDPDLPPNQGPFT YYLLSTGPATSYFSLSTAGVLSTTREIDREQIADFYLSVVTKDSGVPQMSSTGTVHIT VIDQNDNPSQSRTVEIFVNYYGNLFPGGILGSVKPQDPDVLDSFHCSLTSGVTSLFSI PGGTCDLNSQPRSTDGTFDLTVLSNDGVHSTVTSNIRVFFAGFSNATVDNSILLRLGV PTVKDFLTNHYLHFLRIASSQLTGLGTAVQLYSAYEENNRTFLLAAVKRNHNQYVNPS GVATFFESIKEILLRQSGVKVESVDHDSCVHGPCQNGGSCLRRLAVSSVLKSRESLPV IIVANEPLQPFLCKCLPGYAGSWCEIDIDECLPSPCHSGGTCHNLVGGFSCSCPDGFT GRACERDINECLQSPCKNGAICQNFPGSFNCVCKTGYTGMTTFVLFSLRLGKCVNLQS ITVNATPALMVVPAKVVWILIIVIVHLVSLEHCELNSYGFEELSYMEFPSLDPNNNYI YVKFATIKSHALLLYNYDNQTGDRAEFLALEIAEERLRFSYNLGSGTYKLTTMKKVSD GHFHTVIARRAGMAASLTVDSCSENQEPGYCTVSNVAVSDDWTLDVQPNRVTVGGIRS LEPILQRRGHVESHDFVGCIMEFAVNGRPLEPSQALAAQGILDQYGDFISYCFKEKKC KKVCFTVTPDTALSLEGKGRLDYHMSQNEKREYLLRQSLRGAMLEPFGVNSLEVKFRT RSENGVLIHIQESSNYTTVKGMCESSVNYCECNPCFNGGSCQSGVDSYYCHCPFGVFG KHCELNSYGFEELSYMEFPSLDPNNNYIYVKFATIKSHALLLYNYDNQTGDRAEFLAL EIAEERLRFSYNLGSGTYKLTTMKKVSDGHFHTVIARRAGMTLDVQPNRVTVGGIRSL EPILQRRGHVESHDFVGCIMEFAVNGRPLEPSQALAAQGILDQYGDFISYCFKEKKCK KYASLGLHLGKHSLASISKTDPSVKIGCRGPNICASNPCWGDLLCINQWYAYRCVPPG DCASHPCQNGGSCEPGLHSGFTCSCPDSHTGRTCEMVVACLGVLCPQGKVCKAGSPAG HVCVLSQGPEEISLPLWAVPAIVGSCATVLALLVLSLILCNQCRGKKAKNPKEEKKPK EKKKKGSENVAFDDPDNIPPYGDDMTVRKQPEGNPKPDIIERENPYLIYDETDIPHNS ETIPSAPLASPEQEIEHYDIDNASSIAPSDADIIQHYKQFRSHTPKFSIQRHSPLGFA RQSPMPLGASSLTYQPSYGQGLRTSSLSHSACPTPNPLSRHSPAPFSKSSTFYRNSPA RELHLPIRDGNTLEMHGDTCQPGIFNYATRLGRRSKSPQAMASHGSRPGSRLKQPIGQ IPLESSPPVGLSIEEVERLNTPRPRNPSICSADHGRSSSEEDCRRPLSRTRNPADGIP APESSSDSDSHESFTCSEMEYDREKPMVYTSRMPKLSQVNESDADDEDNYGARLKPRR YHGRRAEGGPVGTQAAAPGTADNTLPMKLGQQAGTFNWDNLLNWGPGFGHYVDVFKDL ASLPEKAAANEEGKAGTTKPVPKDGEAEQYV

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

Table 17B. Protein Sequence Properties NOV17a

PSort analysis:	0.8000 probability located in nucleus; 0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 43 and 44

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 17C.

Table 17C. Geneseq Results for NOV17a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV17a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABG22977	Novel human diagnostic protein #22968 - Homo sapiens, 4591 aa. [WO200175067-A2, 11-OCT-2001]	119..4188 134..4099	1159/4320 (26%) 1853/4320 (42%)	0.0
ABG22977	Novel human diagnostic protein #22968 - Homo sapiens, 4591 aa. [WO200175067-A2, 11-OCT-2001]	119..4188 134..4099	1159/4320 (26%) 1853/4320 (42%)	0.0
AAM52106	Rat fat 3 protein SEQ ID NO 3 - Rattus norvegicus, 4555 aa. [JP2001258573-A, 25-SEP-2001]	70..3771 192..3829	1054/3877 (27%) 1742/3877 (44%)	0.0
AAU07054	Human Flamingo protein encoded by cDNA splice variant - Homo sapiens, 2923 aa. [WO200161003-A1, 23-AUG-2001]	2627..4042 162..1566	427/1490 (28%) 665/1490 (43%)	e-137
AAU07053	Human Flamingo polypeptide - Homo sapiens, 2956 aa. [WO200161003-A1, 23-AUG-2001]	2627..4042 162..1566	427/1490 (28%) 665/1490 (43%)	e-137

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

5

Table 17D. Public BLASTP Results for NOV17a				
Protein Accession Number	**Protein/Organism/Length**	**NOV17a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
P33450	Cadherin-related tumor suppressor precursor (Fat protein) - Drosophila melanogaster (Fruit fly), 5147 aa.	49..1878 71..1951	748/1932 (38%) 1066/1932 (54%)	0.0
IJFFTM	cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster), 5147 aa.	49..1878 71..1951	744/1932 (38%) 1065/1932 (54%)	0.0

Q96JQ0	Protocadherin 16 precursor (Cadherin 19) (Cadherin fibroblast 1) - Homo sapiens (Human), 3298 aa.	5..1877 12..1883	621/1950 (31%) 925/1950 (46%)	0.0
Q99PF4	Cadherin 23 precursor (Otocadherin) - Mus musculus (Mouse), 3354 aa.	53..1877 588..2430	606/1927 (31%) 896/1927 (46%)	0.0
P58365	Cadherin 23 precursor (Otocadherin) - Rattus norvegicus (Rat), 3317 aa.	53..1877 586..2428	613/1928 (31%) 897/1928 (45%)	0.0

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

Table 17E. Domain Analysis of NOV17a

Pfam Domain	NOV17a Match Region	Identities/ Similarities for the Matched Region	Expect Value
cadherin: domain 1 of 30	47..126	19/107 (18%) 59/107 (55%)	0.012
cadherin: domain 2 of 30	140..241	36/113 (32%) 76/113 (67%)	4.6e-15
cadherin: domain 3 of 30	255..344	43/107 (40%) 75/107 (70%)	1.3e-27
cadherin: domain 4 of 30	363..466	43/113 (38%) 79/113 (70%)	2.9e-23
cadherin: domain 5 of 30	480..573	36/109 (33%) 75/109 (69%)	3.5e-24
cadherin: domain 6 of 30	588..680	46/108 (43%) 76/108 (70%)	5.3e-27
cadherin: domain 7 of 30	694..784	43/107 (40%) 67/107 (63%)	7.2e-27
cadherin: domain 8 of 30	798..884	34/107 (32%) 66/107 (62%)	4.2e-18
cadherin: domain 9 of 30	898..987	31/107 (29%) 66/107 (62%)	1.3e-18
cadherin: domain 10 of 30	1001..1071	26/107 (24%) 59/107 (55%)	1.6e-08
cadherin: domain 11 of 30	1085..1181	34/111 (31%) 72/111 (65%)	1.6e-14
Isochorismatase: domain 1 of 1	1038..1206	27/213 (13%) 112/213 (53%)	8.8

cadherin: domain 12 of 30	1195..1286	38/107 (36%) 76/107 (71%)	3.2e-27
cadherin: domain 13 of 30	1300..1391	41/107 (38%) 70/107 (65%)	7.3e-27
cadherin: domain 14 of 30	1405..1500	38/108 (35%) 73/108 (68%)	5.3e-19
cadherin: domain 15 of 30	1506..1602	29/114 (25%) 70/114 (61%)	6.8e-12
cadherin: domain 16 of 30	1614..1711	28/112 (25%) 63/112 (56%)	0.014
S-AdoMet_syntD2: domain 1 of 1	1789..1803	8/15 (53%) 12/15 (80%)	3.8
cadherin: domain 17 of 30	1754..1840	31/107 (29%) 64/107 (60%)	2.2e-14
cadherin: domain 18 of 30	2107..2198	45/107 (42%) 76/107 (71%)	6.4e-31
cadherin: domain 19 of 30	2244..2334	44/107 (41%) 72/107 (67%)	2.2e-28
cadherin: domain 20 of 30	2348..2436	34/107 (32%) 66/107 (62%)	1.8e-12
cadherin: domain 21 of 30	2449..2537	36/107 (34%) 65/107 (61%)	1.2e-11
cadherin: domain 22 of 30	2551..2641	37/107 (35%) 72/107 (67%)	9.7e-26
cadherin: domain 23 of 30	2654..2740	38/107 (36%) 63/107 (59%)	3.2e-17
cadherin: domain 24 of 30	2754..2851	31/116 (27%) 74/116 (64%)	5.3e-16
cadherin: domain 25 of 30	2865..2957	40/107 (37%) 68/107 (64%)	4.3e-16
cadherin: domain 26 of 30	2971..3062	37/107 (35%) 74/107 (69%)	3.2e-26
cadherin: domain 27 of 30	3076..3164	36/108 (33%) 67/108 (62%)	4.2e-21
cadherin: domain 28 of 30	3180..3273	33/107 (31%) 71/107 (66%)	1.8e-16
cadherin: domain 29 of 30	3286..3378	38/107 (36%) 78/107 (73%)	2e-27

PEP-utilizers: domain 1 of 1	3326..3392	17/107 (16%) 43/107 (40%)	5.2
cadherin: domain 30 of 30	3390..3482	37/109 (34%) 75/109 (69%)	1e-21
EGF: domain 1 of 5	3683..3736	14/64 (22%) 33/64 (52%)	7.1
EGF: domain 2 of 5	3743..3774	18/47 (38%) 25/47 (53%)	2.1e-08
metalthio: domain 1 of 1	3745..3805	16/70 (23%) 29/70 (41%)	7.8
EGF: domain 3 of 5	3781..3812	15/47 (32%) 25/47 (53%)	8.9e-05
EB: domain 1 of 1	3765..3823	16/70 (23%) 41/70 (59%)	3.7
laminin_G: domain 1 of 4	3890..4033	48/163 (29%) 99/163 (61%)	5.8e-19
laminin_G: domain 2 of 4	4116..4125	5/10 (50%) 10/10 (100%)	5.6
EGF: domain 4 of 5	4148..4179	16/47 (34%) 24/47 (51%)	0.013
laminin_EGF: domain 1 of 1	4148..4195	15/64 (23%) 33/64 (52%)	2
laminin_G: domain 3 of 4	4208..4272	26/77 (34%) 48/77 (62%)	4.5e-09
laminin_G: domain 4 of 4	4286..4322	14/37 (38%) 27/37 (73%)	0.0066
EGF: domain 5 of 5	4410..4442	16/47 (34%) 26/47 (55%)	3e-06

Example 18.

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: 65	2118 bp
NOV18a, CG92844-01 DNA Sequence	GGAGTGAACTAACACAAAGTTAGGAATTCTTATTGGTTTACAAAAACA**ATG**TTGATTA GTGATTGGCTATATACATTGTTAAGGTATAGGGTGTGGCTCGCCGGTGCCCTCGACCT GCCCGCTGGGTCCTGTGCCTTTGAAGAGAGCACTTGCGGCTTTGACTCCGTGTTGGCC TCTCTGCCGTGGATTTTAAATGAGGAAGGCCATTACATTTATGTGGATACCTCCTTTG	

	GCAAGCAGGGGGAGAAAGCTGTGCTGCTAAGTCCTGACTTACAGGCTGAGGAATGGAG CTGCCTCCGTTTGGTCTACCAGATAACCACATCTTCGGAGTCTCTGTCAGATCCCAGC CAGCTGAACCTCTACATGAGATTTGAAGATGAAAGCTTTGATCGCTTGCTTTGGTCAG CTAAGGAACCTTCAGACAGCTGGCTCATAGCCAGCTTGGATTTGCAAAACAGTTCCAA GAAATTCAAGATTTTAATAGAAGGTGTACTAGGACAGGGAAACACAGCCAGCATCGCA CTATTTGAAATCAAGATGACAACCGGCTACTGTATTGAATGTGACTTTGAAGAAAATC ATCTCTGTGGCTTTGTGAACCGCTGGAATCCCAATGTGAACTGGTTTGTTGGAGGAGG AAGTATTCGGAATGTCCACTCCATTCTCCCACAGGATCACACCTTCAAGAGTGAACTG GGTCACTACATGTACGTGGACTCAGTTTATGTGAAGCACTTCCAGGAGGTGGCACAGC TCATCTCCCCGTTGACCACGGCCCCCATGGCTGGCTGCCTGTCATTTTATTACCAGAT CCAGCAGGGGAATGACAATGTCTTTTCCCTTTACACTCGGGATGTGGCTGGCCTTTAC GAGGAAATCTGGAAAGCAGACAGGCCAGGGAATGCTGCCTGGAACCTTGCGGAGGTCG AGTTCAGTGCTCATTTTCCTCTGCAGGTTATTTTTGAAGTTGCTTTCAATGGTCCCAA GGGAGGTTATGTTGCCCTGGATGATATTTCATTCTCTCCTGTTCACTGCCAGAATCAG ACAGGTCTTCTGTTCAGTGCCGTGGAAGCCAGCTGCAATTTTGAGCAAGATCTCTGCA ACTTTTACCAAGATAAAGAAGGTCCAGGTTGGACCCGAGTGAAAGTAAAACCAAACAT GTATCGGGCTGGAGACCACACTACAGGCTTAGGTTATTACCTGCTAGCCAACACAAAG TTCACATCTCAGCCTGGCTACATTGGAAGGCTCTATGGGCCCTCCCTACCAGGAAACT TGCAGTATTGTCTGCGTTTTCATTATGCCATCTATGGATTTTTAAAAATGAGTGACAC CCTAGCAGTTTACATCTTTGAAGAGAACCATGTGGTTCAAGAGAAGATCTGGTCTGTG TTGGAGTCCCCAAGGGGTGTTTGGATGCAAGCTGAAATCACCTTTAAGAAGCCCATGC CTTTTCAGGTGGTTTTCATGAGCCTATGCAAAAGTTTCTGGGACTGTGGGCTTGTAGC CCTGGATGACATTACAATACAATTGGGAAGCTGCTCATCTTCAGAGAAACTTCCACCT CCACCTGGAGAGTGTACTTTCGAGCAAGATGAATGTACATTTACTCAGGAGAAAAGAA ACCGGAGCAGCTGGCACAGGAGGAGGGGAGAAACTCCCACTTCCTACACAGGACCAAA GGGAGATCACACTACTGGGGTAGGCTACTACATGTACATTGAGGCCTCCCATATGGTG TATGGACAAAAAGCACGCCTCTTGTCCAGGCCTCTGCGAGGAGTCTCTGGAAAACACT GCTTGACCTTTTTCTACCACATGTATGGAGGGGGCACTGGCCTGCTGAGTGTTTATCT GAAAAAGGAAGAAGACAGTGAAGAGTCCCTCTTATGGAGGAGAAGAGGTGAACAGAGC ATTTCCTGGCTACGAGCACTGATTGAATACAGCTGTGAGAGGCAACACCAGATAATTT TTGAAGCCATTCGAGGAGTATCAATAAGAAGTGATATTGCCATTGATGATGTTAAATT TCAGGCAGGACCCTGTCAATCATCAGGATATTCTGAGGACTTAAATGAAATTGAGTAT **TAA**GAAATGATCTGCATTGGATTTACTAGA
	ORF Start: ATG at 49 ORF Stop: TAA at 2089
	SEQ ID NO: 66 680 aa MW at 77231.5kD
NOV18a, CG92844-01 Protein Sequence	MLISDWLYTLLRYRVWLAGALDLPAGSCAFEESTCGFDSVLASLPWILNEEGHYIYVD TSFGKQGEKAVLLSPDLQAEEWSCLRLVYQITTSSESLSDPSQLNLYMRFEDESFDRL LWSAKEPSDSWLIASLDLQNSSKKFKILIEGVLGQGNTASIALFEIKMTTGYCIECDF EENHLCGFVNRWNPNVNWFVGGGSIRNVHSILPQDHTFKSELGHYMYVDSVYVKHFQE VAQLISPLTTAPMAGCLSFYYQIQQGNDNVFSLYTRDVAGLYEEIWKADRPGNAAWNL AEVEFSAHFPLQVIFEVAFNGPKGGYVALDDISFSPVHCQNQTGLLFSAVEASCNFEQ DLCNFYQDKEGPGWTRVKVKPNMYRAGDHTTGLGYYLLANTKFTSQPGYIGRLYGPSL PGNLQYCLRFHYAIYGFLKMSDTLAVYIFEENHVVQEKIWSVLESPRGVWMQAEITFK KPMPFQVVFMSLCKSFWDCGLVALDDITIQLGSCSSSEKLPPPPGECTFEQDECTFTQ EKRNRSSWHRRRGETPTSYTGPKGDHTTGVGYYMYIEASHMVYGQKARLLSRPLRGVS GKHCLTFFYHMYGGGTGLLSVYLKKEEDSEESLLWRRRGEQSISWLRALIEYSCERQH QIIFEAIRGVSIRSDIAIDDVKFQAGPCQSSGYSEDLNEIEY
	SEQ ID NO: 67 2023 bp
NOV18b, 174308357 DNA Sequence	**GGA**TCCTTTGAAGAGAGCACTTGCGGCTTTGACTCCGTGTTGGCCTCTCTGCCGTGGA TTTTAAATGAGGAAGGCCATTACATTTATGTGGATACCTCCTTTGGCAAGCAGGGGGA GAAAGCTGTGCTGCTAAGTCCTGACTTACAGGCTGAGGAATGGAGCTGCCTCCGTTTG GTCTACCAGATAACCACATCTTCGGAGTCTCTGTCAGATCCCAGCCAGCTGAACCTCT ACATGAGATTTGAAGATGAAAGCTTTGATCGCTTGCTTTGGTCAGCTAAGGAACCTTC AGACAGCTGGCTCATAGCCAGCTTGGATTTGCAAAACAGTTCCAAGAAATTCAAGATT TTAATAGAAGGTGTACTAGGACAGGGAAACACAGCCAGCATCGCACTATTTGAAATCA AGATGACAACCGGCTACTGTATTGAATGTGACTTTGAAGAAAATCATCTCTGTGGCTT TGTGAACCGCTGGAATCCCAATGTGAACTGGTTTGTTGGAGGAGGAAGTATTCGGAAT GTCCACTCCATTCTCCCACAGGATCACACCTTCAAGAGTGAACTGGGCCACTACATGT

	ACGTGGACTCAGTTTATGTGAAGCACTTCCAGGAGGTGGCACAGCTCATCTCCCCGTT GACCACGGCCCCCATGGCTGGCTGCCCGTCATTTTATTACCAGATCCAGCAGGGGAAT GACAATGTCTTTTCCCTTTACACTCGGGATGTGGCTGGCCTTTACGAGGAAATCTGGA AAGCAGACAGGCCAGGGAATGCTGCCTGGAACCTTGCGGAGGTCGAGTTCAATGCTCC TTACCCCATGGAGGTTATTTTTGAAGTTGCTTTCAATGGTCCCAAGGGAGGTTATGTT GCCCTGGATGATATTTCATTCTCTCCTGTTCACTGCCAGAATCAGACAGAACTTCTGT TCAGTGCCGTGGAAGCCAGCTGCAATTTTGAGCAAGATCTCTGCAACTTTTACCAAGA TAAAGAAGGTCCAGGTTGGACCCGAGTGAAAGTAAAACCAAACATGTATCGGGCTGGA GACCACACTACAGGCTTAGGGTATTACCTGCTAGCCAACACAAAGTTCACATCTCAGC CTGGCTACATTGGAAGGCTCTATGGGCCCTCCCTACCAGGAAACTTGCAGTATTGTCT GCGTTTTCATTATGCCATCTATGGATTTTTAAAAATGAGTGACACCCTAGCAGTTTAC ATCTTTGAAGAGAACCATGTGGTTCAAGAGAAGATCTGGTCTGTGTTGGAGTCCCCAA GGGGTGTTTGGATGCAAGCTGAAATCACCTTTAAGAAGCCCATGCCTACCAAGGTGGT TTTCATGAGCCTATGCAAAAGTTTCTGGGACTGTGGGCTTGTAGCCCTGGATGACATT ACAATACAATTGGGAAGCTGCTCATCTTCAGAGAAACTTCCACCTCCACCTGGAGAGT GTACTTTCGAGCAAGATGAATGTACATTTACTCAGGAGAAAAGAAACCGGAGCAGCTG GCACAGGAGGAGGGGAGAAACTCCCACTTCCTACACAGGACCAAAGGGAGATCACACT ACTGGGGTAGGCTACTACATGTACATTGAGGCCTCCCATATGGTGTATGGACAAAAAG CACGCCTCTTGTCCAGGCCTCTGCGAGGAGTCTCTGGAAAACACTGCTTGACCTTTTT CTACCACATGTATGGAGGGGCACTGGCCTGCTGAGTGTTTATCTGAAAAAGGAAGAA GACAGTGAAGAGTCCCTCTTATGGAGGAGAAGAGGTGAACAGAGCATTTCCTGGCTAC GAGCACTGATTGAATACAGCTGTGAGAGGCAACACCAGATGATTTTTGAAGCCATTCG AGGAGTATCAATAAGAAGTGATATTGCCATTGATGATGTTAAATTTCAGGCAGGACCC TGTGGAGAAATGGAAGATACAACTCAACAATCATCAGGATATTCTGAGGACTTAAATG AAATTGAGTATCTCGAGGAGTCAAAGCCGCAAGTGCTCTCTTCAAAGGTCC
	ORF Start: at 1	ORF Stop: at 2023
	SEQ ID NO: 68	674 aa	MW at 76492.5kD
NOV18b, 174308357 Protein Sequence	GSFEESTCGFDSVLASLPWILNEEGHYIYVDTSFGKQGEKAVLLSPDLQAEEWSCLRL VYQITTSSESLSDPSQLNLYMRFEDESFDRLLWSAKEPSDSWLIASLDLQNSSKKFKI LIEGVLGQGNTASIALFEIKMTTGYCIECDFEENHLCGFVNRWNPNVNWFVGGGSIRN VHSILPQDHTFKSELGHYMYVDSVYVKHFQEVAQLISPLTTAPMAGCPSFYYQIQQGN DNVFSLYTRDVAGLYEEIWKADRPGNAAWNLAEVEFNAPYPMEVIFEVAFNGPKGGYV ALDDISFSPVHCQNQTELLFSAVEASCNFEQDLCNFYQDKEGPGWTRVKVKPNMYRAG DHTTGLGYYLLANTKFTSQPGYIGRLYGPSLPGNLQYCLRFHYAIYGFLKMSDTLAVY IFEENHVVQEKIWSVLESPRGVWMQAEITFKKPMPTKVVFMSLCKSFWDCGLVALDDI TIQLGSCSSSEKLPPPPGECTFEQDECTFTQEKRNRSSWHRRRGETPTSYTGPKGDHT TGVGYYMYIEASHMVYGQKARLLSRPLRGVSGKHCLTFFYHMYGGGTGLLSVYLKKEE DSEESLLWRRRGEQSISWLRALIEYSCERQHQMIFEAIRGVSIRSDIAIDDVKFQAGP CGEMEDTTQQSSGYSEDLNEIEYLEESKPQVLSSKV

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

Table 18B. Comparison of NOV18a against NOV18b.

