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

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2443770 A1 2002/10/31

(21) 2 443 770

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2002/04/11
(87) Date publication PCT/PCT Publication Date: 2002/10/31
(85) Entrée phase nationale/National Entry: 2003/10/15
(86) N° demande PCT/PCT Application No.: US 2002/011634
(87) N° publication PCT/PCT Publication No.: 2002/085922
(30) Priorités/Priorities: 2001/04/23 (60/285,748) US;
2001/04/24 (60/286,068) US; 2001/04/25 (60/286,292) US;
2001/05/03 (60/288,334) US; 2001/05/16 (60/291,241) US;
2001/09/14 (60/322,284) US

(51) Cl.Int.^{7}/Int.Cl.^{7} A61K 39/00

(71) Demandeurs/Applicants:
CURAGEN CORPORATION, US;
MILLENNIUM PHARMACEUTICALS, INC., US

(72) Inventeurs/Inventors:
PENA, CAROL E. A., US;
GUO, XIAOJIA, US;
SHIMKETS, RICHARD A., US;
PADIGARU, MURALIDHARA, US;
KEKUDA, RAMESH, US;
SPYTEK, KIMBERLY A., US;
...

(74) Agent: GOUDREAU GAGE DUBUC

(54) Titre : PROTEINES ET ACIDES NUCLEIQUES CODANT POUR CELLES-CI
(54) Title: PROTEINS AND NUCLEIC ACIDS ENCODING SAME

(57) Abrégé/Abstract:
Disclosed are polypeptides and nucleic acids encoding same. Also disclosed are vectors, host cells, antibodies and recombinant methods for producing the polypeptides and polynucleotides, as well as methods for using same.

(21) 2 443 770

(13) A1

(72) **Inventeurs(suite)/Inventors(continued)**: MEHRABAN, FUAD, US; TOPPER, JAMES N., US; MALYANKAR, URIEL M., US; WASSERMAN, SCOTT, US; EDINGER, R. SHLOMIT, US; SMITHSON, GLENNDA, US; GUNTHER, ERIK, US; KOMUVES, LASZLO, US

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 October 2002 (31.10.2002)

PCT

(10) International Publication Number
WO 02/085922 A3

(51) International Patent Classification[7]: A61K 39/00

(21) International Application Number: PCT/US02/11634

(22) International Filing Date: 11 April 2002 (11.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/285,748	23 April 2001 (23.04.2001)	US
60/286,068	24 April 2001 (24.04.2001)	US
60/286,292	25 April 2001 (25.04.2001)	US
60/288,334	3 May 2001 (03.05.2001)	US
60/291,241	16 May 2001 (16.05.2001)	US
60/322,284	14 September 2001 (14.09.2001)	US

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

US	60/285,748 (CIP)
Filed on	23 April 2001 (23.04.2001)
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	60/288,334 (CIP)
Filed on	3 May 2001 (03.05.2001)
US	60/291,241 (CIP)
Filed on	16 May 2001 (16.05.2001)
US	60/322,284 (CIP)
Filed on	14 September 2001 (14.09.2001)

(71) Applicants *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US). **MILLENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sydney Street, Cambridge, MA 02139 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PENA, Carol, E., A.** [US/US]; 604 Orange Street, #2, New Haven, CT 06511 (US). **GUO, Xiaojia** [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). **SHIMKETS, Richard, A.** [US/US]; 5 Ludian Meadows Drive, Guilford, CT 06437 (US). **PADIGARU, Muralidhara** [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). **KEKUDA, Ramesh** [IN/US]; 1213 Avalon Valley Drive, Danbury, CT 06810 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street, #1, New Haven, CT 06511 (US). **MEHRABAN, Fuad** [US/US]; 6095 Main Street, Trumbell, CT 06611 (US). **TOPPER, James, N.** [US/US]; 420 Juanita Way, Los Altos, CA 94022 (US). **MALYANKAR, Uriel, M.** [IN/US]; 229 Branford Road, #330, Branford, CT 06405 (US). **WASSERMAN, Scott** [US/US]; 699 Mississippi Street, #222, San Francisco, CA 94107 (US). **EDINGER, R., Shlomit** [US/US]; 766 Edgewood Avenue, New Haven, CT 06515 (US). **SMITHSON, Glennda** [US/US]; 125 Michael Drive, Guildford, CT 06435 (US). **GUNTHER, Erik** [US/US]; 34 Bryan Road, Branford, CT 06405 (US). **KOMUVES, Laszlo** [HU/US]; 554 Magellan Avenue, San Fransisco, CA 94116 (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*