Protein Sequence	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV18b	29..680 2..661	629/660 (95%) 636/660 (96%)

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

Table 18C. Protein Sequence Properties NOV18a	
PSort analysis:	0.7480 probability located in microbody (peroxisome); 0.6736 probability located in nucleus; 0.6415 probability located in mitochondrial matrix space; 0.3377 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 30 and 31

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

Table 18D. Geneseq Results for NOV18a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB53298	Human polypeptide #38 - Homo sapiens, 686 aa. [WO200181363-A1, 01-NOV-2001]	17..680 15..686	659/672 (98%) 661/672 (98%)	0.0
AAB01432	Human TANGO 239 (form 2) - Homo sapiens, 686 aa. [WO200039284-A1, 06-JUL-2000]	17..680 15..686	655/672 (97%) 660/672 (97%)	0.0
ABB53297	Human polypeptide #37 - Homo sapiens, 640 aa. [WO200181363-A1, 01-NOV-2001]	52..680 4..640	624/637 (97%) 626/637 (97%)	0.0
AAB01426	Human TANGO 239 - Homo sapiens, 549 aa. [WO200039284-A1, 06-JUL-2000]	17..506 15..504	482/490 (98%) 487/490 (99%)	0.0
AAB00036	Human TANGO 239 partial sequence - Homo sapiens, 465 aa. [WO200039284-A1, 06-JUL-2000]	26..500 1..465	456/475 (96%) 461/475 (97%)	0.0

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

Table 18E. Public BLASTP Results for NOV18a				
Protein Accession Number	**Protein/Organism/Length**	**NOV18a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q91641	Thyroid hormone-induced protein B precursor - Xenopus laevis (African clawed frog), 688 aa.	1..680 1..688	438/689 (63%) 547/689 (78%)	0.0
Q96BM4	HYPOTHETICAL 26.4 KDA PROTEIN - Homo sapiens (Human), 232 aa.	457..680 1..232	222/232 (95%) 223/232 (95%)	e-129
CAD13324	BA373A9.2 (NOVEL PROTEIN (ORTHOLOG OF X.LAEVIS THYROID HORMONE-INDUCED PROTEIN B)) - Homo sapiens (Human), 135 aa (fragment).	554..680 1..135	127/135 (94%) 127/135 (94%)	4e-67
Q63191	Apical endosomal glycoprotein precursor - Rattus norvegicus (Rat), 1216 aa.	272..670 587..975	108/412 (26%) 182/412 (43%)	5e-29
Q9GMT4	HYPOTHETICAL 51.2 KDA PROTEIN - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 448 aa.	511..666 240..411	66/172 (38%) 87/172 (50%)	1e-20

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

Table 18F. Domain Analysis of NOV18a			
Pfam Domain	**NOV18a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
MAM: domain 1 of 4	28..171	54/176 (31%) 123/176 (70%)	1.3e-34
MAM: domain 2 of 4	172..331	49/175 (28%) 115/175 (66%)	3.5e-35
TonB_boxC: domain 1 of 1	448..463	2/16 (12%) 14/16 (88%)	5
MAM: domain 3 of 4	344..500	59/174 (34%) 120/174 (69%)	7.3e-43

5

pili_assembly_C: domain 1 of 1	607..623	5/17 (29%) 15/17 (88%)	6.8
MAM: domain 4 of 4	511..668	68/174 (39%) 129/174 (74%)	6.8e-59

Example 19.

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: 69	3815 bp
NOV19a, CG93088-01 DNA Sequence	CGGCCGCGATCCCCACCACACCACCAGCCCGGCCGCACGGGGCACTGAGCCGGGTGCT GAGCACCGGAGGCCCCGCCGAGGCCGGGACTCAGATGTTGAAAGTTAATTTGTGTAAA GACTTATGCACGTGGTGACATGAGTTCTGCCCAGTGCTCTGAAATCAAAGTGAAGAAA TAAATCCATGGAAGCCCAGGCAAATGATGGGTGTAGCTATGACTCTCTGAAGGACCTG CAGAGAAACGCCTCCTGATTTTGTCTTACA**ATG**GAACTTAAAAAGTCGCCTGACGGTG GATGGGGCTGGGTGATTGTGTTTGTCTCCTTCCTTACTCAGTTTTTGTGTTACGGATC CCCACTAGCTGTTGGAGTCCTGTACATAGAATGGCTGGATGCCTTTGGTGAAGGAAAA GGAAAAACAGCCTGGGTTGGATCCCTGGCAAGTGGAGTTGGCTTGCTTGCAAGTCCTG TCTGCAGTCTCTGTGTCTCATCTTTTGGAGCAAGACCTGTCACAATCTTCAGTGGCTT CATGGTGGCTGGAGGCCTGATGTTGAGCAGTTTTGCTCCCAATATCTACTTTCTGTTT TTTTCCTATGGCATTGTTGTAGGTCTTGGATGTGGTTTATTATACACTGCAACAGTGA CCATTACGTGCCAGTATTTTGACGATCGCCGAGGCCTAGCGCTTGGCCTGATTTCAAC AGGTTCAAGCGTTGGCCTTTTCATATATGCTGCTCTGCAGAGGATGCTGGTTGAGTTC TATGGACTGGATGGATGCTTGCTGATTGTGGGTGCTTTAGCTTTAAATATATTAGCCT GTGGCAGTCTGATGAGACCCCTCCAATCTTCTGATTGTCCTTTGCCTAAAAAAATAGC TCCAGAAGATCTACCAGATAAATACTCCATTTACAATGAAAAAGGAAAGAATCTGGAA GAAAACATAAACATTCTTGACAAGAGCTACAGTAGTGAGGAAAAATGCAGGATCACGT TAGCCAATGGTGACTGGAAACAAGACAGCCTACTTCATAAAAACCCCACAGTGACACA CACAAAAGAGCCTGAAACGTACAAAAAGAAAGTTGCAGAACAGACATATTTTTGCAAA CAGCTTGCCAAGAGGAAGTGGCAGTTATATAAAAACTACTGTGGTGAAACTGTGGCTC TTTTTAAAAACAAAGTATTTTCAGCCCTTTTCATTGCTATCTTACTCTTTGACATCGG AGGGTTTCCACCTTCATTACTTATGGAAGATGTAGCAAGAAGTTCAAACGTGAAAGAA GAAGAGTTTATTATGCCACTTATTTCCATTATAGGCATTATGACAGCAGTTGGTAAAC TGCTTTTAGGGATACTGGCTGACTTCAAGTGGATTAATACCTTGTATCTTTATGTTGC TACCTTAATCATCATGGGCCTAGCCTTGTGTGCAATTCCATTTGCCAAAAGCTATGTC ACATTGGCGTTGCTTTCTGGGATCCTAGGGTTTCTTACTGGTAATTGGTCCATCTTTC CATATGTGACCACGAAGACTGTGGGAATTGAAAAATTAGCCCATGCCTATGGGATATT AATGTTCTTTGCTGGACTTGGAAATAGCCTAGGACCACCATCGTTGGGTTGGTTTTAT GACTGGACCCAGACCTATGATATTGCATTTTATTTTAGTGGCTTCTGCGTCCTGCTGG GAGGTTTTATTCTGCTGCTGGCAGCCTTGCCCTCTTGGGATACATGCAACAAGCAACT CCCCAAGCCAGCTCCAACAACTTTCTTGTACAAAGTTGCCTCTAATGTT**TAG**AAGAAT ATTGGAAGACACTATTTTTGCTATTTTATACCATATAGCAACGATATTTTAACAGATT CTCAAGCAAATTTTCTAGAGTCAAGACTATTTTCTCATAGCAAAATTTCACAATGACT GACTCTGAATGAATTATTTTTTTTTATATATCCTATTTTTTATGTAGTGTATGCGTAG CCTCTATCTCGTATTTTTTTCTATTTCTCCTCCCCACACCATCAATGGGACTATTCTG TTTTGCTGTTATTCACTAGTTCTTAACATTGTAAAAAGTTTGACCAGCCTCAGAAGGC TTTCTCTGTGTAAAGAAGTATAATTTCTCTGCTGACTCCATTTAATCCACTGCAAGGC ACCTAGAGAGACTGCTCCTATTTTAAAAGTGATGCAAGCATCATGATAAGATATGTGT GAAGCCCACTAGGAAATAAATCATTCTCTTCTCTATGTTTGACTTGCTAGTAAACAGA AGACTTCAAGCCAGCCAGGAAATTAAAGTGGCGACTAAAACAGCCTTAAGAATTGCAG TGGAGCAAATTGGTCATTTTTTAAAAAAATATATTTTAACCTACAGTCACCAGTTTTC

	ATTATTCTATTTACCTCACTGAAGTACTCGCATGTTGTTTGGTACCCACTGAGCAACT GTTTCAGTTCCTAAGGTATTTGCTGAGATGTGGGTGAACTCCAAATGGAGAAGTAGTC ACTGTAGACTTTCTTCATGGTTGACCACTCCAACCTTGCTCACTTTTGCTTCTTGGCC ATCCACTCAGCTGATGTTTCCTGGAAGTGCTAATTTTACCTGTTTCCAAATTGGAAAC ACATTTCTCAATCATTCCGTTCTGGCAAATGGGAAACATCCATTTGCTTTGGGCACAG TGGGGATGGGCTGCAAGTTCTTGCATATCCTCCCAGTGAAGCATTTATTTGCTACTAT CAGATTTTACCACTATCAAATATAATTCAAGGGCAGAATTAAACGTGAGTGTGTGTGT GTGTGTGTGTGTGTGTGCTATGCATGCTCTAAGTCTGCATGGGATATGGGAATGGAAA AGGGCAATAAGAAATTAATACCCTTATGCAGTTGCATTTAACCTTAAGAAAAATGTCC TTGGGATAAACTCCAATGTTTAATACATTGATTTTTTTTCTAAAGAAATGGGTTTTAA ACTTTGGTATGCATCAGAATTCCCTATAGATCTTTTTGAAAATATAGGTACCTGGGTA TCACACATAGAACTTTTAATTCTGCTGGTGTAGGCTGTTGCCCAAACATCTATAATTT TACTGAGCTCTTCAAGTGATTCTGATAACACAGCCTGGATTGAGAATTTTTATAAGAT TGGCAATGGAAAAACATTTATTCTTTTAAATAATAATTTTTTTAAAACCCAAGAGGTC AGGGGATTTTATAAACCAATAGCCAAGTGTTCTTTAAATAGGAGGCACCCTTCCCATT GTGCCAAAATCATCTTTTCATTTATTTTGAAATTTGTATGATTATTTTATACTTGTAT GTTGCCTTTCTTCGAAGGCGCCTGAAGCACTTTATAAACACAAATCCTCACAATACCT CTGTGAGGTAGGTAAATAGTACTTTTCTATGTAGTAAACCTGGAATATGGAGAATTTC ATAACAGTTCATTCTACTTAATAATGCAATAATGGAGCTCCAAGTTGTCTTGGACTTC TACACCACACTCAGACTTCTGGAAAGTTTTCTGTACCTCATTCTTTAGTCCCTGTCAA GGTTAGTAAATAAAATAAGTGACATAAAAAAAAAAAAAAAACTAAACTACTTGTTGTG TTGAAAGTTCCTTTTTGCCAGTTATGTTCAGGAAACCCAATAACCTGAAAAAGTTTGA CTTTGATGTGACATCTTCATATTCATCAATGCTGATAATTGTCCAAAGGCATCTTCAC TATGTCTGCTAAATAACATCCAATGTGGGCGTTATCTGTTGTCTAGGGGATGAATTTT AAGTTACAATAAAATATTTTTCTTTGTTTTGCATCAAAAAAAAAA		
	ORF Start: ATG at 263	ORF Stop: TAG at 1790	
	SEQ ID NO: 70	509 aa	MW at 55780.8kD
NOV19a, CG93088-01 Protein Sequence	MELKKSPDGGWGWVIVFVSFLTQFLCYGSPLAVGVLYIEWLDAFGEGKGKTAWVGSLA SGVGLLASPVCSLCVSSFGARPVTIFSGFMVAGGLMLSSFAPNIYFLFFSYGIVVGLG CGLLYTATVTITCQYFDDRRGLALGLISTGSSVGLFIYAALQRMLVEFYGLDGCLLIV GALALNILACGSLMRPLQSSDCPLPKKIAPEDLPDKYSIYNEKGKNLEENINILDKSY SSEEKCRITLANGDWKQDSLLHKNPTVTHTKEPETYKKKVAEQTYFCKQLAKRKWQLY KNYCGETVALFKNKVFSALFIAILLFDIGGFPPSLLMEDVARSSNVKEEEFIMPLISI IGIMTAVGKLLLGILADFKWINTLYLYVATLIIMGLALCAIPFAKSYVTLALLSGILG FLTGNWSIFPYVTTKTVGIEKLAHAYGILMFFAGLGNSLGPPSLGWFYDWTQTYDIAF YFSGFCVLLGGFILLLAALPSWDTCNKQLPKPAPTTFLYKVASNV		

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

Table 19B. Protein Sequence Properties NOV19a

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 29 and 30

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.

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**
AAY31642	Human transport-associated protein-4 (TRANP-4) - Homo sapiens, 465 aa. [WO9941373-A2, 19-AUG-1999]	5..218 14..227	75/214 (35%) 116/214 (54%)	1e-33
AAM93737	Human polypeptide, SEQ ID NO: 3705 - Homo sapiens, 471 aa. [EP1130094-A2, 05-SEP-2001]	7..201 33..228	67/196 (34%) 109/196 (55%)	5e-30
AAB88570	Human hydrophobic domain containing protein clone HP03612 #34 - Homo sapiens, 375 aa. [WO200112660-A2, 22-FEB-2001]	7..201 9..204	67/196 (34%) 108/196 (54%)	6e-30
AAE06594	Human protein having hydrophobic domain, HP03949 - Homo sapiens, 390 aa. [WO200149728-A2, 12-JUL-2001]	67..451 13..384	95/403 (23%) 175/403 (42%)	3e-25
AAO07132	Human polypeptide SEQ ID NO 21024 - Homo sapiens, 107 aa. [WO200164835-A2, 07-SEP-2001]	398..480 5..87	38/83 (45%) 51/83 (60%)	1e-14

In a BLAST search of public sequence datbases, 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**
Q9UFH8	HYPOTHETICAL 17.1 KDA PROTEIN - Homo sapiens (Human), 157 aa (fragment).	353..509 1..157	155/157 (98%) 156/157 (98%)	2e-87
Q9CPZ7	4930425B13RIK PROTEIN (1200003C15RIK PROTEIN) - Mus musculus (Mouse), 159 aa.	352..509 1..159	148/159 (93%) 152/159 (95%)	2e-82
O15374	Monocarboxylate transporter 5 (MCT 5) (MCT 4) - Homo sapiens (Human), 487 aa.	5..473 13..468	128/487 (26%) 222/487 (45%)	7e-47

O15403	Monocarboxylate transporter 7 (MCT 7) (MCT 6) - Homo sapiens (Human), 523 aa.	7..491 19..481	124/490 (25%) 223/490 (45%)	3e-40
Q9W509	MCT1 PROTEIN - Drosophila melanogaster (Fruit fly), 626 aa.	7..230 29..255	85/227 (37%) 124/227 (54%)	1e-38

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

Table 19E. Domain Analysis of NOV19a

Pfam Domain	NOV19a Match Region	Identities/ Similarities for the Matched Region	Expect Value
sugar_tr: domain 1 of 1	11..456	74/547 (14%) 276/547 (50%)	0.27

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: 71	724 bp	
NOV20a, CG93335-01 DNA Sequence	CAGGAGGCGGGTGGGTCAAGGTAACTCTGGGCTACAGAGTCCTTGCTGGGGGTTCGGG GAGCGCTTGGACCCCGGCTTCTGGGACGCGTCAGAATATTATCCAGCAATGCAAATGA ACAAACTATAACTACACACAGCTGC**ATG**GATAAATGTCAGAAACATGACGTTGAGTGT GAGAAGCCAGATGCAAACGAGGACTCACTGTGCAATTCTGTGCATGTACAGTGGCCAG GAGAAGGAGCACTGGCTTTGCTTTCATCAGGCCAAAGATGCCTTTCTTTGGGAATAC GTTCAGTCCGAAGAAGACACCTCCTCGGAAGTCGGCATCTCTCTCCAACCTGCATTCT TTGGATCGATCAACCCGGGAGGTGGAGCTGGGCTTGGAATACGGATCCCCGACTATGA ACCTGGCAGGGCAAAGCCTGAAGTTTGAAAATGGCCAGTGGATAGCAGAGACAGGGGT TAGTGGCGGTGTGGACCGGAGGGAGGTTCAGCGCCTTCGCAGGCGGAACCAGCAGTTG GAGGAAGAGAACAATCTCTTGCGGCTGAAAGTGGACATCTTATTAGACATGCTTTCAG AGTCCACTGCTGAATCCCACTTAATGGAGAAGGAACTGGATGAACTGAGGATCAGCCG GAAGAGAAAA**TGA**AGACCCCAGAGACATTTATTGGGGAGTAGGATGTGGCTGAGTGCT TTTTTTTTGGCCAGACTAGCGGATTCAG		
	ORF Start: ATG at 142	ORF Stop: TGA at 649	
	SEQ ID NO: 72	169 aa	MW at 19286.6kD
NOV20a, CG93335-01 Protein Sequence	MDKCQKHDVECEKPDANEDSLCNSVHVQWPGEGSTGFAFIRPKMPFFGNTFSPKKTPP RKSASLSNLHSLDRSTREVELGLEYGSPTMNLAGQSLKFENGQWIAETGVSGGVDRRE VQRLRRRNQQLEEENNLLRLKVDILLDMLSESTAESHLMEKELDELRISRKRK		

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

Table 20B. Protein Sequence Properties NOV20a	
PSort analysis:	0.4600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence

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 20C.

5

Table 20C. Geneseq Results for NOV20a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV20a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAM00955	Human bone marrow protein, SEQ ID NO: 431 - Homo sapiens, 175 aa. [WO200153453-A2, 26-JUL-2001]	31..169 37..175	139/139 (100%) 139/139 (100%)	1e-74
AAY86201	Nuclear transport protein clone hfb2025 protein sequence - Homo sapiens, 67 aa. [WO9964455-A1, 16-DEC-1999]	103..169 1..67	67/67 (100%) 67/67 (100%)	6e-30
ABB23535	Protein #5534 encoded by probe for measuring heart cell gene expression - Homo sapiens, 26 aa. [WO200157274-A2, 09-AUG-2001]	44..69 1..26	26/26 (100%) 26/26 (100%)	2e-08
AAB69070	Human male enhanced antigen-2 (MEA-2) protein sequence SEQ ID NO:2 - Homo sapiens, 1374 aa. [JP2000316580-A, 21-NOV-2000]	62..163 768..868	25/102 (24%) 45/102 (43%)	1.1
AAU36216	Pseudomonas aeruginosa cellular proliferation protein #206 - Pseudomonas aeruginosa, 874 aa. [WO200170955-A2, 27-SEP-2001]	104..163 683..749	22/67 (32%) 35/67 (51%)	1.9

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

10

Table 20D. Public BLASTP Results for NOV20a				
Protein Accession Number	**Protein/Organism/Length**	**NOV20a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9Y3M2	HYPOTHETICAL 14.5 KDA PROTEIN (CHROMOSOME 22 OPEN READING FRAME 2) - Homo sapiens (Human), 126 aa.	44..169 1..126	126/126 (100%) 126/126 (100%)	4e-66
AAL56062	CYTOSOLIC LEUCINE-RICH PROTEIN - Homo sapiens (Human), 126 aa.	44..169 1..126	125/126 (99%) 126/126 (99%)	1e-65
Q9D1C2	1110014P06RIK PROTEIN (RIKEN CDNA 1110014P06 GENE) - Mus musculus (Mouse), 127 aa.	44..169 1..126	104/126 (82%) 120/126 (94%)	1e-56
Q9UIK9	HRIHFB2025 PROTEIN - Homo sapiens (Human), 67 aa (fragment).	103..169 1..67	67/67 (100%) 67/67 (100%)	1e-29
Q9CVN6	1700121K02RIK PROTEIN - Mus musculus (Mouse), 226 aa (fragment).	47..160 70..191	45/122 (36%) 69/122 (55%)	2e-15

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

Table 20E. Domain Analysis of NOV20a			
Pfam Domain	**NOV20a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Transposase_8: domain 1 of 1	54..149	22/99 (22%) 64/99 (65%)	2.9

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: 73	1310 bp
NOV21a, CG93345-01 DNA Sequence	GATTCCAGTTGAAGTCAGTTTGACTTA**ATG**AGCTCTTCTCTATTTTCCTACTCAAACC TCTATTCCACCATGTCACCACTCAACCAAACTACTGAGAACCACCAGAGCTTCTTCAC CCTGACTGGGATTCCAGGAATGCCAGAGAAAGACTTATGGATGGCCTTGCCCCTCTGT CTTCTTTATAGCACCACGATCTTGGGAAATGTCACCATCCTTGTTGTCATCAAAGTTG	

	AGCAAAGTCTCCATGAGCCCATGTATTTTTTTCTAGCCATGTTAGCTGCCACTGACCT CAGCCTTTCACTGTCTTCCATGCCTACCATGGTCAGTGTTCACTGGTTCAACTGGCGT TCAATAACTTTTAATGGCTGCCTTATCCAGATGTTCTTCATCCACACATTTGGGGGAG TGGAATCAGGTGTTCTGGTGGCCATGGCCTTTGATCGCTTTGTGGCCATCCGCTTTCC TTTGCACTATGCTACAATTCTCACTCACAGTGTCATCAGCAAGATTGCAGCAGCCATC CTGCTACGGAGTGTGGGGGCTGTGCTCCCTGTGCCTTTTCTCATCAAAAGGTTACCTT TCTGTCACTCCAATGTCCTCTCCCATGCATACTGCCTCCATCAGGATGCCATGAGGCT TGCCTGTGCTGACACTGGTGTCAATAGCATCTATGGCCTGTTGGCTGTGATCTTCATC ATTGTACTAGATGCCTTAATACTTTTGGCCTCTTACATTCTAATCCTGCAGGCAGTAT TGAGCATTGCTTCCCAGGAAGACAGGCTCAAGGCTCTCAACACCTGTGTCTCTCTCAT ATCTGCAGTGCTGCTTTTCTATGTGCCTCTCATTGGTATGACCCTAATTCATCGCTAT GGGAAGCATTTGTCACCACTAATACACACATTCATGGCCAATATCTACCTGCTTCTCC CTCCTGTGCTCAATCCCATTGTGTACAGTGTTAGGACCAAGCAGATC**TGA**TAGCAGAT TGTCCAGGCCTTTTGTGGGGCTAGGGTTAGCCCTTAATGGCATCTACTATTTCCAAGT AAATGCAATCAAGTTAGAGAAGAGTATCAAATACAGCACTATCCAATAGAAATTCCCA CAGAAGTGGATATTTTCTATTTCTCTGCTGTTTAGTAACTAGTAGCTGTACATGGCTA TTAATTGCTTGAAATTTTGCTAGTGCAAGCTGAGGAACTGAATTTTAAATGTACTTAA TTTTAATTGATTTAAATGTAAATTTAAGTAGTCATATGTAACTAGTAGCTGCCGTATC AAATAGTACAAATACAATGGGTAGTGATATGAAA		
	ORF Start: ATG at 28	ORF Stop: TGA at 976	
	SEQ ID NO: 74	316 aa	MW at 35115.4kD
NOV21a, CG93345-01 Protein Sequence	MSSSLFSYSNLYSTMSPLNQTTENHQSFFTLTGIPGMPEKDLWMALPLCLLYSTTILG NVTILVVIKVEQSLHEPMYFFLAMLAATDLSLSLSSMPTMVSVHWFNWRSITFNGCLI QMFFIHTFGGVESGVLVAMAFDRFVAIRFPLHYATILTHSVISKIAAAILLRSVGAVL PVPFLIKRLPFCHSNVLSHAYCLHQDAMRLACADTGVNSIYGLLAVIFIIVLDALILL ASYILILQAVLSIASQEDRLKALNTCVSLISAVLLFYVPLIGMTLIHRYGKHLSPLIH TFMANIYLLLPPVLNPIVYSVRTKQI		

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

Table 21B. Protein Sequence Properties NOV21a

PSort analysis:	0.6000 probability located in plasma membrane; 0.4905 probability located in mitochondrial inner membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 59 and 60

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**
AAG71700	Human olfactory receptor polypeptide, SEQ ID NO: 1381 - Homo sapiens, 323 aa. [WO200127158-A2, 19-APR-2001]	15..316 1..302	298/302 (98%) 301/302 (98%)	e-169
AAG71602	Human olfactory receptor polypeptide, SEQ ID NO: 1283 - Homo sapiens, 302 aa. [WO200127158-A2, 19-APR-2001]	15..316 1..302	193/303 (63%) 244/303 (79%)	e-110
AAU24684	Human olfactory receptor AOLFR183 - Homo sapiens, 302 aa. [WO200168805-A2, 20-SEP-2001]	15..316 1..302	193/303 (63%) 244/303 (79%)	e-110
AAG71516	Human olfactory receptor polypeptide, SEQ ID NO: 1197 - Homo sapiens, 315 aa. [WO200127158-A2, 19-APR-2001]	26..316 11..301	170/291 (58%) 220/291 (75%)	9e-99
AAU24569	Human olfactory receptor AOLFR59 - Homo sapiens, 315 aa. [WO200168805-A2, 20-SEP-2001]	26..316 11..301	170/291 (58%) 220/291 (75%)	9e-99

In a BLAST search of public sequence datbases, 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**
AAL60646	OLFACTORY RECEPTOR MOR10-1 - Mus musculus (Mouse), 315 aa.	19..316 4..300	210/298 (70%) 248/298 (82%)	e-119
AAL60660	OLFACTORY RECEPTOR MOR10-2 - Mus musculus (Mouse), 318 aa.	22..316 7..301	181/295 (61%) 228/295 (76%)	e-104

AAL60631	OLFACTORY RECEPTOR MOR5-2 - Mus musculus (Mouse), 321 aa.	13..316 5..307	172/304 (56%) 232/304 (75%)	2e-98
AAL60629	OLFACTORY RECEPTOR MOR5-1 - Mus musculus (Mouse), 321 aa.	13..316 5..307	173/304 (56%) 230/304 (74%)	2e-98
AAL60640	OLFACTORY RECEPTOR MOR7-2 - Mus musculus (Mouse), 312 aa.	27..316 13..302	170/290 (58%) 221/290 (75%)	1e-97

PFam analysis indicates that the NOV21a protein contains the domains shown in Table 21E.