(88) Date of publication of the international search report:
30 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.

WO 02/085922 A3

(54) Title: PROTEINS AND NUCLEIC ACIDS ENCODING SAME

(57) Abstract: Disclosed are polypeptides and nucleic acids encoding same. Also disclosed are vectors, host cells, antibodies and recombinant methods for producing the polypeptides and polynucleotides, as well as methods for using same.

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 192

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 192

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

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

PROTEINS AND NUCLEIC ACIDS ENCODING SAME

FIELD OF THE INVENTION

The invention relates to polynucleotides and the polypeptides encoded by such polynucleotides, as well as vectors, host cells, antibodies and recombinant methods for producing the polypeptides and polynucleotides, and methods for using the same.

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom. More specifically, the invention relates to nucleic acids encoding cytoplasmic, nuclear, membrane bound, and secreted polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

Heart disease is the primary cause of death in most western societies. Death from heart disease is often induced by platelet-dependent ischemic syndromes which are initiated by atherosclerosis and arteriosclerosis and include, but are not limited to, acute myocardial infarction, chronic unstable angina, transient ischemic attacks and strokes, peripheral vascular disease, arterial thrombosis, preeclampsia, embolism, restenosis and/or thrombosis following angioplasty, carotid endarterectomy, anastomosis of vascular grafts, and chronic cardiovascular devices (e.g., in-dwelling catheters or shunts "extracorporeal circulating devices"). These syndromes represent a variety of stenotic and occlusive vascular disorders thought to be initiated by platelet activation either on vessel walls or within the lumen by blood-borne mediators but are manifested by platelet aggregates which form thrombi that restrict blood flow.

For example, Thrombospondin-1-like proteins associate with the extracellular matrix and inhibits angiogenesis in vivo. In vitro, Thrombospondin-like proteins block capillary-like tube formation and endothelial cell proliferation. The antiangiogenic activity is mediated by a region that contains 3 type 1 (properdin or thrombospondin) repeats.

In addition, Selectin-like proteins such as P-selectin, also called GMP-140, CD62, or selectin P, is a 140-kD adhesion molecule, expressed at the surface of activated cells, that mediates the interaction of activated endothelial cells or platelets with leukocytes. In endothelial cells, the protein is localized to the membranes of Weibel-Palade bodies, the intracellular storage granules for von Willebrand factor.

Many disease states are characterized by uncontrolled cell proliferation. These diseases involve a variety of cell types and include disorders such as cancer, psoriasis, pulmonary fibrosis, glomerulonephritis, atherosclerosis and restenosis following angioplasty. Vital cellular functions such as cell proliferation and signal transduction are regulated in part by the balance between the activities of protein-tyrosine kinases (PTK) and protein-tyrosine phosphatases (PTPase). Oncogenesis can result from an imbalance.

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, NOV4, NOV5, NOV6, NOV7, NOV8, NOV9, NOV10a, NOV10b, NOV11, NOV12, NOV13, NOV14, NOV15, and NOV16 nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as variants, 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 NOVX nucleic acid molecule encoding a NOVX polypeptide that includes a nucleic acid sequence that has identity to the nucleic acids disclosed in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33. In some embodiments, the NOVX nucleic acid molecule will hybridize under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule that includes a protein-coding sequence of a NOVX nucleic acid sequence. The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes the nucleic acid sequence of any of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33.

Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide which includes at least 6 contiguous nucleotides of a NOVX nucleic acid (*e.g.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33) or a complement of said oligonucleotide.

Also included in the invention are substantially purified NOVX polypeptides (SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34). In certain embodiments,

the NOVX polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human NOVX polypeptide.

The invention also features antibodies that immunoselectively bind to NOVX polypeptides, or fragments, homologs, analogs or derivatives thereof.

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 a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a NOVX nucleic acid, under conditions allowing for expression of the NOVX polypeptide encoded by the DNA. If desired, the NOVX polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a NOVX polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex between the polypeptide and the compound. The complex is detected, if present, thereby identifying the NOVX polypeptide within the sample.

The invention also includes methods to identify specific cell or tissue types based on their expression of a NOVX.

Also included in the invention is a method of detecting the presence of a NOVX nucleic acid molecule in a sample by contacting the sample with a NOVX nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a NOVX nucleic acid molecule in the sample.

In a further aspect, the invention provides a method for modulating the activity of a NOVX polypeptide by contacting a cell sample that includes the NOVX polypeptide with a compound that binds to the NOVX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon containing) or inorganic molecule, as further described herein.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes including, *e.g.*, those

described for the individual NOVX nucleotides and polypeptides herein, and/or other pathologies and disorders of the like.

The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or a NOVX-specific antibody, or biologically-active derivatives or fragments thereof. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed below and/or other pathologies and disorders of the like. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding NOVX may be useful in gene therapy, and NOVX may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

The invention further includes a method for screening for a modulator of disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The method includes contacting a test compound with a NOVX polypeptide and determining if the test compound binds to said NOVX polypeptide. Binding of the test compound to the NOVX polypeptide indicates the test compound is a modulator of activity, or of latency or predisposition to the aforementioned disorders or syndromes.

Also within the scope of the invention is a method for screening for a modulator of activity, or of latency or predisposition to an disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like by administering a test compound to a test animal at increased risk for the aforementioned disorders or syndromes. The test animal expresses a recombinant polypeptide encoded by a NOVX nucleic acid. Expression or activity of NOVX polypeptide is then measured in the test animal, as is expression or activity of the protein in a control animal which recombinantly-expresses NOVX polypeptide and is not at increased risk for the disorder or syndrome. Next, the expression of NOVX polypeptide in both the test animal and the control animal is compared. A change in the activity of NOVX polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of the disorder or syndrome.

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, a NOVX nucleic acid, or both, in a subject (*e.g.*, a human subject). The method includes measuring the

amount of the NOVX polypeptide in a test sample from the subject and comparing the amount of the polypeptide in the test sample to the amount of the NOVX polypeptide present in a control sample. An alteration in the level of the NOVX polypeptide in the test sample as compared to the control sample indicates the presence of or predisposition to a disease in the subject. Preferably, the predisposition includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. Also, the expression levels of the new polypeptides of the invention can be used in a method to screen for various cancers as well as to determine the stage of cancers.

In a further aspect, the invention includes a method of treating or preventing a pathological condition associated with a disorder in a mammal by administering to the subject a NOVX polypeptide, a NOVX nucleic acid, or a NOVX-specific antibody to a subject (*e.g.*, a human subject), in an amount sufficient to alleviate or prevent the pathological condition. In preferred embodiments, the disorder, includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

In yet another aspect, the invention can be used in a method to identity the cellular receptors and downstream effectors of the invention by any one of a number of techniques commonly employed in the art. These include but are not limited to the two-hybrid system, affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

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

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

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 and their polypeptides. The sequences are collectively referred to as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or