Table 21E. Domain Analysis of NOV21a

Pfam Domain	NOV21a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1: domain 1 of 1	58..309	50/270 (19%) 169/270 (63%)	1.2e-22

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: 75	999 bp	
NOV22a, CG93400-01 DNA Sequence	ACGAGTGGATAAGCTAGTGACCTATCTGTG**ATG**TTTCTGCTCAATACCTCAGAAGTTG AAGTCTCCACATTCCTATTGATTGGGATACCAGGACTTGAGCATGCACACATTTGGAT CTCTATCCCCATCTGCCTTATGTACCTCATGGCCATCCTGGGCAACTGCACCATCCTA TTTGTTATCAGAACAGAGCATTCCCTGCAAGAGCCCATGTACTATTTCCTCTCCATGC TGGCCCTGTCCGACCTGGGCCTGTCTTTCTCCTCCCTACCCACGATGCTGAGAATCTT CTTGTTCAACAACATGGGGATTTCTGCTGATACATGCATTGCCCAGGAATTCTTCATC CATGGATTCACAGACATGGAGTCTTCAGTTCTCCTAATCATGTCCTTTGATCACTTAG TAGCCATTTGCAACCCCCTAAGATATAGCTCTATTCTCACCAGCTTCAGGGTTTTGCA AATTGGACTGGCTTTTGCCATTAAAAGCATTCTCCTAGTGCTACCCCTTCCTTTTACT TTAAAGAGACTCAGATACTGTAATAAACACCTTTTATCCCACTCCTACTGCCTTCACC AGGATGTAATGAAGCTGGCCTGCTCTGACAACAGGGTTAACTTTTACTATGGTTTGTT CGTTGCACTCTGCATGATGTCAGACAGTTTTTATTGCTATTTCCTATATGTGTTCATC CTGAAGACTGTGTTGGGTATTGCATCCCATGGGGAGTGCCTCGAAGCTCTTGACACCT GTGTGTCTCATATCTGTGCTGTACTCGTCTTCTATGTGCCCATCATCACCTTGGCTAC CATGCGTCGCTTTGCTAAGCATAAATCCCCTTTAGCTATGATTCTGATAGCAGATGCA TTCTTGCTGGTACCACCCTTGATGAATCCCATTGTGTATTGTGTAAAAACTCGGCAGA TTAGAGTAAAGGTCCTGGAAAAATTGGCTCTGAAGCCTAAA**TGA**TGGGGCAAAGGTGG AAATTCTATTTTT		
	ORF Start: ATG at 31	ORF Stop: TGA at 970	
	SEQ ID NO: 76	313 aa	MW at 35541.3kD

NOV22a, CG93400-01 Protein Sequence	MFLLNTSEVEVSTFLLIGIPGLEHAHIWISIPICLMYLMAILGNCTILFVIRTEHSLQ EPMYYFLSMLALSDLGLSFSSLPTMLRIFLFNNMGISADTCIAQEFFIHGFTDMESSV LLIMSFDHLVAICNPLRYSSILTSFRVLQIGLAFAIKSILLVLPLPFTLKRLRYCNKH LLSHSYCLHQDVMKLACSDNRVNFYYGLFVALCMMSDSFYCYFLYVFILKTVLGIASH GECLEALDTCVSHICAVLVFYVPIITLATMRRFAKHKSPLAMILIADAFLLVPPLMNP IVYCVKTRQIRVKVLEKLALKPK

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

Table 22B. Protein Sequence Properties NOV22a

PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.2414 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 44 and 45

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
AAG71721	Human olfactory receptor polypeptide, SEQ ID NO: 1402 - Homo sapiens, 316 aa. [WO200127158-A2, 19-APR-2001]	1..313 1..313	305/314 (97%) 307/314 (97%)	e-172
AAG71564	Human olfactory receptor polypeptide, SEQ ID NO: 1245 - Homo sapiens, 322 aa. [WO200127158-A2, 19-APR-2001]	1..311 5..315	231/311 (74%) 264/311 (84%)	e-131
AAG71701	Human olfactory receptor polypeptide, SEQ ID NO: 1382 - Homo sapiens, 312 aa. [WO200127158-A2, 19-APR-2001]	1..308 1..306	230/308 (74%) 257/308 (82%)	e-129
AAU24682	Human olfactory receptor AOLFR181 - Homo sapiens, 312 aa. [WO200168805-A2, 20-SEP-2001]	1..308 1..306	230/308 (74%) 257/308 (82%)	e-129

AAG72486	Human OR-like polypeptide query sequence, SEQ ID NO: 2167 - Homo sapiens, 345 aa. [WO200127158-A2, 19-APR-2001]	1..313 1..338	234/338 (69%) 265/338 (78%)	e-127

In a BLAST search of public sequence datbases, 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
AAL60639	OLFACTORY RECEPTOR MOR8-3 - Mus musculus (Mouse), 312 aa.	3..313 2..312	273/311 (87%) 286/311 (91%)	e-154
AAL60635	OLFACTORY RECEPTOR MOR8-1 - Mus musculus (Mouse), 318 aa.	1..311 8..318	248/311 (79%) 271/311 (86%)	e-141
AAL60638	OLFACTORY RECEPTOR MOR8-2 - Mus musculus (Mouse), 317 aa.	1..311 5..315	214/311 (68%) 252/311 (80%)	e-121
AAL60640	OLFACTORY RECEPTOR MOR7-2 - Mus musculus (Mouse), 312 aa.	12..304 13..306	175/294 (59%) 215/294 (72%)	1e-94
AAL60634	OLFACTORY RECEPTOR MOR7-1 - Mus musculus (Mouse), 313 aa.	1..307 1..308	171/308 (55%) 221/308 (71%)	4e-94

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

Table 22E. Domain Analysis of NOV22a

Pfam Domain	NOV22a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1: domain 1 of 2	43..81	12/39 (31%) 30/39 (77%)	6e-09
7tm_1: domain 2 of 2	217..293	13/88 (15%) 51/88 (58%)	1.3

Example 23.

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: 77	2715 bp
NOV23a, CG93410-01 DNA Sequence	GATGGAGCACGGCACACTCCTCGCCCAGCCCGGGCTCTGGACCAGGGACACCAGCTGG GCACTCCTCTATTTCCTCTGCTATATCCTCCCTCAGACCGCCCCGCAAGTACTCAGGA TCGGAGGGATTTTTGAAACAGTGGAAAATGAGCCTGTTAATGTTGAAGAATTAGCTTT CAAGTTTGCAGTCACCAGCATTAACAGAAACCGAACCCTGATGCCTAACACCACATTA ACCTATGACATCCAGAGAATTAACCTTTTTGATAGTTTTGAGGCCTCGCGGAGAGCAT GTGACCAGCTGGCTCTTGGTGTGGCTGCTCTCTTTGGCCCTTCCCATAGCTCCTCCGT CAGTGCTGTGCAGTCTATTTGCAATGCTCTCGAAGTTCCACACATACAGACCCGCTGG AAACACCCCTCGGTGGACAACAAAGATTTGTTTTACATCAACCTTTACCCAGATTATG CAGCTATCAGCAGGGCGATCCTGGATCTGGTCCTCTATTACAACTGGAAAACAGTGGC AGTGGTGTATGAAGACAGCACAGGTCTAATTCGTCTACAAGAGCTCATCAAAGCTCCC TCCAGATATAATATTAAAATCAAAATCCGCCAGCTGCCCTCTGGGAATAAAGATGCCA AGCCTTTACTCAAGGAGATGAAGAAAGGCAAGGAGTTCTATGTGATATTTGATTGTTC ACATGAAACAGCCGCTGAAATCCTTAAGCAGATTCTGTTCATGGGCATGATGACCGAG TACTATCACTACTTTTTCACAACCCTGGACTTATTTGCTTTGGATCTGGAACTCTATA GGTACAGTGGCGTAAACATGACCGGGTTTCGGCTGCTTAACATTGACAACCCTCACGT GTCATCCATCATTGAGAAGTGGTCCATGGAGAGACTGCAGGCCCCACCCAGGCCCGAG ACTGGCCTTTTGGATGGCATGATGACAACTGAAGCGGCTCTGATGTACGATGCTGTGT ACATGGTGGCCATTGCCTCGCACCGGGCATCCCAGCTGACCGTCAGCTCCCTGCAGTG CCATAGACATAAGCCATGGCGCCTCGGACCCAGATTTATGAACCTGATCAAAGAGGCC CGGTGGGATGGCTTGACTGGGCATATCACCTTTAATAAAACCAATGGCTTGAGGAAGG ATTTTGATCTGGATATTATTAGTCTCAAAGAGGAAGGAACTGAAAAGATTGGGATTTG GAATTCCAACAGTGGGCTTAACATGACGGACAGCAACAAAGACAAGTCCAGCAATATC ACTGATTCATTGGCCAACAGAACACTCATTGTCACCACCATTCTGGAAGAACCCTATG TTATGTACAGGAAATCTGATAAGCCTCTATATGGAAATGACAGATTTGAAGGATATTG CCTAGACCTGTTGAAAGAATTGTCAAACATCCTGGGTTTCATTTATGATGTTAAACTA GTTCCCGATGGCAAATATGGGGCCCAGAATGACAAAGGGGAGTGGAACGGGATGGTTA AAGAACTCATAGATCACAGGGCTGACCTGGCAGTGGCTCCTCTTACCATCACCTACGT GCGGGAGAAAGTCATTGACTTCTCCAAACCCTTCATGACCCTAGGCATCAGCATTCTC TACCGGAAGCCCAATGGTACCAATCCAGGCGTTTTCTCCTTCCTCAACCCCCTGTCTC CAGATATTTGGATGTATGTGCTCTTAGCCTGCTTGGGAGTCAGCTGTGTACTCTTTGT GATTGCAAGGTTTACACCCTACGAGTGGTATAACCCCCACCCATGCAACCCTGACTCA GACGTGGTGGAAAACAATTTTACTTTACTAAATAGTTTCTGGTTTGGAGTTGGAGCTT TCATGCAGCAAGGATCAGAGCTGATGCCCAAAGCTCTATCGACCAGAATAGTTGGAGG GATATGGTGGTTTTTCACCCTAATCATCATTTCATCCTACACGGCCAATCTGGCTGCC TTCTTGACAGTAGAGAGAATGGAATCCCCCATAGATTCGGCAGATGATCTGGCAAAGC AAACCAAGATAGAATATGGGGCGGTTAGAGATGGATCAACAATGACCTTCTTCAAGAA ATCAAAAATCTCCACCTATGAGAAGATGTGGGCTTTCATGAGCAGCAGGCAGCAGACC GCCCTGGCAAGAAACAGTGATGAGGGGATCCAGAGAGTGCTCACCACAGACTACGCGC TGCTGATGGAGTCCACCAGCATTGAGTATGTGACGCAGAGAAACTGCAACCTCACTCA GATCGGGGGCCTCATTGACTCCAAAGGTTACGGAGTGGGAACACCTATTGGTTCTCCT TACCGGGATAAAATTACTATTGCTATTCTTCAACTCCAAGAAGAAGGGAAGCTGCATA TGATGAAAGAGAAGTGGTGGCGTGGGAATGGCTGCCCCGAGGAAGACAACAAAGAAGC CAGTGCCCTGGGAGTGGAAAATATTGGAGGCATCTTCATTGTTCTGGCTGCCGGACTG GTCCTTTCTGTATTTGTAGCTATTGGAGAATTCATATACAAATCACGGAAGAATAATG ATATTGAACAGGCTTTTTGTTTCTTTTATGGACTGCAATGTAAGCAAACCCATCCAAC CAACTCCACTTCTGGAACTACTTTATCTACGGATTTAGAATGTGGTAAATTAATTCGA GAGGAGAGAGGGATTCGAAAACAGTCCTCAGTTCATACTGTGTAATC
	ORF Start: ATG at 2	ORF Stop: TAA at 2711
	SEQ ID NO: 78	903 aa	MW at 102229.3kD

NOV23a, CG93410-01 Protein Sequence	MEHGTLLAQPGLWTRDTSWALLYFLCYILPQTAPQVLRIGGIFETVENEPVNVEELAF KFAVTSINRNRTLMPNTTLTYDIQRINLFDSFEASRRACDQLALGVAALFGPSHSSSV SAVQSICNALEVPHIQTRWKHPSVDNKDLFYINLYPDYAAISRAILDLVLYYNWKTVA VVYEDSTGLIRLQELIKAPSRYNIKIKIRQLPSGNKDAKPLLKEMKKGKEFYVIFDCS HETAAEILKQILFMGMMTEYYHYFFTTLDLFALDLELYRYSGVNMTGFRLLNIDNPHV SSIIEKWSMERLQAPPRPETGLLDGMMTTEAALMYDAVYMVAIASHRASQLTVSSLQC HRHKPWRLGPRFMNLIKEARWDGLTGHITFNKTNGLRKDFDLDIISLKEEGTEKIGIW NSNSGLNMTDSNKDKSSNITDSLANRTLIVTTILEEPYVMYRKSDKPLYGNDRFEGYC LDLLKELSNILGFIYDVKLVPDGKYGAQNDKGEWNGMVKELIDHRADLAVAPLTITYV REKVIDFSKPFMTLGISILYRKPNGTNPGVFSFLNPLSPDIWMYVLLACLGVSCVLFV IARFTPYEWYNPHPCNPDSDVVENNFTLLNSFWFGVGAFMQQGSELMPKALSTRIVGG IWWFFTLIIISSYTANLAAFLTVERMESPIDSADDLAKQTKIEYGAVRDGSTMTFFKK SKISTYEKMWAFMSSRQQTALARNSDEGIQRVLTTDYALLMESTSIEYVTQRNCNLTQ IGGLIDSKGYGVGTPIGSPYRDKITIAILQLQEEGKLHMMKEKWWRGNGCPEEDNKEA SALGVENIGGIFIVLAAGLVLSVFVAIGEFIYKSRKNNDIEQAFCFFYGLQCKQTHPT NSTSGTTLSTDLECGKLIREERGIRKQSSVHTV	
	SEQ ID NO: 79	1602 bp
NOV23b, 188822752 DNA Sequence	**AGA**TCTCAAGTACTCAGGATCGGAGGGATTTTTGAAACAGTGGAAAATGAGCCTGTTA ATGTTGAAGAATTAGCTTTCAAGTTTGCAGTCACCAGCATTAACAGAAACCGAACCCT GATGCCTAACACCACATTAACCTATGACATCCAGAGAATTAACCTTTTTGATAGTTTT GAGGCCTCGCGGAGAGCATGTGACCAGCTGGCTCTTGGTGTGGCTGCTCTCTTTGGCC CTTCCCATAGCTCCTCCGTCAGTGCTGTGCAGTCTATTTGCAATGCTCTCGAAGTTCC ACACATACAGACCCGCTGGAAACACCCCTCGGTGGACAACAAAGATTTGTTTTACATC AACCTTTACCCAGATTATGCAGCTATCAGCAGGGCGATCCTGGATCTGGTCCTCTATT ACAACTGGAAAACAGTGACAGTGGTGTATGAAGACAGCACAGGTCTAATTCGTCTACA AGAGCTCATCAAAGCTCCCTCCAGATATAATATTAAAATCAAAATCCGCCAGCTGCCC TCTGGGAATAAAGATGCCAAGCCTTTACTCAAGGAGATGAAGAAAGGCAAGGAGTTCT ATGTGATATTTGATTGTTCACATGAAACAGCCGCTGAAATCCTTAAGCAGATTCTGTT CATGGGCATGATGACCGAGTACTATCACTACTTTTTCACAACCCTGGACTTATTTGCT TTGGATCTGGAACTCTATAGGTACAGTGGCGTAAACATGACCGGGTTTCGGCTGCTTA ACATTGACAACCCTCACGTGTCATCCATCATTGAGAAGTGGTCCATGGAGAGACTGCA GGCCCCACCCAGGCCCGAGACTGGCCTTTTGGATGGCATGATGACAACTGAAGCGGCT CTGATGTACGATGCTGTGTACATGGTGGCCATTGCCTCGCACCGGGCATCCCAGCTGA CCGTCAGCTCCCTGCAGTGCCATAGACATAAGCCATGGCGCCTCGGACCCAGATTTAT GAACCTGATCAAAGAGGCCCGGTGGGATGGCTTGACTGGGCATATCACCTTTAATAAA ACCAATGGCTTGAGGAAGGATTTTGATCTGGATATTATTAGTCTCAAAGAGGAAGGAA CTGAAAAGATTGGGATTTGGAATTCCAACAGTGGGCTTAACATGACGGACAGCAACAA AGACAAGTCCAGCAATATCACTGATTCATTGGCCAACAGAACACTCATTGTCACCACC ATTCTGGAAGAACCCTATGTTATGTACAGGAAATCTGATAAGCCTCTATATGGAAATG ACAGATTTGAAGGATATTGCCTAGACCTGTTGAAAGAATTGTCAAACATCCTGGGTTT CATTTATGATGTTAAACTAGTTCCCGATGGCAAATATGGGGCCCAGAATGACAAAGGG GAGTGGAACGGGATGGTTAAAGAACTCATAGATCACAGGGCTGACCTGGCAGTGGCTC CTCTTACCATCACCTACGTGCGGGAGAAAGTCATTGACTTCTCCAAACCCTTCATGAC CCTAGGCATCAGCATTCTCTACCGGAAGCCCAATGGTACCAATCCAGGCGTTTTCTCC TTCCTCAACCCCCTGTCTCCAGATATTTGGCTCGAG	
	ORF Start: at 1	ORF Stop: end of sequence
	SEQ ID NO: 80	534 aa	MW at 60947.3kD
NOV23b, 188822752 Protein Sequence	RSQVLRIGGIFETVENEPVNVEELAFKFAVTSINRNRTLMPNTTLTYDIQRINLFDSF EASRRACDQLALGVAALFGPSHSSSVSAVQSICNALEVPHIQTRWKHPSVDNKDLFYI NLYPDYAAISRAILDLVLYYNWKTVTVVYEDSTGLIRLQELIKAPSRYNIKIKIRQLP SGNKDAKPLLKEMKKGKEFYVIFDCSHETAAEILKQILFMGMMTEYYHYFFTTLDLFA LDLELYRYSGVNMTGFRLLNIDNPHVSSIIEKWSMERLQAPPRPETGLLDGMMTTEAA LMYDAVYMVAIASHRASQLTVSSLQCHRHKPWRLGPRFMNLIKEARWDGLTGHITFNK TNGLRKDFDLDIISLKEEGTEKIGIWNSNSGLNMTDSNKDKSSNITDSLANRTLIVTT ILEEPYVMYRKSDKPLYGNDRFEGYCLDLLKELSNILGFIYDVKLVPDGKYGAQNDKG EWNGMVKELIDHRADLAVAPLTITYVREKVIDFSKPFMTLGISILYRKPNGTNPGVFS FLNPLSPDIWLE	

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

Table 23B. Comparison of NOV23a against NOV23b.		
Protein Sequence	**NOV23a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV23b	35..565 3..533	492/531 (92%) 493/531 (92%)

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

Table 23C. Protein Sequence Properties NOV23a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 35 and 36

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

Table 23D. Geneseq Results for NOV23a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV23a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAR63069	Human EAA3c excitatory amino acid receptor - Homo sapiens, 865 aa. [CA2110933-A, 12-JUN-1994]	1..862 1..862	853/862 (98%) 853/862 (98%)	0.0
AAB19496	The Q591 form of the human EAA3 receptor - Homo sapiens, 905 aa. [US6136544-A, 24-OCT-2000]	1..858 1..858	851/858 (99%) 851/858 (99%)	0.0
AAR75883	Human EAA3 receptor (Q-591) - Homo sapiens, 905 aa. [WO9517508-A2, 29-JUN-1995]	1..858 1..858	851/858 (99%) 851/858 (99%)	0.0

AAR60112	Human EAA3a excitatory amino acid receptor - Homo sapiens, 905 aa. [CA2110933-A, 12-JUN-1994]	1..858 1..858	851/858 (99%) 851/858 (99%)	0.0
AAB19499	Amino acid sequence of the R591 form of the human EAA3 receptor - Homo sapiens, 905 aa. [US6136544-A, 24-OCT-2000]	1..858 1..858	850/858 (99%) 851/858 (99%)	0.0

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

Table 23E. Public BLASTP Results for NOV23a

Protein Accession Number	Protein/Organism/Length	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P39086	Glutamate receptor, ionotropic kainate 1 precursor (Glutamate receptor 5) (GLUR-5) (GluR5) (Excitatory amino acid receptor 3) (EAA3) - Homo sapiens (Human), 918 aa.	1..903 1..918	900/918 (98%) 900/918 (98%)	0.0
CAC80546	GLUTAMATE RECEPTOR SUBUNIT GLUR5 - Homo sapiens (Human), 905 aa.	1..858 1..858	852/858 (99%) 852/858 (99%)	0.0
P22756	Glutamate receptor, ionotropic kainate 1 precursor (Glutamate receptor 5) (GLUR-5) (GluR5) - Rattus norvegicus (Rat), 949 aa.	1..854 1..869	823/869 (94%) 838/869 (95%)	0.0
Q9DGM1	GLUTAMATE RECEPTOR 5 - Danio aequipinnatus (Giant danio) (Brachydanio aequipinnatus), 880 aa.	32..854 32..868	735/837 (87%) 789/837 (93%)	0.0
Q60934	Glutamate receptor, ionotropic kainate 1 precursor (Glutamate receptor 5) (GLUR-5) (GluR5) - Mus musculus (Mouse), 836 aa.	1..756 1..758	707/758 (93%) 727/758 (95%)	0.0

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

Table 23F. Domain Analysis of NOV23a			
Pfam Domain	**NOV23a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ANF_receptor: domain 1 of 1	25..415	95/466 (20%) 351/466 (75%)	8.9e-114
SBP_bac_3: domain 1 of 1	434..801	46/425 (11%) 216/425 (51%)	0.79
lig_chan: domain 1 of 1	560..841	161/322 (50%) 272/322 (84%)	4.8e-161

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: 81	1443 bp	
NOV24a, CG93722-01 DNA Sequence	AGCATTACCCTACTCCTAGAGGGATTGTAGAGATTAGTGCCATCATCAAGGACTTCAA AGATGCCAGAGTGGCTAG**ATG**CAATGGGTCGTGCCTGTAATCCCAACACTTTGGGAGG CCAAGGCAGGAGGATTGCTTGAGGTCAGGAGTTTGAGCTTGGGCAACATAGTGAGACT CTGCCTCTACAAAAAGTCTTTGAAAAAAGATACCAGAGTGGTGATTCCTAAAACATCC CCATTTAGTTTGCCTATTTGGCAGAAGACAGATGGATCTTGGAGAAGACTCCACTGCA CTTCCACCACCAGTGCACACACACATGGACCCAGCTGCACTGCTGCCCTGCTACTGCT GGCACATGCACACGATTGTGGAACAGCACCGCTTAAGGATGTGTTGCAAGGGTCTCGG ATTATAGGGGGCACCGAAGCACAAGCTGGCGCATGGCCGTGGGTGGTGAGCCTGCAGA TTAAATATGGCCGTGTTCTTGTTCATGTATGTGGGGGAACCCTAGTGAGAGAGAGGTG GGTCCTCACAGCTGCCCACTGCACTAAAGACACTAGGTACGTATTCAGAACACAACTA TTTAGCGATCCTTTAATGTGGACAGCTGTGATTGGAACTAATAATATACATGGACGCT ATCCTCATACCAAGAAGATAAAAATTAAAGCAATCATTATTCATCCAAACTTCATTTT GGAATCTTATGTAAATGATATTGCACTTTTTCACTTAAAAAAAGCAGTGAGGTATAAT GACTATATTCAGCCTATTTGCCTACCTTTTGATGTTTTCCAAATCCTGGACGGAAACA CAAAGTGTTTTATAAGTGGCTGGGGAAGAACAAAAGAAGAAGGTAATTTGCAGCCGCT TTGTTTACCTACTCAAGCCTCCGCAATGGTGTGCTCCAAGATAACTTACTGGTATTTT TTGTTAACAGGTAACGCTACAAATATTTTACAAGATGCAGAAGTGCATTATATTTCTC GAGAGATGTGTAATTCTGAGAGGAGTTATGGGGGAATAATTCCTAACACTTCATTTTG TGCAGGTGATGAAGATGGAGCTTTTGATACTTGCAGGGGTGACAGTGGGGGACCATTA ATGTGCTACTTACCAGAATATAAAAGATTTTTTGTAATGGGAATTACCAGTTACGGAC ATGGCTGTGGTCGAAGAGGTTTTCCTGGTGTCTATATTGGGCCATCCTTCTACCAAAA GTGGCTGACAGAGCATTTCTTCCATGCAAGCACTCAAGGCATACTTACTATAAATATT TTACGTGGCCAGATCCTCATAGCTTTATGTTTTGTCATCTTACTAGCAACAACA**TAA**A GAAATTCTGAAGGCTTTCATATCTTTATTTTGCATTGTGTCCCTTTCTATGTTCTATA TAATGAACATCATTTATTCTTCTAGCAATTAATTGCCTACATTAGAGATTT		
	ORF Start: ATG at 77	ORF Stop: TAA at 1331	
	SEQ ID NO: 82	418 aa	MW at 46709.8kD
NOV24a, CG93722-01	MQWVVPVIPTLWEAKAGGLLEVRSLSLGNIVRLCLYKKSLKKDTRVVIPKTSPFSLPI WQKTDGSWRRLHCTSTTSAHTHGPSCTAALLLLAHAHDCGTAPLKDVLQGSRIIGGTE		

Protein Sequence	AQAGAWPWVVSLQIKYGRVLVHVCGGTLVRERWVLTAAHCTKDTRYVFRTQLFSDPLM WTAVIGTNNIHGRYPHTKKIKIKAIIIHPNFILESYVNDIALFHLKKAVRYNDYIQPI CLPFDVFQILDGNTKCFISGWGRTKEEGNLQPLCLPTQASAMVCSKITYWYFLLTGNA TNILQDAEVHYISREMCNSERSYGGIIPNTSFCAGDEDGAFDTCRGDSGGPLMCYLPE YKRFFVMGITSYGHGCGRRGFPGVYIGPSFYQKWLTEHFFHASTQGILTINILRGQIL IALCFVILLATT

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

Table 24B. Protein Sequence Properties NOV24a

PSort analysis:	0.9325 probability located in endoplasmic reticulum (membrane); 0.6976 probability located in plasma membrane; 0.3200 probability located in microbody (peroxisome); 0.1900 probability located in Golgi body
SignalP analysis:	Cleavage site between residues 17 and 18

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 24C.

Table 24C. Geneseq Results for NOV24a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU03900	Human protease-like polypeptide #2 - Homo sapiens, 348 aa. [WO200149864-A1, 12-JUL-2001]	92..418 59..348	288/327 (88%) 288/327 (88%)	e-168
AAU03901	Human protease-like polypeptide #3 - Homo sapiens, 288 aa. [WO200149864-A1, 12-JUL-2001]	96..418 3..288	284/323 (87%) 285/323 (87%)	e-166
AAU03899	Human protease-like polypeptide #1 - Homo sapiens, 217 aa. [WO200149864-A1, 12-JUL-2001]	174..418 1..217	217/245 (88%) 217/245 (88%)	e-126
AAW96812	A mouse serine protease called hepsin - Mus musculus, 416 aa. [WO9854307-A1, 03-DEC-1998]	56..397 62..414	119/398 (29%) 178/398 (43%)	4e-41
AAY43325	Mouse hepsin protein sequence - Mus musculus, 416 aa. [US5981830-A, 09-NOV-1999]	56..397 62..414	119/398 (29%) 178/398 (43%)	4e-41

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

Table 24D. Public BLASTP Results for NOV24a				
Protein Accession Number	**Protein/Organism/Length**	**NOV24a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAL50817	AIRWAY TRYPSIN-LIKE PROTEASE - Rattus norvegicus (Rat), 417 aa.	96..385 171..413	104/290 (35%) 146/290 (49%)	1e-42
Q29015	PREPROACROSIN - Sus sp, 415 aa.	85..383 4..283	116/318 (36%) 150/318 (46%)	2e-42
P08001	Acrosin precursor (EC 3.4.21.10) (53 kDa fucose-binding protein) - Sus scrofa (Pig), 415 aa.	85..383 4..283	116/318 (36%) 150/318 (46%)	2e-42
Q9QZ74	ADRENAL SECRETORY SERINE PROTEASE PRECURSOR - Rattus norvegicus (Rat), 279 aa.	96..385 33..275	104/290 (35%) 145/290 (49%)	3e-42
O35453	Serine protease hepsin (EC 3.4.21.-) - Mus musculus (Mouse), 416 aa.	84..397 139..414	104/324 (32%) 156/324 (48%)	2e-40

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

Table 24E. Domain Analysis of NOV24a			
Pfam Domain	**NOV24a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
trypsin: domain 1 of 2	111..263	67/174 (39%) 115/174 (66%)	1.2e-41
trypsin: domain 2 of 2	287..383	41/105 (39%) 72/105 (69%)	4.1e-25

5

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: 83	2867 bp
NOV25a, CG93858-01 DNA Sequence	TCATTTTAGGGGCTCTGTTTTCATCTCAGATTATTCTGTCTTGTAGCCCATGGTAACT GGAGTCCTTGGAGTGGCTGGGGAACATGCAGCCGGACGTGTAACGGAGGGCAG**ATG**CG GCGGTACCGCACATGTGATAACCCTCCTCCCTCCAATGGGGGAAGAGCTTGTGGGGGA CCAGACTCCCAGATCCAGAGGTGCAACACTGACATGTGTCCTGTGGATGGAAGTTGGG GAAGCTGGCATAGTTGGAGCCAGTGCTCTGCCTCCTGTGGAGGAGGTGAAAAGACTCG

	GAAGCGGCTGTGCGACCATCCTGTGCCAGTTAAAGGTGGCCGTCCTTGTCCCGGAGAC ACTACTCAGGTGACCAGGTGCAATGTACAAGCATGTCCAGGTGGGCCCCAGCGAGCCA GAGGAAGTGTTATTGGAAATATTAATGATGTTGAATTTGGAATTGCTTTCCTTAATGC CACAATAACTGATAGCCCTAACTCTGATACTAGAATAATACGTGCCAAAATTACCAAT GTACCTCGTAGTCTTGGTTCAGCAATGAGAAAGATAGTTTCTATTCTAAATCCCATTT ATTGGACAACAGCAAAGGAAATAGGAGAAGCAGTCAATGGCTTTACCCTCACCAATGC AGTCTTCAAAAGAGAAACTCAAGTGGAATTTGCAACTGGAGAAATCTTGCAGATGAGT CATATTGCCCGGGGCTTGGATTCCGATGGTTCTTTGCTGCTAGATATCGTTGTGAGTG GCTATGTCCTACAGCTTCAGTCACCTGCTGAAGTCACTGTAAAGGATTACACAGAGGA CTACATTCAAACAGGTCCTGGGCAGCTGTACGCCTACTCAACCCGGCTGTTCACCATT GATGGCATCAGCATCCCATACACATGGAACCACACCGTTTTCTATGATCAGGCACAGG GAAGAATGCCTTTCTTGGTTGAAACACTTCATGCATCCTCTGTGGAATCTGACTATAA CCAGATAGAAGAGACACTGGGTTTTAAAATTCATGCTTCAATATCCAAAGGAGATCGC AGTAATCAGTGCCCCTCCGGGTTTACCTTAGACTCAGTTGGACCTTTTTGTGCTGATG AGGATGAATGTGCAGCAGGGAATCCCTGCTCCCATAGCTGCCACAATGCCATGGGGAC TTACTACTGCTCCTGCCCTAAAGGCCTCACCATAGCTGCAGATGGAAGAACTTGTCAA GATATTGATGAGTGTGCTTTGGGTAGGCATACCTGCCACGCTGGTCAGGACTGTGACA ATACGATTGGATCTTATCGCTGTGTGGTCCGTTGTGGAAGTGGCTTTCGAAGAACCTC TGATGGGCTGAGTTGTCAAGATATTAATGAATGTCAAGAATCCAGCCCCTGTCACCAG CGCTGTTTCAATGCCATAGGAAGTTTCCATTGTGGATGTGAACCTGGGTATCAGCTCA AAGGCAGAAAATGCATGGATGTGAACGAGTGTAGACAAAATGTATGCAGACCAGATCA GCACTGTAAGAACACCCGTGGTGGCTATAAGTGCATTGATCTTTGTCCAAATGGAATG ACCAAGGCAGAAAATGGAACCTGTATTGATATTGATGAATGTAAAGATGGGACCCATC AGTGCAGATATAACCAGATATGTGAGAATACAAGAGGCAGCTATCGTTGTGTATGCCC AAGAGGTTATCGGTCTCAAGGAGTTGGAAGACCCTGCATGGATATTGATGAATGTGAA AATACAGATGCCTGCCAGCATGAGTGTAAGAATACCTTTGGAAGTTATCAGTGCATCT GCCCACCTGGCTATCAACTCACACACAATGGAAAGACATGCCAAGATATCGATGAATG TCTGGAGCAGAATGTGCACTGTGGACCCAATCGCATGTGCTTCAACATGAGAGGAAGC TACCAGTGCATCGATACACCCTGTCCACCCAACTACCAACGGGATCCTGTTTCAGGGT TCTGCCTCAAGAACTGTCCACCCAATGATTTGGAATGTGCCTTGAGCCCATATGCCTT GGAATACAAACTCGTCTCCCTCCCATTTGGAATAGCCACCAATCAAGATTTAATCCGG CTGGTTGCATACACACAGGATGGAGTGATGCATCCCAGGACAACTTTCCTCATGGTAG ATGAGGAACAGACTGTTCCTTTTGCCTTGAGGGATGAAAACCTGAAAGGAGTGGTGTA TACAACACGACCACTACGAGAAGCAGAGACCTACCGCATGAGGGTCCGAGCCTCATCC TACAGTGCCAATGGGACCATTGAATATCAGACCACATTCATAGTTTATATAGCTGTGT CCGCCTATCCATAC**TAA**GGAACTCTCCAAAGCCTATTCCACATATTTAAACCGCATTA ATCATGGCAATCAAGCCCCCTTCCAGATTACTGTCTCTTGAACAGTTGCAATCTTGGC AGCTTGAAAATGGTGCTACACTCTGTTTTGTGTGCCTTCCTTGGTACTTCTGAGGTAT TTTCATGATCCCACCATGGTCATATCTTGAAGTATGGTCTAGAAAAGTCCCTTATTAT TTTATTTATTACACTGGAGCAGTTACTTCCCAAAGATTATTCTGAACATCTAACAGGA CATATCAGTGATGGTTTACAGTAGTGTAGTACCTAAGATCATTTTCCTGAAAGCCAAA CCAAACAACGAAAAACAAGAACAACTAATTCAGAATCAAATAGAGTTTTTGAGCATTT GACTATTTTTAGAATCATAAAATTAGTTACTAAGTATTTTGATCAAAGCTTATAAAAT AACTTACGGAGATTTTTGTAAGTATTGATACATTATAATAGGACTTGCCTATTTTCAT TTTTAAGAAGAAAAACACCACTCAT
	ORF Start: ATG at 112 | ORF Stop: TAA at 2335
	SEQ ID NO: 84 | 741 aa | MW at 81868.0kD
NOV25a, CG93858-01 Protein Sequence	MRRYRTCDNPPPSNGGRACGGPDSQIQRCNTDMCPVDGSWGSWHSWSQCSASCGGGEK TRKRLCDHPVPVKGGRPCPGDTTQVTRCNVQACPGGPQRARGSVIGNINDVEFGIAFL NATITDSPNSDTRIIRAKITNVPRSLGSAMRKIVSILNPIYWTTAKEIGEAVNGFTLT NAVFKRETQVEFATGEILQMSHIARGLDSDGSLLLDIVVSGYVLQLQSPAEVTVKDYT EDYIQTGPGQLYAYSTRLFTIDGISIPYTWNHTVFYDQAQGRMPFLVETLHASSVESD YNQIEETLGFKIHASISKGDRSNQCPSGFTLDSVGPFCADEDECAAGNPCSHSCHNAM GTYYCSCPKGLTIAADGRTCQDIDECALGRHTCHAGQDCDNTIGSYRCVVRCGSGFRR TSDGLSCQDINECQESSPCHQRCFNAIGSFHCGCEPGYQLKGRKCMDVNECRQNVCRP DQHCKNTRGGYKCIDLCPNGMTKAENGTCIDIDECKDGTHQCRYNQICENTRGSYRCV CPRGYRSQGVGRPCMDIDECENTDACQHECKNTFGSYQCICPPGYQLTHNGKTCQDID ECLEQNVHCGPNRMCFNMRGSYQCIDTPCPPNYQRDPVSGFCLKNCPPNDLECALSPY ALEYKLVSLPFGIATNQDLIRLVAYTQDGVMHPRTTFLMVDEEQTVPFALRDENLKGV

	VYTTRPLREAETYRMRVRASSYSANGTIEYQTTFIVYIAVSAYPY	
	SEQ ID NO: 85	8243 bp
NOV25b, CG93858-02 DNA Sequence	GCAGAGTACAGTGGTTGGATTTATATTTAGTAAATGGGAATAT**ATG**TTGATAACACCT GCTTTCACTTTTAATATATTTACTATTATAGTTCCTCCAAGTGTCATTGGTCCTAAAT CTGAAAATCTTACCGTCGTGGTGAACAATTTCATCTCTTTGACCTGTGAGGTCTCTGG TTTTCCACCTCCTGACCTCAGCTGGCTCAAGAATGAACAGCCCATCAAACTGAACACA AATACTCTCATTGTGCCTGGTGGTCGAACTCTACAGATTATTCGGGCCAAGGTATCAG ATGGTGGTGAATACACTTGTATAGCTATCAATCAAGCTGGCGAAAGCAAGAAAAAGTT TTCCCTGACTGTTTATGTGCCCCCAAGCATTAAAGACCATGACAGTGAATCTCTTTCT GTAGTTAATGTAAGAGAGGGAACTTCTGTGTCTTTGGAGTGTGAGTCGAACGCTGTGC CACCTCCAGTCATCACTTGGTATAAGAATGGGCGGATGATAACAGAGTCTACTCATGT GGAGATTTTAGCTGATGGACAAATGCTACACATTAAGAAAGCTGAGGTATCTGACACA GGCCAGTATGTATGTAGAGCTATAAATGTAGCAGGACGGGATGATAAAAATTTCCACC TCAATGTATATGTGCCACCCAGTATTGAAGGACCTGAAAGAGAAGTGATTGTGGAGAC GATCAGCAATCCTGTGACATTAACATGTGATGCCACTGGGATCCCACCTCCCACGATA GCATGGTTAAAGAACCACAAGCGCATAGAAAATTCTGACTCACTGGAAGTTCGTATTT TGTCTGGAGGTAGCAAACTCCAGATTGCCCGGTCTCAGCATTCAGATAGTGGAAACTA TACATGTATTGCTTCAAATATGGAGGGAAAAGCCCAGAAATATTACTTTCTTTCAATT CAAGTTCCTCCAAGTGTTGCTGGTGCTGAAATTCCAAGTGATGTCAGTGTCCTTCTAG GAGAAAATGTTGAGCTGGTCTGCAATGCAAATGGCATTCCTACTCCACTTATTCAATG GCTTAAAGATGGAAAGCCCATAGCTAGTGGTGAAACAGAAAGAATCCGAGTGAGTGCA AATGGCAGCACATTAAACATTTATGGAGCTCTTACATCTGACACGGGGAAATACACAT GTGTTGCTACTAATCCCGCTGGAGAAGAAGACCGAATTTTTAACTTGAATGTCTATGT TACACCTACAATTAGGGGTAATAAAGATGAAGCAGAGAAACTAATGACTTTAGTGGAT ACTTCAATAAATATTGAATGCAGAGCCACAGGGACGCCTCCACCACAGATAAACTGGC TGAAGAATGGACTTCCTCTGCCTCTCTCCTCCCATATCCGGTTACTGGCAGCAGGACA AGTTATCAGGATTGTGAGAGCTCAGGTGTCTGATGTCGCTGTGTATACTTGTGTGGCC TCCAACAGAGCTGGGGTGGATAATAAGCATTACAATCTTCAAGTGTTTGCACCACCAA ATATGGACAATTCAATGGGGACAGAGGAAATCACAGTTCTCAAAGGTAGTTCCACCTC TATGGCATGCATTACTGATGGAACCCCAGCTCCCAGTATGGCCTGGCTTAGAGATGGC CAGCCTCTGGGGCTTGATGCCCATCTGACAGTCAGCACCCATGGAATGGTCCTGCAGC TCCTCAAAGCAGAGACTGAAGATTCGGGAAAGTACACCTGCATTGCCTCAAATGAAGC TGGAGAAGTCAGCAAGCACTTTATCCTCAAGGTCCTAGAACCACCTCACATTAATGGA TCTGAAGAACATGAAGAGATATCAGTAATTGTTAATAACCCACTTGAACTTACCTGCA TTGCTTCTGGAATCCCAGCCCCTAAAATGACCTGGATGAAAGATGGCCGGCCCCTTCC ACAGACGGATCAAGTGCAAACTCTAGGAGGAGGAGAGGTTCTTCGAATTTCTACTGCT CAGGTGGAGGATACAGGAAGATATACATGTCTGGCATCCAGTCCTGCAGGAGATGATG ATAAGGAATATCTAGTGAGAGTGCATGTACCTCCTAATATTGCTGGAACTGATGAGCC CCGGGATATCACTGTGTTACGGAACAGACAAGTGACATTGGAATGCAAGTCAGATGCA GTGCCCCCACCTGTAATTACTTGGCTCAGAAATGGAGAACGGTTACAGGCAACACCTC GAGTGCGAATCCTATCTGGAGGGAGATACTTGCAAATCAACAATGCTGACCTAGGTGA TACAGCCAATTATACCTGTGTTGCCAGCAACATTGCAGGAAAGACTACAAGAGAATTT ATTCTCACTGTAAATGTTCCTCCAAACATAAAGGGGGGCCCCCAGAGCCTTGTAATTC TTTTAAATAAGTCAACTGTATTGGAATGCATCGCTGAAGGTGTCCCAACTCCAAGGAT AACATGGAGAAAGGATGGAGCTGTTCTAGCTGGGAATCATGCAAGATATTCCATCTTG GAAAATGGATTCCTTCATATTCAATCAGCACATGTCACTGACACTGGACGGTATTTGT GTATGGCCACCAATGCTGCTGGAACAGATCGCAGGCGAATAGATTTACAGGTCCATGT TCCTCCATCTATTGCTCCGGGTCCTACCAACATGACTGTAATAGTAAATGTTCAAACT ACTCTGGCTTGTGAGGCTACTGGGATACCAAAACCATCAATCAATTGGAGAAAAAATG GGCATCTTCTTAATGTGGATCAAAATCAGAACTCATACAGGCTCCTTTCTTCAGGTTC ACTAGTAATTATTTCCCCTTCTGTGGATGACACTGCAACCTATGAATGTACTGTGACA AACGGTGCTGGAGATGATAAAAGAACTGTGGATCTCACTGTCCAAGTTCCACCTTCCA TAGCTGATGAGCCTACAGATTTCCTAGTAACCAAACATGCCCCAGCAGTAATTACCTG CACTGCTTCGGGAGTTCCATTTCCCTCAATTCACTGGACCAAAAATGGTATAAGACTG CTTCCCAGGGGAGATGGCTATAGAATTCTGTCCTCAGGAGCAATTGAAATACTTGCCA CCCAATTAAACCATGCTGGAAGATACACTTGTGTCGCTAGGAATGCGGCTGGCTCTGC ACATCGACACGTGACCCTTCATGTTCATGAGCCTCCAGTCATTCAGCCCCAACCAAGT GAACTACACGTCATTCTGAACAATCCTATTTTATTACCATGTGAAGCAACAGGGACAC CCAGTCCTTTCATTACTTGGCAAAAAGAAGGCATCAATGTTAACACTTCAGGCAGAAA CCATGCAGTTCTTCCTAGTGGCGGCTTACAGATCTCCAGAGCTGTCCGAGAGGATGCT	

GGCACTTACATGTGTGTGGCCCAGAACCCGGCTGGTACAGCCTTGGGCAAAATCAAGT
TAAATGTCCAAGTTCCTCCAGTCATTAGCCCTCATCTAAAGGAATATGTTATTGCTGT
GGACAAGCCCATCACGTTATCCTGTGAAGCAGATGGCCTCCCTCCGCCTGACATTACA
TGGCATAAAGATGGGCGTGCAATTGTGGAATCTATCCGCCAGCGCGTCCTCAGCTCTG
GCTCTCTGCAAATAACATTTGTCCAGCCTGGTGATGCTGGCCATTACACGTGCATGGC
AGCCAATGTAGCAGGATCAAGCAGCACAAGCACCAAGCTCACCGTCCATGTACCACCC
AGGATCAGAAGTACAGAAGGACACTACACGGTCAATGAGAATTCACAAGCCATTCTTC
CATGCGTAGCTGATGGAATCCCCACACCAGCAATTAACTGGAAAAAAGACAATGTTCT
TTTAGCTAACTTGTTAGGAAAATACACTGCTGAACCATATGGAGAACTCATTTTAGAA
AATGTTGTGCTGGAGGATTCTGGCTTCTATACCTGTGTTGCTAACAATGCTGCAGGTG
AAGATACACACACTGTCAGCCTGACTGTGCATGTTCTCCCCACTTTTACTGAACTTCC
TGGAGACGTGTCATTAAATAAAGGAGAACAGCTACGATTAAGCTGTAAAGCTACTGGT
ATTCCATTGCCCAAATTAACATGGACCTTCAATAACAATATTATTCCAGCCCACTTTG
ACAGTGTGAATGGACACAGTGAACTTGTTATTGAAAGAGTGTCAAAAGAGGATTCAGG
TACTTATGTGTGCACCGCAGAGAACAGCGTTGGCTTTGTGAAGGCAATTGGATTTGTG
TATGTGAAAGAACCTCCAGTCTTCAAAGGTGATTATCCTTCTCACTGGATTGAACCAC
TTGGTGGGAATGCAATCCTGAATTGTGAGGTGAAAGGAGACCCCACCCCAACCATCCA
GTGGAACAGAAAGGGAGTGGATATTGAAATTAGCCACAGAATCCGGCAACTGGGCAAT
GGCTCCCTGGCCATCTATGGCACTGTTAATGAAGATGCCGGTGACTATACATGTGTAG
CTACCAATGAAGCTGGGGTGGTGGAGCGCAGCATGAGTCTGACTCTGCAAAGTCCTCC
TATTATCACTCTTGAGCCAGTGGAAACTGTTATTAATGCTGGTGGCAAAATCATATTG
AATTGTCAGGCAACTGGAGAGCCTCAACCAACCATTACATGGTCCCGTCAAGGGCACT
CTATTTCCTGGGATGACCGGGTTAACGTGTTGTCCAACAACTCATTATATATTGCTGA
TGCTCAGAAAGAAGATACCTCTGAATTTGAATGTGTTGCTCGAAACTTAATGGGTTCT
GTCCTTGTCAGAGTGCCAGTCATAGTCCAGGTTCATGGTGGATTTTCCCAGTGGTCTG
CATGGAGAGCCTGCAGTGTCACCTGTGGAAAAGGCATCCAAAAGAGGAGTCGTCTGTG
CAACCAGCCCCTTCCAGCCAATGGTGGGAAGCCCTGCCAAGGTTCAGATTTGGAAATG
CGAAACTGTCAAAATAAGCCTTGTCCAGTGGATGGTAGCTGGTCGGAATGGAGTCTTT
GGGAAGAATGCACAAGGAGCTGTGGACGCGGCAACCAAACCAGGACCAGGACTTGCAA
TAATCCATCAGTTCAGCATGGTGGGCGGCCATGTGAAGGGAATGCTGTGGAAATAATT
ATGTGCAACATTAGGCCTTGCCCAGTTCATGGAGCATGGAGCGCTTGGCAGCCTTGGG
GAACATGCAGCGAAAGTTGTGGGAAAGGTACTCAGACAAGAGCAAGACTTTGTAATAA
CCCACCACCAGCGTTTGGTGGGTCCTACTGTGATGGAGCAGAAACACAGATGCAAGTT
TGCAATGAAAGAAATTGTCCAGTTCATGGCAAGTGGGCGACTTGGGCCAGTTGGAGTG
CCTGTTCTGTGTCATGTGGAGGAGGTGCCAGACAGAGAACAAGGGGCTGCTCCGACCC
TGTGCCCCAGTATGGAGGAAGGAAATGCGAAGGGAGTGATGTCCAGAGTGATTTTTGC
AACAGTGACCCTTGCCCAACCCATGGTAACTGGAGTCCTTGGAGTGGCTGGGGAACAT
GCAGCCGGACGTGTAACGGAGGGCAGATGCGGCGGTACCGCACATGTGATAACCCTCC
TCCCTCCAATGGGGGAAGAGCTTGTGGGGGACCAGACTCCCAGATCCAGAGGTGCAAC
ACTGACATGTGTCCTGTGGATGGAAGTTGGGGAAGCTGGCATAGTTGGAGCCAGTGCT
CTGCCTCCTGTGGAGGAGGTGAAAAGACTCGGAAGCGGCTGTGCGACCATCCTGTGCC
AGTTAAAGGTGGCCGTCCCTGTCCCGGAGACACTACTCAGGTGACCAGGTGCAATGTA
CAAGCATGTCCAGGTGGGCCCCAGCGAGCCAGAGGAAGTGTTATTGGAAATATTAATG
ATGTTGAATTTGGAATTGCTTTCCTTAATGCCACAATAACTGATAGCCCTAACTCTGA
TACTAGAATAATACGTGCCAAAATTACCAATGTACCTCGTAGTCTTGGTTCAGCAATG
AGAAAGATAGTTTCTATTCTAAATCCCATTTATTGGACAACAGCAAAGGAAATAGGAG
AAGCAGTCAATGGCTTTACCCTCACCAATGCAGTCTTCAAAAGAGAAACTCAAGTGGA
ATTTGCAACTGGAGAAATCTTGCAGATGAGTCATATTGCCCGGGGCTTGGATTCCGAT
GGTTCTTTGCTGCTAGATATCGTTGTGAGTGGCTATGTCCTACAGCTTCAGTCACCTG
CTGAAGTCACTGTAAAGGATTACACAGAGGACTACATTCAAACAGGTCCTGGGCAGCT
GTACGCCTACTCAACCCGGCTGTTCACCATTGATGGCATCAGCATCCCATACACATGG
AACCACACCGTTTTCTATGATCAGGCACAGGGAAGAATGCCTTTCTTGGTTGAAACAC
TTCATGCATCCTCTGTGGAATCTGACTATAACCAGATAGAAGAGACACTGGGTTTTAA
AATTCATGCTTCAATATCCAAAGGAGATCGCAGTAATCAGTGCCCCTCCGGGTTTACC
TTAGACTCAGTTGGACCTTTTTGTGCTGATGAGGATGAATGTGCAGCAGGGAATCCCT
GCTCCCATAGCTGCCACAATGCCATGGGGACTTACTACTGCTCCTGCCCTAAAGGCCT
CACCATAGCTGCAGATGGAAGAACTTGTCAAGATATTGATGAGTGTGCTTTGGGTAGG
CATACCTGCCACGCTGGTCAGGACTGTGACAATACGATTGGATCTTATCGCTGTGTGG
TCCGTTGTGGAAGTGGCTTTCGAAGAACCTCTGATGGGCTGAGTTGTCAAGATATTAA
TGAATGTCAAGAATCCAGCCCCTGTCACCAGCGCTGTTTCAATGCCATAGGAAGTTTC
CATTGTGGATGTGAACCTGGGTATCAGCTCAAAGGCAGAAAATGCATGGATGTGAACG