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

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX	Internal Identification	SEQ ID NO (nt)	SEQ ID NO (aa)	Homology
1	CG93221-01	1	2	Paladin
2	CG93210-01	3	4	Plasma membrane ring finger protein
3	CG93275-01	5	6	Thrombospondin-1 domain containing protein
4	CG93187-01	7	8	Protocadherin alpha C2 short form
5	COR CG95083-01	9	10	Nuclear protein
6	COR CG94989-01	11	12	Secretory protein
7	COR CG94978-01	13	14	Transmission-blocking target antigen S230 precursor
8	COR CG94713-01	15	16	Nuclear protein
9	COR CG94702-01	17	18	Hemicentin precursor
10a	COR CG94661-01	19	20	Selectin
10b	COR CG94661-02	21	22	Selectin
11	COR CG94325-01	23	24	Nuclear protein
12	COR CG94282-01	25	26	Plasma membrane protein
13	COR CG94399-01	27	28	BHLH Factor MATH6
14	COR CG94366-01	29	30	Putative protein-tyrosine phosphatase
15	CG95387-02	31	32	LRR protein
16	CG95419-02	33	34	RhoGEF

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 sixteen 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 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, *e.g.*, cell growth, cell metabolism, cell differentiation, cell proliferation, and/or cell signaling.

In one embodiment of the present invention, NOVX or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NOVX. Examples of such disorders include, but are not limited to, cancers such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; neurological disorders such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; and disorders of vesicular transport such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers, other conditions associated with abnormal vesicle trafficking including acquired immunodeficiency syndrome (AIDS), allergic reactions, autoimmune hemolytic anemia, proliferative glomerulonephritis, inflammatory bowel disease, multiple sclerosis, myasthenia gravis, rheumatoid arthritis, osteoarthritis, scleroderma, Chediak-Higashi

syndrome, Sjogren's syndrome, systemic lupus erythiematosus, toxic shock syndrome, traumatic tissue damage, and viral, bacterial, fungal, helminthic, and protozoal infections, as well as additional indications listed for the individual NOVX clones.

The NOVX nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These also include 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), (v) an agent promoting tissue regeneration *in vitro* and *in vivo*, and (vi) a biological defense weapon. Additional utilities for the NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOV1

A NOV1 polypeptide has been identified as a Paladin-like protein (also referred to as CG93221-01). The disclosed novel NOV1 nucleic acid (SEQ ID NO:1) of 2600 nucleotides is shown in Table 1A. The novel NOV1 nucleic acid sequences maps to the chromosome 10.

An ORF begins with an ATG initiation codon at nucleotides 15-17 and ends with a TAG codon at nucleotides 2583-2585. A putative untranslated region and/or downstream from the termination codon is underlined in Table 1A, and the start and stop codons are in bold letters.

Table 1A. NOV1 Nucleotide Sequence (SEQ ID NO:1)