	AGTGTAGACAAAATGTATGCAGACCAGATCAGCACTGTAAGAACACCCGTGGTGGCTA TAAGTGCATTGATCTTTGTCCAAATGGAATGACCAAGGCAGAAAATGGAACCTGTATT GATATTGATGAATGTAAAGATGGGACCCATCAGTGCAGATATAACCAGATATGTGAGA ATACAAGAGGCAGCTATCGTTGTGTATGCCCAAGAGGTTATCGGTCTCAAGGAGTTGG AAGACCCTGCATGGATATTGATGAATGTGAAAATACAGATGCCTGCCAGCATGAGTGT AAGAATACCTTTGGAAGTTATCAGTGCATCTGCCCACCTGGCTATCAACTCACACACA ATGGAAAGACATGCCAAGATATCGATGAATGTCTGGAGCAGAATGTGCACTGTGGACC CAATCGCATGTGCTTCAACATGAGAGGAAGCTACCAGTGCATCGATACACCCTGTCCA CCCAACTACCAACGGGATCCTGTTTCAGGGTTCTGCCTCAAGAACTGTCCACCCAATG ATTTGGAATGTGCCTTGAGCCCATATGCCTTGGAATACAAACTCGTCTCCCTCCCATT TGGAATAGCCACCAATCAAGATTTAATCCGGCTGGTTGCATACACACAGGATGGAGTG ATGCATCCCAGGACAACTTTCCTCATGGTAGATGAGGAACAGACTGTTCCTTTTGCCT TGAGGGATGAAAACCTGAAAGGAGTGGTGTATACAACACGACCACTACGAGAAGCAGA GACCTACCGCATGAGGGTCCGAGCCTCATCCTACAGTGCCAATGGGACCATTGAATAT CAGACCACATTCATAGTTTATATAGCTGTGTCCGCCTATCCATACTAAGGAACTCTCC AAAGCCTATTCCACATATTTAAACCGCATTAATCATGGCAATCAAGCCCCCTTCCAGA TTACTGTCTCTTGAACAGTTGCAATCTTGGCAGCTTGAAAATGGTGCTACACTCTGTT TTGTGTGCCTTCCTTGGTACTTCTGAGGTATTTTCATGATCCCACCATGGTCATATCT TGAAGTATGGTCTAGAAAAGTCCCTTATTATTTTATTTATTACACTGGAGCAGTTACT TCCCAAAGATTATTCTGAACATCTAACAGGACATATCAGTGATGGTTTACAGTAGTGT AGTACCTAAGATCATTTTCCTGAAAGCCAAACCAAACAACGAAAAACAAGAACAACTA ATTCAGAATCAAATAGAGTTTTTGAGCATTTGACTATTTTTAGAATCATAAAATTAGT TACTAAGTATTTTGATCAAAGCTTATAAAATAACTTACGGAGAATTTTGTAAGTATTG ATACATT		
	ORF Start: ATG at 44	ORF Stop: TAA at 7760	
	SEQ ID NO: 86	2572 aa	MW at 279540.0kD
NOV25b, CG93858-02 Protein Sequence	MLITPAFTFNIFTIIVPPSVIGPKSENLTVVVNNFISLTCEVSGFPPPDLSWLKNEQP IKLNTNTLIVPGGRTLQIIRAKVSDGGEYTCIAINQAGESKKKFSLTVYVPPSIKDHD SESLSVVNVREGTSVSLECESNAVPPPVITWYKNGRMITESTHVEILADGQMLHIKKA EVSDTGQYVCRAINVAGRDDKNFHLNVYVPPSIEGPEREVIVETISNPVTLTCDATGI PPPTIAWLKNHKRIENSDSLEVRILSGGSKLQIARSQHSDSGNYTCIASNMEGKAQKY YFLSIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIQWLKDGKPIASGETER IRVSANGSTLNIYGALTSDTGKYTCVATNPAGEEDRIFNLNVYVTPTIRGNKDEAEKL MTLVDTSINIECRATGTPPPQINWLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAV YTCVASNRAGVDNKHYNLQVFAPPNMDNSMGTEEITVLKGSSTSMACITDGTPAPSMA WLRDGQPLGLDAHLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFILKVLEP PHINGSEEHEEISVIVNNPLELTCIASGIPAPKMTWMKDGRPLPQTDQVQTLGGGEVL RISTAQVEDTGRYTCLASSPAGDDDKEYLVRVHVPPNIAGTDEPRDITVLRNRQVTLE CKSDAVPPPVITWLRNGERLQATPRVRILSGGRYLQINNADLGDTANYTCVASNIAGK TTREFILTVNVPPNIKGGPQSLVILLNKSTVLECIAEGVPTPRITWRKDGAVLAGNHA RYSILENGFLHIQSAHVTDTGRYLCMATNAAGTDRRRIDLQVHVPPSIAPGPTNMTVI VNVQTTLACEATGIPKPSINWRKNGHLLNVDQNQNSYRLLSSGSLVIISPSVDDTATY ECTVTNGAGDDKRTVDLTVQVPPSIADEPTDFLVTKHAPAVITCTASGVPFPSIHWTK NGIRLLPRGDGYRILSSGAIEILATQLNHAGRYTCVARNAAGSAHRHVTLHVHEPPVI QPQPSELHVILNNPILLPCEATGTPSPFITWQKEGINVNTSGRNHAVLPSGGLQISRA VREDAGTYMCVAQNPAGTALGKIKLNVQVPPVISPHLKEYVIAVDKPITLSCEADGLP PPDITWHKDGRAIVESIRQRVLSSGSLQITFVQPGDAGHYTCMAANVAGSSSTSTKLT VHVPPRIRSTEGHYTVNENSQAILPCVADGIPTPAINWKKDNVLLANLLGKYTAEPYG ELILENVVLEDSGFYTCVANNAAGEDTHTVSLTVHVLPTFTELPGDVSLNKGEQLRLS CKATGIPLPKLTWTFNNNIIPAHFDSVNGHSELVIERVSKEDSGTYVCTAENSVGFVK AIGFVYVKEPPVFKGDYPSHWIEPLGGNAILNCEVKGDPTPTIQWNRKGVDIEISHRI RQLGNGSLAIYGTVNEDAGDYTCVATNEAGVVERSMSLTLQSPPIITLEPVETVINAG GKIILNCQATGEPQPTITWSRQGHSISWDDRVNVLSNNSLYIADAQKEDTSEFECVAR NLMGSVLVRVPVIVQVHGGFSQWSAWRACSVTCGKGIQKRSRLCNQPLPANGGKPCQG SDLEMRNCQNKPCPVDGSWSEWSLWEECTRSCGRGNQTRTRTCNNPSVQHGGRPCEGN AVEIIMCNIRPCPVHGAWSAWQPWGTCSESCGKGTQTRARLCNNPPPAFGGSYCDGAE TQMQVCNERNCPVHGKWATWASWSACSVSCGGGARQRTRGCSDPVPQYGGRKCEGSDV QSDFCNSDPCPTHGNWSPWSGWGTCSRTCNGGQMRRYRTCDNPPPSNGGRACGGPDSQ IQRCNTDMCPVDGSWGSWHSWSQCSASCGGGEKTRKRLCDHPVPVKGGRPCPGDTTQV		

	TRCNVQACPGGPQRARGSVIGNINDVEFGIAFLNATITDSPNSDTRIIRAKITNVPRS LGSAMRKIVSILNPIYWTTAKEIGEAVNGFTLTNAVFKRETQVEFATGEILQMSHIAR GLDSDGSLLLDIVVSGYVLQLQSPAEVTVKDYTEDYIQTGPGQLYAYSTRLFTIDGIS IPYTWNHTVFYDQAQGRMPFLVETLHASSVESDYNQIEETLGFKIHASISKGDRSNQC PSGFTLDSVGPFCADEDECAAGNPCSHSCHNAMGTYYCSCPKGLTIAADGRTCQDIDE CALGRHTCHAGQDCDNTIGSYRCVVRCGSGFRRTSDGLSCQDINECQESSPCHQRCFN AIGSFHCGCEPGYQLKGRKCMDVNECRQNVCRPDQHCKNTRGGYKCIDLCPNGMTKAE NGTCIDIDECKDGTHQCRYNQICENTRGSYRCVCPRGYRSQGVGRPCMDIDECENTDA CQHECKNTFGSYQCICPPGYQLTHNGKTCQDIDECLEQNVHCGPNRMCFNMRGSYQCI DTPCPPNYQRDPVSGFCLKNCPPNDLECALSPYALEYKLVSLPFGIATNQDLIRLVAY TQDGVMHPRTTFLMVDEEQTVPFALRDENLKGVVYTTRPLREAETYRMRVRASSYSAN GTIEYQTTFIVYIAVSAYPY
	SEQ ID NO: 87 6343 bp
NOV25c, CG56914-03 DNA Sequence	AACCACCTCACATTAATGGATCTGAAGAACATGAAGAGATATCAGTAATTGTTAATAA CCCACTTGAACTTACCTGCATTGCTTCTGGAATCCCAGCCCCTAAA**ATG**ACCTGGATG AAAGATGGCCGGCCCCTTCCACAGACGGATCAAGTGCAAACTCTAGGAGGAGGAGAGG TTCTTCGAATTTCTACTGCTCAGGTGGAGGATACAGGAAGATATACATGTCTGGCATC CAGTCCTGCAGGAGATGATGATAAGGAATATCTAGTGAGAGTGCATGTACCTCCTAAT ATTGCTGGAACTGATGAGCCCCGGGATATCACTGTGTTACGGAACAGACAAGTGACAT TGGAATGCAAGTCAGATGCAGTGCCCCCACCTGTAATTACTTGGCTCAGAAATGGAGA ACGGTTACAGGCAACACCTCGAGTGCGAATCCTATCTGGAGGGAGATACTTGCAAATC AACAATGCTGACCTAGGTGATACAGCCAATTATACCTGTGTTGCCAGCAACATTGCAG GAAAGACTACAAGAGAATTTATTCTCACTGTAAATGTTCCTCCAAACATAAAGGGGGG CCCCCAGAGCCTTGTAATTCTTTTAAATAAGTCAACTGTATTGGAATGCATCGCTGAA GGTGTGCCAACTCCAAGGATAACATGGAGAAAGGATGGAGCTGTTCTAGCTGGGAATC ATGCAAGATATTCCATCTTGGAAAATGGATTCCTTCATATTCAATCAGCACATGTCAC TGACACTGGACGGTATTTGTGTATGGCCACCAATGCTGCTGGAACAGATCGCAGGCGA ATAGATTTACAGGTCCATGGTTCACTAGTAATTATTTCCCCTTCTGTGGATGACACTG CAACCTATGAATGTACTGTGACAAACGGTGCTGGAGATGATAAAAGAACTGTGGATCT CACTGTCCAAGTTCCACCTTCCATAGCTGATGAGCCTACAGATTTCCTAGTAACCAAA CATGCCCCAGCAGTAATTACCTGCACTGCTTCGGGAGTTCCATTTCCCTCAATTCACT GGACCAAAAATGGTATAAGACTGCTTCCCAGGGGAGATGGCTATAGAATTCTGTCCTC AGGAGCAATTGAAATACTTGCCACCCAATTAAACCATGCTGGAAGATACACTTGTGTC GCTAGGAATGCGGCTGGCTCTGCACATCGACACGTGACCCTTCATGTTCATGAGCCTC CAGTCATTCAGCCCCAACCAAGTGAACTACACGTCATTCTGAACAATCCTATTTTATT ACCATGTGAAGCAACAGGGACACCCAGTCCTTTCATTACTTGGCAAAAAGAAGGCATC AATGTTAACACTTCAGGCAGAAACCATGCAGTTCTTCCTAGTGGCGGCTTACAGATCT CCAGAGCTGTCCGAGAGGATGCTGGCACTTACATGTGTGTGGCCCAGAACCCGGCTGG TACAGCCTTGGGCAAAATCAAGTTAAATGTCCAAGTTCCTCCAGTCATTAGCCCTCAT CTAAAGGAATATGTTATTGCTGTGGACAAGCCCATCACGTTATCCTGTGAAGCAGATG GCCTCCCTCCGCCTGACATTACATGGCATAAAGATGGGCGTGCAATTGTGGAATCTAT CCGCCAGCGCGTCCTCAGCTCTGGCTCTCTGCAAATAGCATTTGTCCAGCCTGGTGAT GCTGGCCATTACACGTGCATGGCAGCCAATGTAGCAGGATCAAGCAGCACAAGCACCA AGCTCACCGTCCATGTACCACCCAGGATCAGAAGTACAGAAGGACACTACACGGTCAA TGAGAATTCACAAGCCATTCTTCCATGCGTAGCTGATGGAATCCCCACACCAGCAATT AACTGGAAAAAAGACAATGTTCTTTTAGCTAACTTGTTAGGAAAATACACTGCTGAAC CATATGGAGAACTCATTTTAGAAAATGTTGTGCTGGAGGATTCTGGCTTCTATACCTG TGTTGCTAACAATGCTGCAGGTGAAGATACACACACTGTCAGCCTGACTGTGCATGTT CTCCCCACTTTTACTGAACTTCCTGGAGACGTGTCATTAAATAAAGGAGAACAGCTAC GATTAAGCTGTAAAGCTACTGGTATTCCATTGCCCAAATTAACATGGACCTTCAATAA CAATATTATTCCAGCCCACTTTGACAGTGTGAATGGACACAGTGAACTTGTTATTGAA AGAGTGTCAAAAGAGGATTCAGGTACTTATGTGTGCACCGCAGAGAACAGCGTTGGCT TTGTGAAGGCAATTGGATTTGTTTATGTGAAAGAACCTCCAGTCTTCAAAGGTGATTA TCCTTCTAACTGGATTGAACCACTTGGTGGGAATGCAATCCTGAATTGTGAGGTGAAA GGAGACCCCACCCCAACCATCCAGTGGAACAGAAAGGGAGTGGATATTGAAATTAGCC ACAGAATCCGGCAACTGGGCAATGGCTCCCTGGCCATCTATGGCACTGTTAATGAAGA TGCCGGTGACTATACATGTGTAGCTACCAATGAAGCTGGGGTGGTGGAGCGCAGCATG AGTCTGACTCTGCAAAGTCCTCCTATTATCACTCTTGAGCCAGTGGAAACTGTTATTA ATGCTGGTGGCAAAATCATATTGAATTGTCAGGCAACTGGAGAGCCTCAACCAACCAT TACATGGTCCCGTCAAGGGCACTCTATTTCCTGGGATGACCGGGTTAACGTGTTGTCC

AACAACTCATTATATATTGCTGATGCTCAGAAAGAAGATACCTCTGAATTTGAATGCG
TTGCTCGAAACTTAATGGGTTCTGTCCTTGTCAGAGTGCCAGTCATAGTCCAGGTTCA
TGGTGGATTTTCCCAGTGGTCTGCATGGAGAGCCTGCAGTGTCACCTGTGGAAAAGGC
ATCCAAAAGAGGAGTCGTCTGTGCAACCAGCCCCTTCCAGCCAATGGTGGGAAGCCCT
GCCAAGGTTCAGATTTGGAAATGCGAAACTGTCAAAATAAGCCTTGTCCAGTGGATGG
TAGCTGGTCGGAATGGAGTCTTTGGGAAGAATGCACAAGGAGCTGTGGACGCGGCAAC
CAAACCAGGACCAGGACTTGCAATAATCCATCAGTTCAGCATGGTGGGCGGCCATGTG
AAGGGAATGCTGTGGAAATAATTATGTGCAACATTAGGCCTTGCCCAGTTCATGGAGC
ATGGAGCGCTTGGCAGCCTTGGGGAACATGCAGCGAAAGTTGTGGGAAAGGTACTCAG
ACAAGAGCAAGACTTTGTAATAACCCACCACCAGCGTTTGGTGGGTCCTACTGTGATG
GAGCAGAAACACAGATGCAAGTTTGCAATGAAAGAAATTGTCCAATTCATGGCAAGTG
GGCGACTTGGGCCAGTTGGAGTGCCTGTTCTGTGTCATGTGGAGGAGGTGCCAGACAG
AGAACAAGGGGCTGCTCCGACCCTGTGCCCCAGTATGGAGGAAGGAAATGCGAAGGGA
GTGATGTCCAGAGTGATTTTTGCAACAGTGACCCTTGCCCAACCCATGGTAACTGGAG
TCCTTGGAGTGGCTGGGGAACATGCAGCCGGACGTGTAACGGAGGGCAGATGCGGCGG
TACCGCACATGTGATAACCCTCCTCCCTCCAATGGGGGAAGAGCTTGTGGGGGACCAG
ACTCCCAGATCCAGAGGTGCAACACTGACATGTGTCCTGTGGATGGAAGTTGGGGAAG
CTGGCATAGTTGGAGCCAGTGCTCTGCCTCCTGTGGAGGAGGTGAAAAGACTCGGAAG
CGGCTGTGCGACCATCCTGTGCCAGTTAAAGGTGGCCGTCCTTGTCCCGGAGACACTA
CTCAGGTGACCAGGTGCAATGTACAAGCATGTCCAGGTGGGCCCCAGCGAGCCAGAGG
AAGTGTTATTGGAAATATTAATGATGTTGAATTTGGAATTGCTTTCCTTAATGCCACA
ATAACTGATAGCCCTAACTCTGATACTAGAATAATACGTGCCAAAATTACCAATGTAC
CTCGTAGTCTTGGTTCAGCAATGAGAAAGATAGTTTCTATTCTAAATCCCATTTATTG
GACAACAGCAAAGGAAATAGGAGAAGCAGTCAATGGCTTTACCCTCACCAATGCAGTC
TTCAAAAGAGAAACTCAAGTGGAATTTGCAACTGGAGAAATCTTGCAGATGAGTCATA
TTGCCCGGGGCTTGGATTCCGATGGTTCTTTGCTGCTAGATATCGTTGTGAGTGGCTA
TGTCCTACAGCTTCAGTCACCTGCTGAAGTCACTGTAAAGGATTACACAGAGGACTAC
ATTCAAACAGGTCCTGGGCAGCTGTACGCCTACTCAACCCGGCTGTTCACCATTGATG
GCATCAGCATCCCATACACATGGAACCACACCGTTTTCTATGATCAGGCACAGGGAAG
AATGCCTTTCTTGGTTGAAACACTTCATGCATCCTCTGTGGAATCTGACTATAACCAG
ATAGAAGAGACACTGGGTTTTAAAATTCATGCTTCAATATCCAAAGGAGATCGCAGTA
ATCAGTGCCCCTCCGGGTTTACCTTAGACTCAGTTGGACCTTTTTGTGCTGATGAGGA
TGAATGTGCAGCAGGGAATCCCTGCTCCCATAGCTGCCACAATGCCATGGGGACTTAC
TACTGCTCCTGCCCTAAAGGCCTCACCATAGCTGCAGATGGAAGAACTTGTCAAGATA
TTGATGAGTGTGCTTTGGGTAGGCATACCTGCCACGCTGGTCAGGACTGTGACAATAC
GATTGGATCTTATCGCTGTGTGGTCCGTTGTGGAAGTGGCTTTCGAAGAACCTCTGAT
GGGCTGAGTTGTCAAGATATTAATGAATGTCAAGAATCCAGCCCCTGTCACCAGCGCT
GTTTCAATGCCATAGGAAGTTTCCATTGTGGATGTGAACCTGGGTATCAGCTCAAAGG
CAGAAAATGCATGGATGTGAACGAGTGTAGACAAAATGTATGCAGACCAGATCAGCAC
TGTAAGAACACCCGTGGTGGCTATAAGTGCATTGATCTTTGTCCAAATGGAATGACCA
AGGCAGAAAATGGAACCTGTATTGATATTGATGAATGTAAAGATGGGACCCATCAGTG
CAGATATAACCAGATATGTGAGAATACAAGAGGCAGCTATCGTTGTGTATGCCCAAGA
GGTTATCGGTCTCAAGGAGTTGGAAGACCCTGCATGGATATTGATGAATGTGAAAATA
CAGATGCCTGCCTGCATGAGTGTAAGAATACCTTTGGAAGTTATCAGTGCATCTGCCC
ACCTGGCTATCAACTCACACACAATGGAAAGACATGCCAAGATATCGATGAATGTCTG
GAGCAGAATGTGCACTGTGGACCCAATCGCATGTGCTTCAACATGAGAGGAAGCTACC
AGTGCATCGATACACCCTGTCCACCCAACTACCAACGGGATCCTGCTTCAGGGTTCTG
CCTCAAGAACTGTCCACCCAATGATTTGGAATGTGCCTTGAGCCCATATGCCTTGGAA
TACAAACTCGTCTCCCTCCCATTTGGAATAGCCACCAATCAAGATTTAATCCGGCTGG
TTGCATACACACAGGATGGAGTGATGCATCCCAGGACAACTTTCCTCATGGTAGATGA
GGAACAGACTGTTCCTTTTGCCTTGAGGGATGAAAACCTGAAAGGAGTGGTGTATACA
ACACGACCACTACGAGAAGCAGAGACCTACCGCATGAGGGTCCGAGCCTCATCCTACA
GTGCCAATGGGACCATTGAATATCAGACCACATTCATAGTTTATATAGCTGTGTCCGC
CTATCCATAC**TAA**GGAACTCTCCAAAGCCTATTCCACATATTTAAACCGCATTAATCA
TGGCAATCAAGCCCCCTTCCAGATTACTGTCTCTTGAACAGTTGCAATCTTGGCAGCT
TGAAAATGGTGCTACACTCTGTTTTGTGTGCCTTCCTTGGTACTTCTGAGGTATTTTC
ATGATCCCACCATGGTCATATCTTGAAGTATGGTCTAGAAAAGTCCCTTATTATTTTA
TTTATTACACTGGAGCAGTTACTTCCCAAAGATTATTCTGAACATCTAACAGGACATA
TCAGTGATGGTTTACAGTAGTGTAGTACCTAAGATCATTTTCCTGAAAGCCAAACCAA
ACAACGAAAAACAAGAACAACTAATTCAGAATCAAATAGAGTTTTTGAGCATTTGACT
ATTTTTAGAATCATAAAATTAGTTACTAAGTATTTTGATCAAAGCTTATAAAATAACT

	TACGGAGATTTTTGTAAGTATTGATACATTATAATAGGACTTGCCTATTTTCATTTTT AAGAAGAAAAACACCACTCAT		
	ORF Start: ATG at 105	ORF Stop: TAA at 5811	
	SEQ ID NO: 88	1902 aa	MW at 207163.2kD
NOV25c, CG56914-03 Protein Sequence	MTWMKDGRPLPQTDQVQTLGGGEVLRISTAQVEDTGRYTCLASSPAGDDDKEYLVRVH VPPNIAGTDEPRDITVLRNRQVTLECKSDAVPPPVITWLRNGERLQATPRVRILSGGR YLQINNADLGDTANYTCVASNIAGKTTREFILTVNVPPNIKGGPQSLVILLNKSTVLE CIAEGVPTPRITWRKDGAVLAGNHARYSILENGFLHIQSAHVTDTGRYLCMATNAAGT DRRRIDLQVHGSLVIISPSVDDTATYECTVTNGAGDDKRTVDLTVQVPPSIADEPTDF LVTKHAPAVITCTASGVPFPSIHWTKNGIRLLPRGDGYRILSSGAIEILATQLNHAGR YTCVARNAAGSAHRHVTLHVHEPPVIQPQPSELHVILNNPILLPCEATGTPSPFITWQ KEGINVNTSGRNHAVLPSGGLQISRAVREDAGTYMCVAQNPAGTALGKIKLNVQVPPV ISPHLKEYVIAVDKPITLSCEADGLPPPDITWHKDGRAIVESIRQRVLSSGSLQIAFV QPGDAGHYTCMAANVAGSSSTSTKLTVHVPPRIRSTEGHYTVNENSQAILPCVADGIP TPAINWKKDNVLLANLLGKYTAEPYGELILENVVLEDSGFYTCVANNAAGEDTHTVSL TVHVLPTFTELPGDVSLNKGEQLRLSCKATGIPLPKLTWTFNNNIIPAHFDSVNGHSE LVIERVSKEDSGTYVCTAENSVGFVKAIGFVYVKEPPVFKGDYPSNWIEPLGGNAILN CEVKGDPTPTIQWNRKGVDIEISHRIRQLGNGSLAIYGTVNEDAGDYTCVATNEAGVV ERSMSLTLQSPPIITLEPVETVINAGGKIILNCQATGEPQPTITWSRQGHSISWDDRV NVLSNNSLYIADAQKEDTSEFECVARNLMGSVLVRVPVIVQVHGGFSQWSAWRACSVT CGKGIQKRSRLCNQPLPANGGKPCQGSDLEMRNCQNKPCPVDGSWSEWSLWEECTRSC GRGNQTRTRTCNNPSVQHGGRPCEGNAVEIIMCNIRPCPVHGAWSAWQPWGTCSESCG KGTQTRARLCNNPPPAFGGSYCDGAETQMQVCNERNCPIHGKWATWASWSACSVSCGG GARQRTRGCSDPVPQYGGRKCEGSDVQSDFCNSDPCPTHGNWSPWSGWGTCSRTCNGG QMRRYRTCDNPPPSNGGRACGGPDSQIQRCNTDMCPVDGSWGSWHSWSQCSASCGGGE KTRKRLCDHPVPVKGGRPCPGDTTQVTRCNVQACPGGPQRARGSVIGNINDVEFGIAF LNATITDSPNSDTRIIRAKITNVPRSLGSAMRKIVSILNPIYWTTAKEIGEAVNGFTL TNAVFKRETQVEFATGEILQMSHIARGLDSDGSLLLDIVVSGYVLQLQSPAEVTVKDY TEDYIQTGPGQLYAYSTRLFTIDGISIPYTWNHTVFYDQAQGRMPFLVETLHASSVES DYNQIEETLGFKIHASISKGDRSNQCPSGFTLDSVGPFCADEDECAAGNPCSHSCHNA MGTYYCSCPKGLTIAADGRTCQDIDECALGRHTCHAGQDCDNTIGSYRCVVRCGSGFR RTSDGLSCQDINECQESSPCHQRCFNAIGSFHCGCEPGYQLKGRKCMDVNECRQNVCR PDQHCKNTRGGYKCIDLCPNGMTKAENGTCIDIDECKDGTHQCRYNQICENTRGSYRC VCPRGYRSQGVGRPCMDIDECENTDACLHECKNTFGSYQCICPPGYQLTHNGKTCQDI DECLEQNVHCGPNRMCFNMRGSYQCIDTPCPPNYQRDPASGFCLKNCPPNDLECALSP YALEYKLVSLPFGIATNQDLIRLVAYTQDGVMHPRTTFLMVDEEQTVPFALRDENLKG VVYTTRPLREAETYRMRVRASSYSANGTIEYQTTFIVYIAVSAYPY

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 25B.

Table 25B. Comparison of NOV25a against NOV25b and NOV25c.

Protein Sequence	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV25b	1..741 1832..2572	730/741 (98%) 730/741 (98%)
NOV25c	1..741 1162..1902	728/741 (98%) 728/741 (98%)

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

Table 25C. Protein Sequence Properties NOV25a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence

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

5

Table 25D. Geneseq Results for NOV25a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV25a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB95002	Human protein sequence SEQ ID NO:16644 - Homo sapiens, 741 aa. [EP1074617-A2, 07-FEB-2001]	1..741 1..741	741/741 (100%) 741/741 (100%)	0.0
AAU16959	Human novel secreted protein, SEQ ID 200 - Homo sapiens, 877 aa. [WO200155441-A2, 02-AUG-2001]	1..741 137..877	741/741 (100%) 741/741 (100%)	0.0
AAG67241	Amino acid sequence of human thrombospondin 1-like protein - Homo sapiens, 780 aa. [WO200109321-A1, 08-FEB-2001]	1..741 40..780	741/741 (100%) 741/741 (100%)	0.0
AAG67244	Amino acid sequence of murine thrombospondin 1-like protein - Mus musculus, 1068 aa. [WO200109321-A1, 08-FEB-2001]	1..741 328..1068	673/741 (90%) 707/741 (94%)	0.0
AAG67243	Amino acid sequence of murine thrombospondin 1-like protein - Mus musculus, 744 aa. [WO200109321-A1, 08-FEB-2001]	1..741 4..744	673/741 (90%) 707/741 (94%)	0.0

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

10

Table 25E. Public BLASTP Results for NOV25a				
Protein Accession Number	**Protein/Organism/Length**	**NOV25a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96K89	CDNA FLJ14438 FIS, CLONE HEMBB1000317, WEAKLY SIMILAR TO FIBULIN-1, ISOFORM D PRECURSOR - Homo sapiens (Human), 741 aa.	1..741 1..741	741/741 (100%) 741/741 (100%)	0.0
Q96SC3	FIBULIN-6 - Homo sapiens (Human), 2673 aa (fragment).	1..580 1816..2396	559/581 (96%) 567/581 (97%)	0.0
Q96RW7	HEMICENTIN - Homo sapiens (Human), 5636 aa.	1..580 4779..5359	559/581 (96%) 566/581 (97%)	0.0
Q95NZ3	F56H11.1B PROTEIN - Caenorhabditis elegans, 689 aa.	311..741 223..689	160/480 (33%) 224/480 (46%)	4e-62
Q9TZS1	FIBULIN-1D - Caenorhabditis elegans, 589 aa (fragment).	311..741 123..589	160/480 (33%) 224/480 (46%)	4e-62

PFam analysis indicates that the NOV25a protein contains the domains shown in Table 25F.

Table 25F. Domain Analysis of NOV25a			
Pfam Domain	**NOV25a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
tsp_1: domain 1 of 1	41..91	23/54 (43%) 39/54 (72%)	6.7e-13
EGF: domain 1 of 7	334..368	16/47 (34%) 25/47 (53%)	8.4e-06
granulin: domain 1 of 1	355..370	7/16 (44%) 11/16 (69%)	4.2
EGF: domain 2 of 7	374..413	14/48 (29%) 25/48 (52%)	2
EGF: domain 3 of 7	419..451	12/47 (26%) 24/47 (51%)	0.0045
EGF: domain 4 of 7	457..493	14/47 (30%) 24/47 (51%)	13
TILa: domain 1 of 1	467..522	20/62 (32%) 32/62 (52%)	7.7

Keratin_B2: domain 1 of 1	383..525	34/191 (18%) 70/191 (37%)	8.7
EGF: domain 5 of 7	499..536	14/47 (30%) 28/47 (60%)	0.0013
EGF: domain 6 of 7	542..576	17/47 (36%) 28/47 (60%)	1.3e-07
EGF: domain 7 of 7	582..622	13/49 (27%) 26/49 (53%)	17
fn2: domain 1 of 1	611..622	7/12 (58%) 8/12 (67%)	7.8
cadherin: domain 1 of 1	643..735	15/107 (14%) 54/107 (50%)	5.2

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: 89	2018 bp
NOV26a, CG93871-01 DNA Sequence	CTCCCCACGGCGCCAGGAGGAGGGGCGAGGGCCGGCAGCCCCCTCTCCCGCGCGCGGC GCAGGAGCCGAGCCCAGCCCGGGGGACCCGCCGCCGCCGGTC**ATG**TGGGCCGGACTGC TCCTTCGGGCCGCCTGTGTCGCGCTCCTGCTGCCGGGGGCACCAGCCCGAGGCTACAC CGGGAGGAAGCCGCCCGGGCACTTCGCGGCCGAGAGGCGCCGACTGGGCCCCCACGTC TGCCTCTCTGGGTTTGGGAGTGGCTGCTGCCCTGGCTGGGCGCCCTCTATGGGTGGTG GGCACTGCACCCTGCGTCTCTGCTCCTTCGGCTGTGGGAGTGGCATCTGCATCGCTCC CAATGTCTGCTCCTGCCAGGATGGAGAGCAAGGGGCCACCTGCCCAGAAACCCATGGA CCATGTGGGGAGTACGGCTGTGACCTTACCTGCAACCATGGAGGCTGTCAGGAGGTGG CCCGAGTGTGCCCCGTGGGCTTCTCGATGACGGAGACAGCTGTTGGCATCAGGTGTGA CATTGACGAATGTGTAACCTCCTCCTGCGAGGGCCACTGTGTGAACACAGAAGGTGGG TTTGTGTGCGAGTGTGGGCCGGGCATGCAGCTGTCTGCCGACCGCCACAGCTGCCAAG ACACTGACGAATGCCTAGGGACTCCCTGTCAGCAGAGATGTAAAAACAGCATTGGCAG CTACAAGTGTTCCTGTCGAACTGGCTTCCACCTTCATGGCAACCGGCACTCCTGTGTA GATGTAAACGAGTGTCGGAGGCCATTGGAGAGGCGAGTCTGTCACCATTCCTGCCACA ACACCGTGGGCAGCTTCCTATGCACATGCCGACCTGGCTTCAGGCTCCGAGCTGACCG CGTGTCCTGTGAAGCTTTCCCGAAAGCCGTGCTGGCCCCATCTGCCATCCTGCAACCC CGGCAACACCCGTCCAAGATGCTTCTGTTGCTTCCTGAGGCCGGCCGGCCTGCCCTGT CCCCAGGACATAGCCCTCCTTCTGGGGCTCCAGGGCCCCCAGCCGGAGTCAGGACCAC CCGCCTGCCATCTCCCACCCCACGACTACCCACATCCTCCCCTTCTGCCCCTGTGTGG CTGCTGTCCACCCTGCTGGCCACCCCAGTGCCTACTGCCTCCCTGCTGGGGAACCTCA GACCCCCCTCACTCCTTCAGGGGGAGGTGATGGGGACCCCTTCCTCACCCAGGGGCCC TGAGTCCCCCCGACTGGCAGCAGGGCCCTCTCCCTGCTGGCACCTGGGAGCCATGCAT GAATCAAGGAGTCGCTGGACAGAGCCTGGGTGTTCCCAGTGCTGGTGCGAGGATGGGA AGGTGACCTGTGAAAAGGTGAGGTGTGAAGCTGCTTGTTCCCACCCAATTCCCTCCAG AGATGGTGGGTGCTGCCCATCGTGCACAGGTTGTTTTCACAGTGGTGTCGTCCGAGCT GAAGGGGATGTGTTTTCACCTCCCAATGAGAACTGCACCGTCTGTGTCTGTCTGGCTG GAAACGTGTCGTGCATGTTTCGTGAGTGTCCTTTTGGCCCGTGTGAGACCCCCCATAA AGACAGATGCTATTTCCACGGCCGGTGGTACGCAGACGGGGCTGTGTTCAGTGGGGGT GGTGACGAGTGTACCACCTGTGTTTGCCAGAATGGGGAGGTGGAGTGCTCCTTCATGC CCTGCCCTGAGCTGGCCTGCCCCCGAGAAGAGTGGCGGCTGGGCCCTGGGCAGTGTTG

	CTTCACCTGCCAGGAGCCCACACCCTCGACAGGTTGCTCTCTTGACGACAACGGGGTT GAGTTTCCGATTGGACAGATCTGGTCGCCTGGTGACCCCTGTAGATGGCTCGGTGAGC TGCAAGAGGACAGACTGTGTGGACTCCTGCCCTCACCCGATCCGGATCCCTGGACAGT GCTGCCCAGACTGTTCAGCAGG**TAA**TCCCCTGCCTCTGCCCCAAGCCCCCAGGGCAGG GCATCTCAGGCATCGGGCTCCTTAAGCCCTATACAGCCTTCATCTC		
	ORF Start: ATG at 101	ORF Stop: TAA at 1937	
	SEQ ID NO: 90	612 aa	MW at 65156.4kD
NOV26a, CG93871-01 Protein Sequence	MWAGLLLRAACVALLLPGAPARGYTGRKPPGHFAAERRRLGPHVCLSGFGSGCCPGWA PSMGGGHCTLRLCSFGCGSGICIAPNVCSCQDGEQGATCPETHGPCGEYGCDLTCNHG GCQEVARVCPVGFSMTETAVGIRCDIDECVTSSCEGHCVNTEGGFVCECGPGMQLSAD RHSCQDTDECLGTPCQQRCKNSIGSYKCSCRTGFHLHGNRHSCVDVNECRRPLERRVC HHSCHNTVGSFLCTCRPGFRLRADRVSCEAFPKAVLAPSAILQPRQHPSKMLLLLPEA GRPALSPGHSPPSGAPGPPAGVRTTRLPSPTPRLPTSSPSAPVWLLSTLLATPVPTAS LLGNLRPPSLLQGEVMGTPSSPRGPESPRLAAGPSPCWHLGAMHESRSRWTEPGCSQC WCEDGKVTCEKVRCEAACSHPIPSRDGGCCPSCTGCFHSGVVRAEGDVFSPPNENCTV CVCLAGNVSCMFRECPFGPCETPHKDRCYFHGRWYADGAVFSGGGDECTTCVCQNGEV ECSFMPCPELACPREEWRLGPGQCCFTCQEPTPSTGCSLDDNGVEFPIGQIWSPGDPC RWLGELQEDRLCGLLPSPDPDPWTVLPRLFSR		

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

Table 26B. Protein Sequence Properties NOV26a

PSort analysis:	0.5947 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 22 and 23

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 26C.

Table 26C. Geneseq Results for NOV26a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB85364	Novel Von Willebrand/thrombosporin-like polypeptide - Homo sapiens, 235 aa. [WO200153485-A1, 26-JUL-2001]	283..490 1..208	201/208 (96%) 203/208 (96%)	e-124
AAM99920	Human polypeptide SEQ ID NO 36 - Homo sapiens, 272 aa. [WO200155173-A2, 02-AUG-2001]	388..580 5..205	185/201 (92%) 188/201 (93%)	e-120

AAM99933	Human polypeptide SEQ ID NO 49 - Homo sapiens, 212 aa. [WO200155173-A2, 02-AUG-2001]	388..580 5..205	181/201 (90%) 185/201 (91%)	e-117
AAB85365	Novel Von Willebrand/thrombosporin-like mature protein sequence - Homo sapiens, 217 aa. [WO200153485-A1, 26-JUL-2001]	301..490 1..190	183/190 (96%) 185/190 (97%)	e-113
ABG15393	Novel human diagnostic protein #15384 - Homo sapiens, 1028 aa. [WO200175067-A2, 11-OCT-2001]	70..138 959..1027	69/69 (100%) 69/69 (100%)	7e-39

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

5

Table 26D. Public BLASTP Results for NOV26a

Protein Accession Number	Protein/Organism/Length	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96DN2	CDNA FLJ32009 FIS, CLONE NT2RP7009498, WEAKLY SIMILAR TO FIBULIN-1, ISOFORM A PRECURSOR - Homo sapiens (Human), 955 aa.	1..580 1..589	570/589 (96%) 573/589 (96%)	0.0
Q9DBE2	1300015B04RIK PROTEIN - Mus musculus (Mouse), 608 aa.	1..606 1..607	507/615 (82%) 538/615 (87%)	0.0
O00274	C1QR(P) - Homo sapiens (Human), 652 aa.	80..375 300..566	104/300 (34%) 134/300 (44%)	5e-32
Q9NPY3	DJ737E23.1 (COMPLEMENT COMPONENT C1Q RECEPTOR) - Homo sapiens (Human), 652 aa.	80..375 300..566	104/300 (34%) 134/300 (44%)	7e-32
Q91V88	POEM (NEPHRONECTIN SHORT ISOFORM) - Mus musculus (Mouse), 561 aa.	44..372 35..383	103/363 (28%) 152/363 (41%)	5e-31

PFam analysis indicates that the NOV26a protein contains the domains shown in Table 26E.

Table 26E. Domain Analysis of NOV26a			
Pfam Domain	**NOV26a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
EGF: domain 1 of 5	71..97	9/47 (19%) 16/47 (34%)	8.1
zf-NF-X1: domain 1 of 1	104..127	8/27 (30%) 13/27 (48%)	8
EGF: domain 2 of 5	109..140	10/47 (21%) 24/47 (51%)	25
EGF: domain 3 of 5	145..178	16/47 (34%) 23/47 (49%)	0.0045
EGF: domain 4 of 5	184..217	12/47 (26%) 25/47 (53%)	0.011
TIL: domain 1 of 1	165..223	13/70 (19%) 40/70 (57%)	0.53
EGF: domain 5 of 5	223..260	12/48 (25%) 26/48 (54%)	0.034
Keratin_B2: domain 1 of 1	93..271	39/213 (18%) 89/213 (42%)	6.2
TILa: domain 1 of 1	384..438	15/59 (25%) 28/59 (47%)	9.4
vwc: domain 1 of 3	385..439	21/84 (25%) 40/84 (48%)	7.8e-08
vwc: domain 2 of 3	442..492	18/84 (21%) 39/84 (46%)	0.00017
vwc: domain 3 of 3	493..550	21/84 (25%) 40/84 (48%)	1.8e-07

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: 91	2173 bp
NOV27a, CG93884-01 DNA	CTGAGGGCTCATCCCTCTGCAGAGCGCGGGGTCACCGGGAGGAGACGCC**ATG**ACGCCC GCCCTCACAGCCCTGCTCTGCCTTGGGCTGAGTCTGGGCCCCAGGACCCGCGTGCAGG CAGGGCCCTTCCCCAAACCCACCCTCTGGGCTGAGCCAGGCTCTGTGATCAGCTGGGG	

Sequence	GAGCCCCGTGACCATCTGGTGTCAGGGGAGCCTGGAGGCCCAGGAGTACCGACTGGAT AAAGAGGGAAGCCCAGAGCCCTTGGACAGAAATAACCCACTGGAACCCAAGAACAAGG CCAGATTCTCCATCCCATCCATGACAGAGCACCATGCGGGGAGATACCGCTGCCACTA TTACAGCTCTGCAGGCTGGTCAGAGCCCAGCGACCCCCTGGAGCTGGTGATGACAGGA TTCTACAACAAACCCACCCTCTCAGCCCTGCCCAGCCCTGTGGTGGCCTCAGGGGGGA ATATGACCCTCCGATGTGGCTCACAGAAGGGATATCACCATTTTGTTCTGATGAAGGA AGGAGAACACCAGCTCCCCCGGACCCTGGACTCACAGCAGCTCCACAGTGGGGGGTTC CAGGCCCTGTTCCCTGTGGGCCCCGTGAACCCCAGCCACAGGTGGAGGTTCACATGCT ATTACTATTATATGAACACCCCCCAGGTGTGGTCCCACCCCAGTGACCCCCTGGAGAT TCTGCCCTCAGGCGTGTCTAGGAAGCCCTCCCTCCTGACCCTGCAGGGCCCTGTCGTG GCCCCTGGGCAGAGCCTGACCCTCCAGTGTGGCTCTGATGTCGGCTACGACAGATTTG TTCTGTATAAGGAGGGGGAACGTGACTTCCTCCAGCGCCCTGGCCAGCAGCCCCAGGC TGGGCTCTCCCAGGCCAACTTCACCCTGGGCCCTGTGAGCCCCTCCCACGGGGGCCAG TACAGGTGCTATGGTGCACACAACCTCTCCTCCGAGTGGTCGGCCCCCAGCGACCCCC TGAACATCCTGATGGCAGGACAGATCTATGACACCGTCTCCCTGTCAGCACAGCCGGG CCCCACAGTGGCCTCAGGAGAGAACGTGACCCTGCTGTGTCAGTCATGGTGGCAGTTT GACACTTTCCTTCTGACCAAAGAAGGGGCAGCCCATCCCCCACTGCGTCTGAGATCAA TGTACGGAGCTCATAAGTACCAGGCTGAATTCCCCATGAGTCCTGTGACCTCAGCCCA CGCGGGGACCTACAGGTGCTACGGCTCATACAGCTCCAACCCCCACCTGCTGTCTTTC CCCAGTGAGCCCCTGGAACTCATGGTCTCAGGACACTCTGGAGGCTCCAGCCTCCCAC CCACAGGGCCGCCCTCCACACCTGGTCTGGGAAGATACCTGGAGGTTTTGATTGGGGT CTCGGTGGCCTTCGTCCTGCTGCTCTTCCTCCTCCTCTTCCTCCTCCTCCGACGTCAG CGTCACAGCAAACACAGGACATCTGACCAGAGAAAGACTGATTTCCAGCGTCCTGCAG GGGCTGCGGAGACAGAGCCCAAGGACAGGGGCCTGCTGAGGAGGTCCAGCCCAGCTGC TGACGTCCAGGAAGAAAACCTCTATGCTGCCGTGAAGGACACACAGTCTGAGGACAGG GTGGAGCTGGACAGTCAGCAGAGCCCACACGATGAAGACCCCCAGGCAGTGACGTATG CCCCGGTGAAACACTCCAGTCCTAGGAGAGAAATGGCCTCTCCTCCCTCCTCACTGTC TGGGGAATTCCTGGACACAAAGGACAGACAGGTGGAAGAGGACAGACAGATGGACACT GAGGCTGCTGCATCTGAAGCCTCCCAGGATGTGACCTACGCCCAGCTGCACAGCTTGA CCCTTAGACGGAAGGCAACTGAGCCTCCTCCATCCCAGGAAGGGGAACCTCCAGCTGA GCCCAGCATCTACGCCACTCTGGCCATCCAC**TAG**CCCGGGGGGTACGCAGACCCCACA CTCAGCAGAAGGAGACTCAGGACTGCTGAAGGCACGGGAGCTGCCCCCAGTGGACACC AGTGAACCCCAGTCAGCCTGGACCCCTAACACAGACCATGAGGAGACGCTGGGAACTT GTGGGACTCACCTGACTCAAAGATGACTAATATCGTCCCATTTTGGAAATAAAGCAAC AGACTTCTCAACAATCAATGAGTTAAT		
	ORF Start: ATG at 50	ORF Stop: TAG at 1946	
	SEQ ID NO: 92	632 aa	MW at 69499.3kD
NOV27a, CG93884-01 Protein Sequence	MTPALTALLCLGLSLGPRTRVQAGPFPKPTLWAEPGSVISWGSPVTIWCQGSLEAQEY RLDKEGSPEPLDRNNPLEPKNKARFSIPSMTEHHAGRYRCHYYSSAGWSEPSDPLELV MTGFYNKPTLSALPSPVVASGGNMTLRCGSQKGYHHFVLMKEGEHQLPRTLDSQQLHS GGFQALFPVGPVNPSHRWRFTCYYYYMNTPQVWSHPSDPLEILPSGVSRKPSLLTLQG PVVAPGQSLTLQCGSDVGYDRFVLYKEGERDFLQRPGQQPQAGLSQANFTLGPVSPSH GGQYRCYGAHNLSSEWSAPSDPLNILMAGQIYDTVSLSAQPGPTVASGENVTLLCQSW WQFDTFLLTKEGAAHPPLRLRSMYGAHKYQAEFPMSPVTSAHAGTYRCYGSYSSNPHL LSFPSEPLELMVSGHSGGSSLPPTGPPSTPGLGRYLEVLIGVSVAFVLLLFLLLFLLL RRQRHSKHRTSDQRKTDFQRPAGAAETEPKDRGLLRRSSPAADVQEENLYAAVKDTQS EDRVELDSQQSPHDEDPQAVTYAPVKHSSPRREMASPPSSLSGEFLDTKDRQVEEDRQ MDTEAAASEASQDVTYAQLHSLTLRRKATEPPPSQEGEPPAEPSIYATLAIH		