GCTGCTGGCAGACT**ATG**GGTACAACGGCCAGCACAGCCCAGCAGACGGTCTCGGCAGGCACCCCATT
TGAGGGCCTACAGGGCAGTGGCACGATGGACAGTCGGCACTCCGTCAGCATCCACTCCTTCCAGAGC
ACTAGCTTGCATAACAGCAAGGCCAAGTCCATCATCCCCAACAAGGTGGCCCCTGTTGTGATCACGT
ACAACTGCAAGGAGGAGTTCCAGATCCATGATGAGCTGCTCAAGGCTCATTACACGTTGGGCCGGCT
CTCGGACAACACCCCTGAGCACTACCTGGTGCAAGGCCGCTACTTCCTGGTGCGGGATGTCACTGAG
AAGATGGATGTGCTGGGCACCGTGGGAAGCTGTGGGGCCCCCAACTTCCGGCAGGTGCAGGGTGGGC
TCACTGTGTTCGGCATGGGACAGCCCAGCCTCTTAGGGTTCAGGCGGGTCCTCCAGAAACTCCAGAA
GGACGGACATAGGGAGTGTGTCATCTTCTGTGTGCGGGAGGAACCTGTGCTTTTCCTGCGTGCAGAT
GAGGACTTTGTGTCCTACACACCTCGAGACAAGCAGAACCTTCATGAGAACCTCCAGGGCCTTGGAC
CCGGGGTCCGGGTGGAGAGCCTGGAGCTGGCCATCCGGAAAGAGATCCACGACTTTGCCCAGCTGAG
CGAGAACACATACCATGTGTACCATAACACCGAGGACCTGTGGGGGGAGCCCCATGCTGTGGCCATC
CATGGTGAGGACGACTTGCATGTGACGGAGGAGGTGTACAAGCGGCCCCTCTTCCTGCAGCCCACCT
ACAGGTACCACCGCCTGCCCCTGCCCGAGCAAGGGAGTCCCCTGGAGGCCCAGTTGGACGCCTTTGT
CAGTGTTCTCCGGGAGACCCCCAGCCTGCTGCAGCTCCGTGATGCCCACGGGCCTCCCCCAGCCCTC
GTCTTCAGCTGCCAGATGGGCGTGGGCAGGACCAACCTGGGCATGGTCCTGGGCACCCTCATCCTGC
TTCACCGCAGTGGGACCACCTCCCAGCCAGAGGCTGCCCCCACGCAGGCCAAGCCCCTGCCTATGGA
GCAGTTCCAGGTGATCCAGAGCTTTCTCCGCATGGTGCCCCAGGGAAGGAGGATGGTGGAAGAGGTG
GACAGAGCCATCACTGCCTGTGCCGAGTTGCATGACCTGAAAGAAGTGGTCTTGGAAAACCAGAAGA
AGTTAGAAGGTATCCGACCGGAGAGCCCAGCCCAGGGAAGCGGCAGCCGACACAGCGTCTGGCAGAG

GGCGCTGTGGAGCCTGGAGCGATACTTCTACCTGATCCTGTTTAACTACTACCTTCATGAGCAGTAC
CCGCTGGCCTTTGCCCTCAGTTTCAGCCGCTGGCTGTGTGCCCACCCTGAGCTGTACCGCCTGCCCG
TGACGCTGAGCTCAGCAGGCCCTGTGGCTCCGAGGGACCTCATCGCCAGGGGCTCCCTACGGGAGGA
CGATCTGGTCTCCCCGGACGCGCTCAGCACTGTCAGAGAGATGGATGTGGCCAACTTCCGGCGGGTG
CCCCGCATGCCCATCTACGGCACGGCCCAGCCCAGCGCCAAGGCCCTGGGGAGCATCCTGGCCTACC
TGACGGACGCCAAGAGGAGGCTGCGGAAGGTTGTCTGGGTGAGCCTTCGGGAGGAGGCCGTGTTGGA
GTGTGACGGGCACACCTACAGCCTGCGGTGGCCTGGGCCCCCTGTGGCTCCTGACCAGCTGGAGACC
CTGGAGGCCCAGCTGAAGGCCCATCTAAGCGAGCCTCCCCCAGGCAAGGAGGGCCCCCTGACCTACA
GGTTCCAGACCTGCCTTACCATGCAGGAGGTCTTCAGCCAGCACCGCAGGGCCTGTCCTGGCCTCAC
CTACCACCGCATCCCCATGCCGGACTTCTGTGCCCCCCGAGAGGAGGACTTTGACCAGCTGCTGGAG
GCCCTGCGGGCCGCCCTCTCCAAGGACCCAGGCACTGGCTTCGTGTTCAGCTGCCTCAGCGGCCAGG
GCCGTACCACAACTGCGATGGTGGTGGCTGTCCTGGCCTTCTGGCACATCCAAGGCTTCCCCGAGGT
GGGTGAGGAGGAGCTCGTGAGTGTGCCTGATGCCAAGTTCACTAAGGGTGAATTTCAGGTAGTAATG
AAGGTGGTGCAGCTGCTACCCGATGGGCACCGTGTGAAGAAGGAGGTGGACGCAGCGCTGGACACTG
TCAGCGAGACCATGACGCCCATGCACTACCACCTGCGGGAGATCATCATCTGCACCTACCGCCAGGC
GAAGGCAGCGAAAGAGGCGCAGGAAATGCGGAGGCTGCAGCTGCGGAGCCTGCAGTACTTGGAGCGC
TATGTCTGCCTGATTCTCTTCAACGCGTACCTCCACCTGGAGAAGGCCGACTCCTGGCAGAGGCCCT
TCAGCACCTGGATGCAGGAGGTGGCATCGAAGGCTGGCATCTACGAGATCCTTAACGAGCTGGGCTT
CCCCGAGCTGGAGAGCGGGGAGGACCAGCCCTTCTCCAGGCTGCGCTACCGGTGGCAGGAGCAGAGC
TGCAGCCTCGAGCCCTCTGCCCCCGAGGACTTGCTG**TAG**GGGGCCTTACTCCCT