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

Table 27B. Protein Sequence Properties NOV27a	
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 24 and 25

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
AAB61263	Human monocyte inhibitory receptor precursor - Homo sapiens, 631 aa. [WO200100810-A1, 04-JAN-2001]	1..632 1..631	629/632 (99%) 630/632 (99%)	0.0
AAB04177	Leukocyte immunoglobulin like receptor pbm17 - Homo sapiens, 631 aa. [WO200068383-A2, 16-NOV-2000]	1..632 1..631	615/632 (97%) 623/632 (98%)	0.0
AAW82552	Human LIR-pbm17 protein - Homo sapiens, 631 aa. [WO9848017-A1, 29-OCT-1998]	1..632 1..631	615/632 (97%) 623/632 (98%)	0.0
ABG11435	Novel human diagnostic protein #11426 - Homo sapiens, 656 aa. [WO200175067-A2, 11-OCT-2001]	1..632 16..656	603/641 (94%) 615/641 (95%)	0.0
ABG11435	Novel human diagnostic protein #11426 - Homo sapiens, 656 aa. [WO200175067-A2, 11-OCT-2001]	1..632 16..656	603/641 (94%) 615/641 (95%)	0.0

In a BLAST search of public sequence datbases, 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
O15471	MONOCYTE INHIBITORY RECEPTOR PRECURSOR - Homo sapiens (Human), 631 aa.	1..632 1..631	630/632 (99%) 631/632 (99%)	0.0

AAC51900	IMMUNOGLOBULIN-LIKE TRANSCRIPT 5 - Homo sapiens (Human), 631 aa.	1..632 1..631	629/632 (99%) 631/632 (99%)	0.0
AAC51887	IMMUNOGLOBULIN-LIKE TRANSCRIPT 5 PROTEIN - Homo sapiens (Human), 631 aa.	1..632 1..631	628/632 (99%) 629/632 (99%)	0.0
AAC51901	IMMUNOGLOBULIN-LIKE TRANSCRIPT 5 - Homo sapiens (Human), 632 aa.	1..632 1..632	623/632 (98%) 628/632 (98%)	0.0
AAC51896	IMMUNOGLOBULIN-LIKE TRANSCRIPT 5 PROTEIN - Homo sapiens (Human), 632 aa.	1..632 1..632	620/632 (98%) 626/632 (98%)	0.0

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

Table 27E. Domain Analysis of NOV27a

Pfam Domain	NOV27a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ig: domain 1 of 4	42..100	12/63 (19%) 44/63 (70%)	0.00012
ig: domain 2 of 4	137..198	9/66 (14%) 41/66 (62%)	1.1e+02
ig: domain 3 of 4	238..298	12/65 (18%) 47/65 (72%)	7.7e-07
ig: domain 4 of 4	338..398	13/65 (20%) 39/65 (60%)	0.0043

Example B: Identification of NOVX clones

The novel NOVX target sequences identified in the present invention may have been subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on *in silico* predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, and uterus.

Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product derived from exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

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

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

1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on metabolic diseases), AI_comprehensive_panel (containing normal tissue and samples from autoimmune diseases), Panel CNSD.01 (containing central nervous system samples from normal and diseased brains) and CNS_neurodegeneration_panel (containing samples from normal and Alzheimer's diseased brains).

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

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

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

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer

concentration = 250 nM, primer melting temperature (Tm) range = 58 °-60 °C, primer optimal Tm = 59 °C, maximum primer difference = 2 °C, probe does not have 5'G, probe Tm must be 10 °C greater than primer Tm, amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

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

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

Panels 1, 1.1, 1.2, and 1.3D

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of

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

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

ca. = carcinoma,
* = established from metastasis,
met = metastasis,
s cell var = small cell variant,
non-s = non-sm = non-small,
squam = squamous,
pl. eff = pl effusion = pleural effusion,
glio = glioma,
astro = astrocytoma, and
neuro = neuroblastoma.

General_screening_panel_v1.4

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

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

Panels 2D and 2.2

The plates for Panels 2D and 2.2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (*i.e.* immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, *etc.*). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

Panel 3D

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are

two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

Panels 4D, 4R, and 4.1D

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

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

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5}M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2µg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium

pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5µg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately $2x10^6$cells/ml in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol ($5.5x10^{-5}$M) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10µg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and plated at 10^6cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5µg/ml anti-CD28 (Pharmingen) and 3ug/ml anti-CD3 (OKT3,

ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

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

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

were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at $5x10^5$cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to $5x10^5$cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1µg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

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

AI_comprehensive panel_v1.0

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

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

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

Surgical specimens of diseased colon from patients with ulcerative colitis and Crohns disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three female and three male Crohn's patients between the ages of 41-69 were used. Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female patients. Four of the patients were taking lebvid and two were on phenobarbital.

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

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

AI = Autoimmunity
Syn = Synovial
Normal = No apparent disease
Rep22 /Rep20 = individual patients
RA = Rheumatoid arthritis
Backus = From Backus Hospital
OA = Osteoarthritis
(SS) (BA) (MF) = Individual patients
Adj = Adjacent tissue
Match control = adjacent tissues
-M = Male
-F = Female
COPD = Chronic obstructive pulmonary disease

Panels 5D and 5I

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

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

Patient 2	Diabetic Hispanic, overweight, not on insulin
Patient 7-9	Nondiabetic Caucasian and obese (BMI>30)
Patient 10	Diabetic Hispanic, overweight, on insulin
Patient 11	Nondiabetic African American and overweight
Patient 12	Diabetic Hispanic on insulin

Adipocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem

cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose
Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated
Donor 2 and 3 AD: Adipose, Adipose Differentiated

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

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

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

GO Adipose = Greater Omentum Adipose
SK = Skeletal Muscle
UT = Uterus
PL = Placenta
AD = Adipose Differentiated
AM = Adipose Midway Differentiated
U = Undifferentiated Stem Cells

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

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

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

PSP = Progressive supranuclear palsy
Sub Nigra = Substantia nigra
Glob Palladus= Globus palladus
Temp Pole = Temporal pole
Cing Gyr = Cingulate gyrus
BA 4 = Brodman Area 4

Panel CNS_Neurodegeneration_V1.0

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

Disease diagnoses are taken from patient records. The panel contains six brains from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no evidence of dementia prior to death. The eight normal control brains are divided into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically

senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the following regions are represented: hippocampus, temporal cortex (Brodman Area 21), parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus; the parietal cortex shows moderate neuronal death in the late stages of the disease; the occipital cortex is spared in AD and therefore acts as a "control" region within AD patients. Not all brain regions are represented in all cases.

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

AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology upon autopsy
Control = Control brains; patient not demented, showing no neuropathology
Control (Path) = Control brains; pateint not demented but showing sever AD-like pathology
SupTemporal Ctx = Superior Temporal Cortex
Inf Temporal Ctx = Inferior Temporal Cortex

A. NOV1a and NOV1b (CG56258-01 and CG56258-02: sodium/calcium exchanger)

Expression of gene CG56258-021 and CG56258-02 was assessed using the primer-probe sets Ag2903, Ag5035 and Ag6163, described in Tables AA, AB and AC. Results of the RTQ-PCR runs are shown in Tables AD, AE, AF, AG, AH and AI.

Table AA. Probe Name Ag2903

Primers	Sequences	Length	Start Position	SEQ ID No:
Forward	5'-gactcgcaagatcaagcatcta-3'	22	641	93
Probe	TET-5'-cttcttcatcaccgctgcttggagta-3'-TAMRA	26	668	94
Reverse	5'-tagagccagatgtaggcaaaga-3'	22	694	95

Table AB. Probe Name Ag5035

Primers	Sequences	Length	Start Position	SEQ ID No:
Forward	5'-gaaagccagtattgggtgaac-3'	21	2023	96
Probe	TET-5'-ccccaaactagaagtcatcattgaaga-3'-TAMRA	27	2045	97
Reverse	5'-tttgtccaccgtagtcttgaac-3'	22	2081	98

Table AC. Probe Name Ag6163

Primers	Sequences	Length	Start Position	SEQ ID No:
Forward	5'-ggggagttggaattcaagaat-3'	21	1815	99
Probe	TET-5'-tgaaactgtcaaaacaattcacatcaag-3'-TAMRA	28	1838	100
Reverse	5'-tctcatatgcctcatcatcaattac-3'	25	1866	101

Table AD. AI_comprehensive panel_v1.0

Tissue Name	Rel. Exp.(%) Ag2903, Run 225410015	Rel. Exp.(%) Ag5035, Run 244570389	Tissue Name	Rel. Exp.(%) Ag2903, Run 225410015	Rel. Exp.(%) Ag5035, Run 244570389
110967 COPD-F	0.6	0.3	112427 Match Control Psoriasis-F	4.1	4.8
110980 COPD-F	0.4	0.6	112418 Psoriasis-M	0.8	0.5
110968 COPD-M	0.9	1.2	112723 Match Control Psoriasis-M	0.1	0.0
110977 COPD-M	0.9	0.8	112419 Psoriasis-M	1.3	0.9
110989 Emphysema-F	0.6	0.0	112424 Match Control Psoriasis-M	1.1	0.5
110992 Emphysema-F	0.4	0.7	112420 Psoriasis-M	2.1	2.2
110993 Emphysema-F	1.1	1.4	112425 Match Control Psoriasis-M	2.9	5.8
110994 Emphysema-F	1.0	0.7	104689 (MF) OA Bone-Backus	68.3	44.4
110995 Emphysema-F	1.1	0.8	104690 (MF) Adj "Normal" Bone-Backus	8.8	5.2

110996 Emphysema-F	0.0	0.0	104691 (MF) OA Synovium-Backus	2.1	1.7
110997 Asthma-M	1.5	0.6	104692 (BA) OA Cartilage-Backus	3.3	3.6
111001 Asthma-F	1.8	1.5	104694 (BA) OA Bone-Backus	**100.0**	**100.0**
111002 Asthma-F	1.9	1.9	104695 (BA) Adj "Normal" Bone-Backus	36.3	28.7
111003 Atopic Asthma-F	2.7	2.4	104696 (BA) OA Synovium-Backus	1.4	0.5
111004 Atopic Asthma-F	0.9	1.2	104700 (SS) OA Bone-Backus	54.0	37.6
111005 Atopic Asthma-F	0.9	1.1	104701 (SS) Adj "Normal" Bone-Backus	60.3	34.9
111006 Atopic Asthma-F	0.4	0.3	104702 (SS) OA Synovium-Backus	2.9	2.1
111417 Allergy-M	2.2	2.8	117093 OA Cartilage Rep7	1.4	0.4
112347 Allergy-M	0.6	0.0	112672 OA Bone5	4.8	3.3
112349 Normal Lung-F	0.9	0.0	112673 OA Synovium5	1.6	1.7
112357 Normal Lung-F	0.1	0.3	112674 OA Synovial Fluid cells5	2.6	2.6
112354 Normal Lung-M	0.0	0.3	117100 OA Cartilage Rep14	0.0	0.0
112374 Crohns-F	0.2	0.0	112756 OA Bone9	5.6	0.4
112389 Match Control Crohns-F	0.2	1.1	112757 OA Synovium9	32.3	37.4
112375 Crohns-F	0.0	0.0	112758 OA Synovial Fluid Cells9	1.1	0.6
112732 Match Control	0.8	0.8	117125 RA Cartilage Rep2	2.8	1.1

Crohns-F					
112725 Crohns-M	0.1	0.0	113492 Bone2 RA	3.0	1.2
112387 Match Control Crohns-M	1.6	1.1	113493 Synovium2 RA	1.6	0.8
112378 Crohns-M	0.6	0.0	113494 Syn Fluid Cells RA	1.8	0.8
112390 Match Control Crohns-M	1.0	0.8	113499 Cartilage4 RA	1.7	1.9
112726 Crohns-M	0.8	0.7	113500 Bone4 RA	2.6	2.2
112731 Match Control Crohns-M	0.9	0.3	113501 Synovium4 RA	2.0	0.7
112380 Ulcer Col-F	0.4	0.5	113502 Syn Fluid Cells4 RA	0.6	0.6
112734 Match Control Ulcer Col-F	3.5	1.8	113495 Cartilage3 RA	1.6	0.6
112384 Ulcer Col-F	2.7	1.9	113496 Bone3 RA	1.9	0.7
112737 Match Control Ulcer Col-F	0.5	0.6	113497 Synovium3 RA	1.4	0.8
112386 Ulcer Col-F	2.0	1.4	113498 Syn Fluid Cells3 RA	2.6	2.7
112738 Match Control Ulcer Col-F	0.1	0.3	117106 Normal Cartilage Rep20	0.4	0.0
112381 Ulcer Col-M	1.3	0.0	113663 Bone3 Normal	0.6	0.0
112735 Match Control Ulcer Col-M	3.3	1.2	113664 Synovium3 Normal	0.2	0.0
112382 Ulcer Col-M	1.2	0.6	113665 Syn Fluid Cells3 Normal	0.2	0.0
112394 Match Control Ulcer Col-M	0.9	0.8	117107 Normal Cartilage Rep22	2.9	0.4

112383 Ulcer Col-M	0.7	0.0	113667 Bone4 Normal	1.1	0.0
112736 Match Control Ulcer Col-M	0.7	0.3	113668 Synovium4 Normal	1.3	0.5
112423 Psoriasis-F	0.7	0.3	113669 Syn Fluid Cells4 Normal	1.1	0.5

Table AE. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2903, Run 209735156	Rel. Exp.(%) Ag5035, Run 224062761	Tissue Name	Rel. Exp.(%) Ag2903, Run 209735156	Rel. Exp.(%) Ag5035, Run 224062761
AD 1 Hippo	9.1	8.1	Control (Path) 3 Temporal Ctx	3.9	3.2
AD 2 Hippo	24.1	35.6	Control (Path) 4 Temporal Ctx	32.8	89.5
AD 3 Hippo	8.2	4.5	AD 1 Occipital Ctx	15.4	8.8
AD 4 Hippo	7.7	9.7	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 Hippo	**100.0**	84.1	AD 3 Occipital Ctx	5.5	1.9
AD 6 Hippo	22.2	20.3	AD 4 Occipital Ctx	18.8	16.3
Control 2 Hippo	31.0	50.3	AD 5 Occipital Ctx	56.6	63.7
Control 4 Hippo	9.1	10.7	AD 6 Occipital Ctx	20.6	13.6
Control (Path) 3 Hippo	5.5	1.4	Control 1 Occipital Ctx	1.7	2.7
AD 1 Temporal Ctx	7.7	7.5	Control 2 Occipital Ctx	63.3	76.3

AD 2 Temporal Ctx	21.6	34.4	Control 3 Occipital Ctx	24.7	14.3
AD 3 Temporal Ctx	4.8	3.8	Control 4 Occipital Ctx	5.6	5.0
AD 4 Temporal Ctx	18.2	24.7	Control (Path) 1 Occipital Ctx	89.5	**100.0**
AD 5 Inf Temporal Ctx	82.9	97.3	Control (Path) 2 Occipital Ctx	15.6	8.8
AD 5 Sup Temporal Ctx	32.1	31.4	Control (Path) 3 Occipital Ctx	1.0	0.9
AD 6 Inf Temporal Ctx	26.6	21.2	Control (Path) 4 Occipital Ctx	19.2	21.0
AD 6 Sup Temporal Ctx	29.5	18.9	Control 1 Parietal Ctx	6.9	5.4
Control 1 Temporal Ctx	4.2	3.1	Control 2 Parietal Ctx	27.0	26.6
Control 2 Temporal Ctx	51.1	50.3	Control 3 Parietal Ctx	22.7	12.3
Control 3 Temporal Ctx	23.8	15.3	Control (Path) 1 Parietal Ctx	**100.0**	87.7
Control 3 Temporal Ctx	6.4	5.6	Control (Path) 2 Parietal Ctx	29.5	19.2
Control (Path) 1 Temporal Ctx	68.3	73.7	Control (Path) 3 Parietal Ctx	3.8	1.3
Control (Path) 2 Temporal Ctx	49.7	27.9	Control (Path) 4 Parietal Ctx	61.6	46.0

Table AF. General_screening_panel_v1.5

Tissue Name	Rel. Exp.(%) Ag5035, Run 228967202	Rel. Exp.(%) Ag5035, Run 244373096	Tissue Name	Rel. Exp.(%) Ag5035, Run 228967202	Rel. Exp.(%) Ag5035, Run 244373096
Adipose	1.6	2.2	Renal ca. TK-10	0.0	0.0
Melanoma* Hs688(A).T	0.0	0.0	Bladder	0.6	0.3
Melanoma* Hs688(B).T	0.0	0.0	Gastric ca. (liver met.) NCI-N87	0.0	0.0
Melanoma* M14	0.0	0.0	Gastric ca. KATO III	0.0	0.0
Melanoma* LOXIMVI	0.0	0.0	Colon ca. SW-948	0.0	0.0
Melanoma* SK-MEL-5	0.0	0.0	Colon ca. SW480	0.0	0.0
Squamous cell carcinoma SCC-4	0.0	0.0	Colon ca.* (SW480 met) SW620	0.0	0.0
Testis Pool	0.0	0.1	Colon ca. HT29	0.0	0.0
Prostate ca.* (bone met) PC-3	0.0	0.0	Colon ca. HCT-116	0.0	0.0
Prostate Pool	1.4	2.0	Colon ca. CaCo-2	0.2	0.2
Placenta	0.3	0.1	Colon cancer tissue	0.7	0.1
Uterus Pool	2.1	1.6	Colon ca. SW1116	0.0	0.0
Ovarian ca. OVCAR-3	0.0	0.0	Colon ca. Colo-205	0.0	0.0
Ovarian ca. SK-OV-3	0.0	0.0	Colon ca. SW-48	0.0	0.0
Ovarian ca. OVCAR-4	0.3	0.0	Colon Pool	3.5	3.1
Ovarian ca. OVCAR-5	0.0	0.0	Small Intestine Pool	1.1	1.3
Ovarian ca. IGROV-1	0.0	0.0	Stomach Pool	0.2	1.4
Ovarian ca. OVCAR-8	0.0	0.0	Bone Marrow Pool	2.3	1.8
Ovary	0.3	0.3	Fetal Heart	0.3	0.6
Breast ca. MCF-7	0.0	0.0	Heart Pool	0.8	0.0
Breast ca. MDA-MB-	0.0	0.0	Lymph Node Pool	2.6	2.0

231					
Breast ca. BT 549	0.0	0.0	Fetal Skeletal Muscle	17.9	22.2
Breast ca. T47D	0.0	0.0	Skeletal Muscle Pool	**100.0**	83.5
Breast ca. MDA-N	0.0	0.0	Spleen Pool	0.6	0.0
Breast Pool	2.6	3.5	Thymus Pool	0.6	0.4
Trachea	0.8	1.1	CNS cancer (glio/astro) U87-MG	0.0	0.0
Lung	0.0	0.0	CNS cancer (glio/astro) U-118-MG	0.2	0.0
Fetal Lung	14.3	14.5	CNS cancer (neuro;met) SK-N-AS	0.0	0.0
Lung ca. NCI-N417	0.0	0.0	CNS cancer (astro) SF-539	0.0	0.0
Lung ca. LX-1	0.0	0.0	CNS cancer (astro) SNB-75	0.0	0.2
Lung ca. NCI-H146	0.0	0.0	CNS cancer (glio) SNB-19	0.0	0.0
Lung ca. SHP-77	9.3	12.7	CNS cancer (glio) SF-295	0.0	0.0
Lung ca. A549	0.0	0.0	Brain (Amygdala) Pool	32.8	31.0
Lung ca. NCI-H526	0.2	0.2	Brain (cerebellum)	69.7	76.3
Lung ca. NCI-H23	0.0	0.0	Brain (fetal)	90.1	**100.0**
Lung ca. NCI-H460	0.0	0.0	Brain (Hippocampus) Pool	27.9	31.0
Lung ca. HOP-62	0.0	0.0	Cerebral Cortex Pool	36.3	48.3
Lung ca. NCI-H522	0.0	0.0	Brain (Substantia nigra) Pool	31.0	32.1
Liver	0.0	0.0	Brain (Thalamus) Pool	50.0	50.3
Fetal Liver	2.1	2.0	Brain (whole)	46.0	38.2
Liver ca. HepG2	0.0	2.0	Spinal Cord Pool	17.6	18.4
Kidney Pool	1.8	0.0	Adrenal Gland	2.5	2.5
Fetal Kidney	0.8	0.7	Pituitary gland	1.3	1.2

			Pool		
Renal ca. 786-0	0.0	0.0	Salivary Gland	0.0	0.2
Renal ca. A498	0.0	0.0	Thyroid (female)	0.0	0.0
Renal ca. ACHN	0.0	0.0	Pancreatic ca. CAPAN2	0.0	0.0
Renal ca. UO-31	0.0	0.0	Pancreas Pool	1.6	7.9

Table AG. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2903, Run 162556420	Tissue Name	Rel. Exp.(%) Ag2903, Run 162556420
Liver adenocarcinoma	0.0	Kidney (fetal)	0.3
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.1	Renal ca. RXF 393	0.0
Thyroid	0.2	Renal ca. ACHN	0.0
Salivary gland	0.1	Renal ca. UO-31	0.1
Pituitary gland	0.4	Renal ca. TK-10	0.0
Brain (fetal)	2.0	Liver	0.0
Brain (whole)	3.9	Liver (fetal)	0.4
Brain (amygdala)	3.7	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	3.3	Lung	0.0
Brain (hippocampus)	5.6	Lung (fetal)	0.4
Brain (substantia nigra)	0.9	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	5.9	Lung ca. (small cell) NCI-H69	0.2
Cerebral Cortex	80.7	Lung ca. (s.cell var.) SHP-77	3.4
Spinal cord	1.7	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.1
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.6	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl)	0.0

		NCI-H522	
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.2
glioma SNB-19	0.1	Mammary gland	0.1
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	5.2	Breast ca.* (pl.ef) T47D	0.0
Heart	0.3	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	**100.0**	Breast ca. MDA-N	0.0
Skeletal muscle	21.2	Ovary	0.2
Bone marrow	0.2	Ovarian ca. OVCAR-3	0.0
Thymus	0.6	Ovarian ca. OVCAR-4	0.3
Spleen	0.0	Ovarian ca. OVCAR-5	0.0
Lymph node	0.3	Ovarian ca. OVCAR-8	0.0
Colorectal	1.1	Ovarian ca. IGROV-1	0.0
Stomach	0.1	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.2	Uterus	0.1
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	0.3
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.5	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.2
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.2	Melanoma LOX IMVI	0.0