Variant sequences of NOV1 are included in Example 3, Table 18. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA.

The NOV1 protein (SEQ ID NO:2) encoded by SEQ ID NO:1 is 856 amino acid residues in length and is presented using the one-letter amino acid code in Table 1B. Psort analysis predicts the NOV1 protein of the invention to be localized in the cytoplasm with a certainty of 0.4500.

Table 1B. Encoded NOV1 protein sequence (SEQ ID NO:2)

MGTTASTAQQTVSAGTPFEGLQGSGTMDSRHSVSIHSFQSTSLHNSKAKSIIPNKVAPVVITYNC
KEEFQIHDELLKAHYTLGRLSDNTPEHYLVQGRYFLVRDVTEKMDVLGTVGSCGAPNFRQVQGGL
TVFGMGQPSLLGFRRVLQKLQKDGHRECVIFCVREEPVLFLRADEDFVSYTPRDKQNLHENLQGL
GPGVRVESLELAIRKEIHDFAQLSENTYHVYHNTEDLWGEPHAVAIHGEDDLHVTEEVYKRPLFL
QPTYRYHRLPLPEQGSPLEAQLDAFVSVLRETPSLLQLRDAHGPPPALVFSCQMGVGRTNLGMVL
GTLILLHRSGTTSQPEAAPTQAKPLPMEQFQVIQSFLRMVPQGRRMVEEVDRAITACAELHDLKE
VVLENQKKLEGIRPESPAQGSGSRHSVWQRALWSLERYFYLILFNYYLHEQYPLAFALSFSRWLC
AHPELYRLPVTLSSAGPVAPRDLIARGSLREDDLVSPDALSTVREMDVANFRRVPRMPIYGTAQP
SAKALGSILAYLTDAKRRLRKVVWVSLREEAVLECDGHTYSLRWPGPPVAPDQLETLEAQLKAHL
SEPPPGKEGPLTYRFQTCLTMQEVFSQHRRACPGLTYHRIPMPDFCAPREEDFDQLLEALRAALS
KDPGTGFVFSCLSGQGRTTTAMVVAVLAFWHIQGFPEVGEEELVSVPDAKFTKGEFQVVMKVVQL
LPDGHRVKKEVDAALDTVSETMTPMHYHLREIIICTYRQAKAAKEAQEMRRLQLRSLQYLERYVC
LILFNAYLHLEKADSWQRPFSTWMQEVASKAGIYEILNELGFPELESGEDQPFSRLRYRWQEQSC
SLEPSAPEDLL

In all BLAST alignments described herein, the “E-value” or “Expect” value is a numeric indication of the probability that the aligned sequences could have achieved their similarity to the BLAST query sequence by chance alone, within the database that was searched. The Expect value (E) is a parameter that describes the number of hits one can

"expect" to see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences.