Trachea	0.3	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	0.3

Table AH. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2903, Run 162345106	Tissue Name	Rel. Exp.(%) Ag2903, Run 162345106
Normal Colon	8.1	Kidney Margin 8120608	0.5
CC Well to Mod Diff (ODO3866)	0.3	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	0.3	Kidney Margin 8120614	0.1
CC Gr.2 rectosigmoid (ODO3868)	0.1	Kidney Cancer 9010320	0.5
CC Margin (ODO3868)	0.2	Kidney Margin 9010321	0.0
CC Mod Diff (ODO3920)	0.3	Normal Uterus	1.0
CC Margin (ODO3920)	0.4	Uterus Cancer 064011	0.5
CC Gr.2 ascend colon (ODO3921)	1.1	Normal Thyroid	1.0
CC Margin (ODO3921)	0.9	Thyroid Cancer 064010	0.0
CC from Partial Hepatectomy (ODO4309) Mets	0.4	Thyroid Cancer A302152	0.0
Liver Margin (ODO4309)	0.1	Thyroid Margin A302153	0.1
Colon mets to lung (OD04451-01)	0.1	Normal Breast	3.0
Lung Margin (OD04451-02)	1.3	Breast Cancer (OD04566)	0.4
Normal Prostate 6546-1	2.3	Breast Cancer (OD04590-01)	0.8
Prostate Cancer (OD04410)	1.2	Breast Cancer Mets (OD04590-03)	1.5
Prostate Margin (OD04410)	4.2	Breast Cancer Metastasis (OD04655-05)	0.2
Prostate Cancer (OD04720-01)	1.2	Breast Cancer 064006	0.5

Prostate Margin (OD04720-02)	4.6	Breast Cancer 1024	1.2
Normal Lung 061010	5.8	Breast Cancer 9100266	1.8
Lung Met to Muscle (ODO4286)	0.0	Breast Margin 9100265	1.1
Muscle Margin (ODO4286)	**100.0**	Breast Cancer A209073	0.6
Lung Malignant Cancer (OD03126)	0.8	Breast Margin A209073	0.0
Lung Margin (OD03126)	7.7	Normal Liver	0.0
Lung Cancer (OD04404)	1.4	Liver Cancer 064003	0.0
Lung Margin (OD04404)	4.1	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	0.3	Liver Cancer 1026	0.0
Lung Margin (OD04565)	1.2	Liver Cancer 6004-T	0.1
Lung Cancer (OD04237-01)	0.7	Liver Tissue 6004-N	0.5
Lung Margin (OD04237-02)	2.4	Liver Cancer 6005-T	0.0
Ocular Mel Met to Liver (ODO4310)	0.0	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.1	Normal Bladder	0.6
Melanoma Mets to Lung (OD04321)	0.1	Bladder Cancer 1023	0.2
Lung Margin (OD04321)	2.4	Bladder Cancer A302173	1.4
Normal Kidney	1.5	Bladder Cancer (OD04718-01)	0.5
Kidney Ca, Nuclear grade 2 (OD04338)	0.5	Bladder Normal Adjacent (OD04718-03)	4.9
Kidney Margin (OD04338)	1.4	Normal Ovary	0.1
Kidney Ca Nuclear grade 1/2 (OD04339)	1.8	Ovarian Cancer 064008	1.9
Kidney Margin (OD04339)	0.6	Ovarian Cancer (OD04768-07)	0.0
Kidney Ca, Clear cell type (OD04340)	1.0	Ovary Margin (OD04768-08)	0.2
Kidney Margin (OD04340)	0.8	Normal Stomach	4.5
Kidney Ca, Nuclear grade 3 (OD04348)	1.3	Gastric Cancer 9060358	1.7
Kidney Margin	0.7	Stomach Margin	1.5

(OD04348)		9060359	
Kidney Cancer (OD04622-01)	0.7	Gastric Cancer 9060395	1.2
Kidney Margin (OD04622-03)	0.0	Stomach Margin 9060394	2.0
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	0.6
Kidney Margin (OD04450-03)	0.1	Stomach Margin 9060396	1.7
Kidney Cancer 8120607	0.1	Gastric Cancer 064005	2.9

Table AI. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5035, Run 223740981	Tissue Name	Rel. Exp.(%) Ag5035, Run 223740981
Secondary Th1 act	0.0	HUVEC IL-1beta	2.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	1.9
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	**100.0**
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	2.0	Microsvasular Dermal EC TNFalpha + IL-1beta	55.5
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	2.2
Primary Th2 rest	0.0	Small airway epithelium none	4.2
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	3.7
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	2.3
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8	0.0	KU-812 (Basophil) rest	10.6

lymphocyte act			
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	13.6
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	8.4	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	2.9	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	3.8	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	2.3	NCI-H292 IL-13	2.5
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	2.1	HPAEC none	0.0
Two Way MLR 5 day	2.3	HPAEC TNF alpha + IL-1 beta	8.8
Two Way MLR 7 day	1.8	Lung fibroblast none	0.0
PBMC rest	6.7	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	6.3	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	9.2	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	1.9	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	2.4	Dermal Fibroblasts rest	2.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	31.6	Colon	5.4
Macrophages rest	1.7	Lung	8.1
Macrophages LPS	6.5	Thymus	5.1
HUVEC none	0.0	Kidney	0.0

HUVEC starved	0.0		

AI_comprehensive panel_v1.0 Summary: Ag2903/Ag5035 Two experiments with two different probe and primer sets produce results that are in very good agreement. Expression of the CG56258-01 gene appears to be more highly associated with synovium and bone samples from patients with osteoarthritis when compared to expression in the control samples. Thus, therapeutic modulation of the expression or function of this gene may be effective in the treatment of osteoarthritis. A third experiment with the probe and primer set Ag6163 shows low/undetectable levels of expression (CTs>35).

CNS_neurodegeneration_v1.0 Summary: Ag2903/Ag5035 Two experiments with two different probes and primers produce results that are in excellent agreement. This panel does not show differential expression of the CG56258-01 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain, with highest expression in the hippocampus of an Alzheimer's patient and the occipital cortex of a control patient (CTs=28-30). Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

General_screening_panel_v1.5 Summary: Ag5035 Two experiments with the same probe and primer produce results that are in excellent agreement, with the CG56258-02 gene showing highly brain preferential expression (CTs=30-31). In addition, moderate levels of expression are seen in fetal and adult skeletal muscle (CTs=30-31). This expression profile is in excellent concordance with the results in Panel 1.3D. Please see Panel 1.3D for further discussion of utility of this gene in the central nervous system and metabolic disease.

Panel 1.3D Summary: Ag2903 Expression of the CG56258-01 gene is highest in fetal skeletal muscle (CT=26.8). In addition, significant levels of expression are also seen in adult skeletal muscle and fetal heart. Thus, expression of this gene could be used to differentiate skeletal muscle derived samples from other samples on this panel and as a marker of skeletal muscle. This gene encodes a putative sodium/calcium exchanger. Altered levels of intracellular calcium have been implicated in many diseases, including type 2 diabetes. Based on its expression profile and homology to a calcium transport protein, therapeutic modulation of the expression or function of this gene or gene product may be effective in the treatment of type 2 diabetes.

In addition, moderate to low levels of expression are seen in all regions of the CNS examined. Inhibition of calcium uptake has been shown to decrease neuronal death in response to cerebral ischemia. Therefore, this gene, a putative calcium transport protein, represents an excellent drug target for the treatment of stroke. Treatment with an antagonist immediately after stroke could decrease total infarct volume and lessen the overall stroke severity.

See, generally,

Balasubramanyam M, Balaji RA, Subashini B, Mohan V. Evidence for mechanistic alterations of Ca2+ homeostasis in Type 2 diabetes mellitus. Int J Exp Diabetes Res 2001;1(4):275-87. PMID: 11467418; and

Matsuda T, Arakawa N, Takuma K, Kishida Y, Kawasaki Y, Sakaue M, Takahashi K, Takahashi T, Suzuki T, Ota T, Hamano-Takahashi A, Onishi M, Tanaka Y, Kameo K, Baba A. SEA0400, a novel and selective inhibitor of the Na+-Ca2+ exchanger, attenuates reperfusion injury in the in vitro and in vivo cerebral ischemic models. J Pharmacol Exp Ther 2001 Jul;298(1):249-56.

Panel 2D Summary: Ag2903 The expression of the CG56258-01 gene in this panel is consistent with the profile seen in Panel 1.3D. Expression is highest and most prominent in a normal muscle sample (CT=28.7). Please see Panel 1.3D for discussion of utility of this gene in metabolic disease.

Panel 4.1D Summary: Ag5035 Expression of the CG56258-02 gene is restricted to TNF-alpha and IL-1 beta treated lung and dermal microvasculature (CTs=33-34). Endothelial cells are known to play important roles in inflammatory responses by altering the expression of surface proteins that are involved in activation and recruitment of effector inflammatory cells. The expression of this gene in dermal microvascular endothelial cells suggests that this protein product may be involved in inflammatory responses to skin disorders, including psoriasis. Expression in lung microvascular endothelial cells suggests that the protein encoded by this transcript may also be involved in lung disorders including asthma, allergies, chronic obstructive pulmonary disease, and emphysema. Therefore, therapeutic modulation of the protein encoded by this gene may lead to amelioration of symptoms associated with psoriasis, asthma, allergies, chronic obstructive pulmonary disease, and emphysema.

Ag5035 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run.

Ag6163 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

B. NOV2a (CG59843-01: fibropellin III-like)

Expression of gene CG59843-01 was assessed using the primer-probe sets Ag2797, Ag3606 and Ag221, described in Tables BA, BB and BC. Results of the RTQ-PCR runs are shown in Tables BD, BE, BF, BG, BH, BI, BJ, BK and BL.

Table BA. Probe Name Ag2797

Primers	Sequences	Length	Start Position	SEQ ID No:
Forward	5'-cagctacaaatgcctctgtgat-3'	22	1488	102
Probe	TET-5'-ccaggttaccatggcctctactgtga-3'-TAMRA	26	1510	103
Reverse	5'-agcggagaggcactcattatat-3'	22	1542	104

Table BB. Probe Name Ag3606

Primers	Sequences	Length	Start Position	SEQ ID No:
Forward	5'-cagctacaaatgcctctgtgat-3'	22	1488	105
Probe	TET-5'-ccaggttaccatggcctctactgtga-3'-TAMRA	26	1510	106
Reverse	5'-agcggagaggcactcattatat-3'	22	1542	107

Table BC. Probe Name Ag221

Primers	Sequences	Length	Start Position	SEQ ID No:
Forward	5'-ctgccaggtaggcagtgtca-3'	20	545	108
Probe	TET-5'-aaaatcctgcctcgctctcaggcaa-3'-TAMRA	25	517	109
Reverse	5'-gcctgttcctgctactcagga-3'	21	489	110

Table BD. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2797, Run 208699245	Rel. Exp.(%) Ag3606, Run 210997602	Tissue Name	Rel. Exp.(%) Ag2797, Run 208699245	Rel. Exp.(%) Ag3606, Run 210997602
AD 1 Hippo	8.0	14.9	Control	2.7	3.4

			(Path) 3 Temporal Ctx		
AD 2 Hippo	19.8	40.9	Control (Path) 4 Temporal Ctx	12.0	21.5
AD 3 Hippo	7.3	11.1	AD 1 Occipital Ctx	9.5	16.3
AD 4 Hippo	7.0	12.1	AD 2 Occipital Ctx (Missing)	92.7	0.0
AD 5 hippo	97.3	61.1	AD 3 Occipital Ctx	5.6	0.1
AD 6 Hippo	31.9	88.9	AD 4 Occipital Ctx	15.1	27.0
Control 2 Hippo	14.6	47.0	AD 5 Occipital Ctx	97.3	25.5
Control 4 Hippo	12.4	19.8	AD 6 Occipital Ctx	99.3	57.8
Control (Path) 3 Hippo	5.1	12.3	Control 1 Occipital Ctx	2.6	3.4
AD 1 Temporal Ctx	10.7	16.7	Control 2 Occipital Ctx	51.8	73.2
AD 2 Temporal Ctx	19.5	41.8	Control 3 Occipital Ctx	0.2	15.3
AD 3 Temporal Ctx	3.2	6.7	Control 4 Occipital Ctx	6.4	13.1
AD 4 Temporal Ctx	14.7	25.0	Control (Path) 1 Occipital Ctx	63.7	97.3
AD 5 Inf Temporal Ctx	**100.0**	**100.0**	Control (Path) 2 Occipital Ctx	7.9	10.9
AD 5	**100.0**	51.4	Control	4.1	2.8

SupTemporal Ctx			(Path) 3 Occipital Ctx		
AD 6 Inf Temporal Ctx	36.1	65.1	Control (Path) 4 Occipital Ctx	4.8	9.7
AD 6 Sup Temporal Ctx	26.2	50.0	Control 1 Parietal Ctx	7.3	7.6
Control 1 Temporal Ctx	6.1	6.9	Control 2 Parietal Ctx	95.9	36.6
Control 2 Temporal Ctx	25.2	54.0	Control 3 Parietal Ctx	10.4	19.8
Control 3 Temporal Ctx	6.0	14.9	Control (Path) 1 Parietal Ctx	34.9	70.2
Control 4 Temporal Ctx	4.3	8.9	Control (Path) 2 Parietal Ctx	12.9	22.7
Control (Path) 1 Temporal Ctx	33.4	50.3	Control (Path) 3 Parietal Ctx	2.2	3.5
Control (Path) 2 Temporal Ctx	12.6	27.0	Control (Path) 4 Parietal Ctx	16.7	28.5

Table BE. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3606, Run 217675868	Tissue Name	Rel. Exp.(%) Ag3606, Run 217675868
Adipose	0.0	Renal ca. TK-10	0.1
Melanoma* Hs688(A).T	0.1	Bladder	0.1
Melanoma* Hs688(B).T	2.9	Gastric ca. (liver met.) NCI-N87	0.3
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	10.4	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	97.9	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	1.4	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.4	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.1	Colon ca. HCT-116	0.6
Prostate Pool	0.3	Colon ca. CaCo-2	0.0

Placenta	0.1	Colon cancer tissue	0.3
Uterus Pool	0.3	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.1	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.2	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.2
Ovarian ca. OVCAR-5	0.7	Small Intestine Pool	0.3
Ovarian ca. IGROV-1	1.2	Stomach Pool	0.4
Ovarian ca. OVCAR-8	0.6	Bone Marrow Pool	0.2
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.2
Breast ca. MDA-MB-231	0.7	Lymph Node Pool	0.3
Breast ca. BT 549	2.9	Fetal Skeletal Muscle	0.0
Breast ca. T47D	1.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	5.4	Spleen Pool	0.8
Breast Pool	0.2	Thymus Pool	0.1
Trachea	1.3	CNS cancer (glio/astro) U87-MG	**100.0**
Lung	0.0	CNS cancer (glio/astro) U-118-MG	42.0
Fetal Lung	0.4	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.3
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	1.2
Lung ca. NCI-H146	12.0	CNS cancer (glio) SNB-19	1.0
Lung ca. SHP-77	40.6	CNS cancer (glio) SF-295	23.7
Lung ca. A549	0.1	Brain (Amygdala) Pool	18.2
Lung ca. NCI-H526	0.0	Brain (cerebellum)	77.4
Lung ca. NCI-H23	2.2	Brain (fetal)	32.1
Lung ca. NCI-H460	34.2	Brain (Hippocampus) Pool	19.2
Lung ca. HOP-62	2.1	Cerebral Cortex Pool	20.2
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	24.7

Liver	0.0	Brain (Thalamus) Pool	25.5
Fetal Liver	0.0	Brain (whole)	26.8
Liver ca. HepG2	0.0	Spinal Cord Pool	24.0
Kidney Pool	1.1	Adrenal Gland	26.4
Fetal Kidney	4.1	Pituitary gland Pool	5.0
Renal ca. 786-0	0.0	Salivary Gland	10.0
Renal ca. A498	0.1	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.3
Renal ca. UO-31	0.3	Pancreas Pool	0.7

Table BF. Panel 1

Tissue Name	Rel. Exp.(%) Ag221, Run 87987754	Tissue Name	Rel. Exp.(%) Ag221, Run 87987754
Endothelial cells	0.0	Renal ca. 786-0	0.0
Endothelial cells (treated)	0.0	Renal ca. A498	0.0
Pancreas	0.0	Renal ca. RXF 393	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	0.0
Adrenal gland	22.8	Renal ca. UO-31	0.0
Thyroid	0.0	Renal ca. TK-10	0.0
Salivary gland	6.7	Liver	0.0
Pituitary gland	0.0	Liver (fetal)	0.0
Brain (fetal)	15.3	Liver ca. (hepatoblast) HepG2	0.0
Brain (whole)	42.6	Lung	0.0
Brain (amygdala)	19.9	Lung (fetal)	0.0
Brain (cerebellum)	**100.0**	Lung ca. (small cell) LX-1	0.0
Brain (hippocampus)	17.9	Lung ca. (small cell) NCI-H69	0.0
Brain (substantia nigra)	40.9	Lung ca. (s.cell var.) SHP-77	0.0
Brain (thalamus)	39.5	Lung ca. (large cell)NCI-H460	0.0
Brain (hypothalamus)	7.9	Lung ca. (non-sm. cell) A549	0.0
Spinal cord	13.6	Lung ca. (non-s.cell) NCI-H23	0.6
glio/astro U87-MG	21.0	Lung ca. (non-s.cell) HOP-62	0.6
glio/astro U-118-MG	10.4	Lung ca. (non-s.cl)	0.0

		NCI-H522	
astrocytoma SW1783	1.8	Lung ca. (squam.) SW 900	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (squam.) NCI-H596	0.0
astrocytoma SF-539	0.0	Mammary gland	0.0
astrocytoma SNB-75	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SNB-19	0.3	Breast ca.* (pl.ef) MDA-MB-231	0.0
glioma U251	0.0	Breast ca.* (pl. ef) T47D	0.0
glioma SF-295	3.5	Breast ca. BT-549	0.0
Heart	0.0	Breast ca. MDA-N	2.8
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	0.0	Ovarian ca. OVCAR-5	0.0
Lymph node	0.0	Ovarian ca. OVCAR-8	0.2
Colon (ascending)	0.2	Ovarian ca. IGROV-1	0.0
Stomach	1.7	Ovarian ca. (ascites) SK-OV-3	0.0
Small intestine	0.3	Uterus	1.5
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620 (SW480 met)	0.0	Prostate	0.9
Colon ca. HT29	0.0	Prostate ca.* (bone met) PC-3	0.0
Colon ca. HCT-116	0.0	Testis	0.2
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. HCT-15	0.0	Melanoma* (met) Hs688(B).T	0.2
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca. * (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.2	Melanoma LOX IMVI	1.7
Trachea	3.1	Melanoma* (met) SK-MEL-5	21.3

Kidney	1.5	Melanoma SK-MEL-28	0.0
Kidney (fetal)	5.9		

Table BG. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2797, Run 165643064	Rel. Exp.(%) Ag2797, Run 165693893	Tissue Name	Rel. Exp.(%) Ag2797, Run 165643064	Rel. Exp.(%) Ag2797, Run 165693893
Liver adenocarcinoma	0.5	0.9	Kidney (fetal)	2.1	4.5
Pancreas	0.0	0.0	Renal ca. 786-0	0.0	0.0
Pancreatic ca. CAPAN 2	0.3	0.2	Renal ca. A498	0.5	0.6
Adrenal gland	11.7	15.6	Renal ca. RXF 393	0.1	0.0
Thyroid	0.0	0.1	Renal ca. ACHN	0.0	0.0
Salivary gland	6.7	9.3	Renal ca. UO-31	3.5	0.4
Pituitary gland	13.7	12.2	Renal ca. TK-10	0.1	0.1
Brain (fetal)	27.9	31.2	Liver	0.0	0.0
Brain (whole)	59.0	66.9	Liver (fetal)	0.1	0.0
Brain (amygdala)	33.4	35.4	Liver ca. (hepatoblast) HepG2	0.0	0.1
Brain (cerebellum)	71.2	82.4	Lung	0.1	0.3
Brain (hippocampus)	29.7	37.4	Lung (fetal)	0.1	0.0
Brain (substantia nigra)	41.8	52.9	Lung ca. (small cell) LX-1	0.0	0.0
Brain (thalamus)	**100.0**	**100.0**	Lung ca. (small cell) NCI-H69	0.0	0.0
Cerebral Cortex	26.6	28.7	Lung ca. (s.cell var.) SHP-77	14.3	15.3
Spinal cord	27.7	38.2	Lung ca. (large cell)NCI-H460	28.3	29.1
glio/astro U87-MG	16.5	21.2	Lung ca. (non-sm. cell) A549	0.1	0.1

DEMANDE OU BREVET VOLUMINEUX

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

CECI EST LE TOME	1	DE	2
CONTENANT LES PAGES	1	À	229

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

JUMBO APPLICATIONS/PATENTS

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

THIS IS VOLUME	1	OF	2
CONTAINING PAGES	1	TO	229

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

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

CLAIMS

We claim:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46;

b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed;

c) the amino acid sequence selected from the group consisting of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46;

d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46, wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and

e) a fragment of any of a) through d).

2. The polypeptide of claim 1 that is a naturally occurring allelic variant of the sequence selected from the group consisting of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46.

3. The polypeptide of claim 2, wherein said allelic variant comprises an amino acid sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46.

4. The polypeptide of claim 1 that is a variant polypeptide described therein, wherein any amino acid specified in the chosen sequence is changed to provide a conservative substitution.

5. A pharmaceutical composition comprising the polypeptide of claim 1 and a pharmaceutically acceptable carrier.

6. A kit comprising in one or more containers, the pharmaceutical composition of claim 5.

7. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein said therapeutic is the polypeptide of claim 1.

8. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:
   (a) providing said sample;
   (b) introducing said sample to an antibody that binds immunospecifically to the polypeptide; and
   (c) determining the presence or amount of antibody bound to said polypeptide, thereby determining the presence or amount of polypeptide in said sample.

9. A method for determining the presence of or predisposition to a disease associated with altered levels of the polypeptide of claim 1 in a first mammalian subject, the method comprising:
   a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and
   b) comparing the amount of said polypeptide in the sample of step (a) to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,
   wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

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

11. The method of claim 10, wherein said subject is a human.

12. 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 of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46;

b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46, 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:2*n*, wherein *n* is an integer between 1-46;

d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO:2*n*, wherein *n* is an integer between 1-46, 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:2*n*, wherein *n* is an integer between 1-46, 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.

13. The nucleic acid molecule of claim 12, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant.

14. The nucleic acid molecule of claim 12 that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.

15. The nucleic acid molecule of claim 12, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46.

16. The nucleic acid molecule of claim 12, 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:2*n*-1, wherein *n* is an integer between 1-46;

b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, 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:2*n*-1, wherein *n* is an integer between 1-46; and

d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, 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.

17. The nucleic acid molecule of claim 12, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-46, or a complement of said nucleotide sequence.

18. The nucleic acid molecule of claim 12, wherein the nucleic acid molecule comprises a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen nucleotide sequence is changed from that selected from the group consisting

of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides in the chosen coding sequence are so changed, an isolated second polynucleotide that is a complement of the first polynucleotide, or a fragment of any of them.

19. A vector comprising the nucleic acid molecule of claim 12.

20. The vector of claim 19, further comprising a promoter operably linked to said nucleic acid molecule.

21. A cell comprising the vector of claim 20.

22. A method for determining the presence or amount of the nucleic acid molecule of claim 12 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.

23. The method of claim 22 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

24. The method of claim 23 wherein the cell or tissue type is cancerous.

25. A method for determining the presence of or predisposition to a disease associated with altered levels of the nucleic acid molecule of claim 12 in a first mammalian subject, the method comprising:

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

b) comparing the amount of said nucleic acid in the sample of step (a) to the amount of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

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

26. An antibody that binds immunospecifically to the polypeptide of claim 1.

27. The antibody of claim 26, wherein said antibody is a monoclonal antibody.

28. The antibody of claim 26, wherein the antibody is a humanized antibody.

29. The antibody of claim 26, wherein the antibody is a fully human antibody

30. The antibody of claim 26, wherein the dissociation constant for the binding of the polypeptide to the antibody is less than 1×10^{-9} M.

31. The antibody of claim 26, wherein the antibody neutralizes an activity of the polypeptide.

32. A pharmaceutical composition comprising the antibody of claim 26 and a pharmaceutically acceptable carrier.

33. A kit comprising in one or more containers, the pharmaceutical composition of claim 29.

34. 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 said therapeutic is a NOVX antibody.

35. A method of treating or preventing a NOVX-associated disorder, said method comprising administering to a subject in which such treatmnet or prevention is desired the antibody of claim 26 in an amount sufficient to treat or prevent said NOVx-associated disorder in said subject.

36. A method of treating a pathological state in a mammal, the method comprising administering to the mammal the antibody of claim 26 in an amount sufficient to alleviate the pathological state.

37. A method of treating or preventing a pathology associated with the polypeptide of claim 1, said method comprising administering to a subject in which such treatment or prevention is desired a NOVX antibody in an amount sufficient to treat or prevent said pathology in said subject.

38. The method of claim 37, wherein the subject is a human.