The Expect value is used to create a significance threshold for reporting results. The default value used for blasting is typically set to 0.0001, with the filter to remove low complexity sequence turned off. In BLAST 2.0, the Expect value is also used instead of the P value (probability) to report the significance of matches. For example, an E value of one assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see one match with a similar score simply by chance. An E value of zero means that one would not expect to see any matches with a similar score simply by chance. See, *e.g.*, http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/. Occasionally, a string of X's or N's will result from a BLAST search. This is a result of automatic filtering of the query for low-complexity sequence that is performed to prevent artifactual hits The filter substitutes any low-complexity sequence that it finds with the letter "N" in nucleotide sequence (e.g., "NNNNNNNN") or the letter "X" in protein sequences (e.g., "XXX"). Low-complexity regions can result in high scores that reflect compositional bias rather than significant position-by-position alignment. Wootton and Federhen, Methods Enzymol 266:554-571, (1996).

A search against the Patp database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 1C.

Table 1C. Patp results for NOV1

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAB41108 Human ORFX ORF872 polypeptide	+1	4187	0.0
>patp:AAB35276 Murine dual specificity phosphatase DSP-11	+1	120	5.2e-06
>patp:AAB73211 Murine phosphatase AA023073_m	+1	120	5.2e-06
>patp:AAB73231 Human phosphatase BAA91172_h	+1	115	1.8e-05
>patp:AAG67455 Amino acid sequence of a human polypeptide	+1	115	1.8e-05

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 2063 of 2508 bases (82%) identical to a gb:GENBANK-ID:MMPAL|acc:X99384.1 mRNA from *Mus musculus* (Paladin gene). The full amino acid sequence of the protein of the invention was found to have 695 of 859 amino acid residues (80%) identical to, and 754 of 859 amino acid residues (87%) similar to, the 859 amino acid residue ptnr:SPTREMBL-ACC:P70261 protein from *Mus musculus* (PALADIN GENE). NOV1 also has homology to the proteins shown in the BLASTP data in Table 1D.

Table 1D. BLAST results for NOV1

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|6331287|dbj|BAA86588.1| (AB033100)	KIAA1274 protein [*Homo sapiens*]	752	752/752 (100%)	752/752 (100%)	0.0
gi|14738662|ref|XP_046314.1| (XM_046314)	KIAA protein (similar to mouse paladin) [*Homo sapiens*]	748	747/748 (99%)	747/748 (99%)	0.0
gi|7305365|ref|NP_038781.1| (NM_013753)	paladin [*Mus musculus*]	859	673/841 (80%)	730/841 (86%)	0.0
gi|15228672|ref|NP_191760.1| (NM_116066)	putative protein [*Arabidopsis thaliana*]	1232	207/821 (25%)	340/821 (41%)	7e-45
gi|12836455|dbj|BAB23663.1|(AK004912)	data source:SPTR, source key:Q9NX48, evidence:ISS~homolog to CDNA FLJ20442 FIS, CLONE KAT04828~putative [*Mus musculus*]	144	24/60 (40%)	33/60 (55%)	2e-04

A multiple sequence alignment is given in Table 1E, with the NOV1 protein being shown on line 1 in Table 1E in a ClustalW analysis, and comparing the NOV1 protein with the related protein sequences shown in Table 1D. This BLASTP data is displayed graphically in the ClustalW in Table 1E.

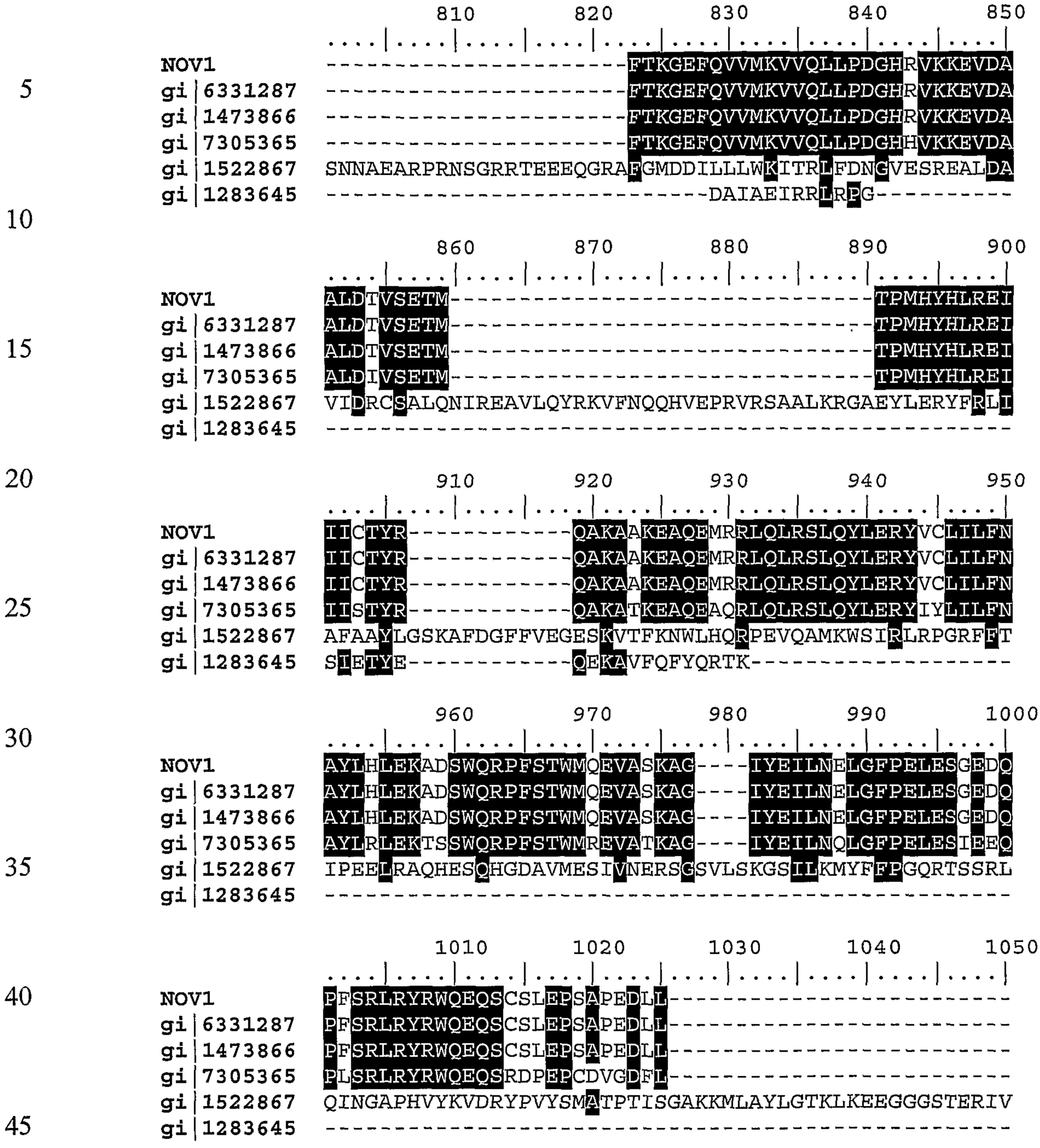
Table 1E. ClustalW Analysis of NOV1

1) > NOV1; SEQ ID NO:2
2) > gi|6331287/ KIAA1274 protein [*Homo sapiens*]; SEQ ID NO:35
3) > gi|1473866/ KIAA protein (similar to paladin) [*Homo sapiens*]; SEQ ID NO:36
4) > gi|7305365/ paladin [*Mus musculus*]; SEQ ID NO:37
5) > gi|1522867/ putative protein [*Arabidopsis thaliana*]; SEQ ID NO:38
6) > gi|1283645/ data source: SPTR, source key: Q9NX48, evidence: ISS-homolog to cDNA FLJ20442 FIS clone KAT04828 putative [*Mus musculus*]; SEQ ID NO:39

```
                                10        20        30        40        50
                     ....|....|....|....|....|....|....|....|....|....|
NOV1                 MGTTASTAQQTVSAGTPFEGLQGSGT--MDSRHSVS-IHSFQSTSLHNSK
gi|6331287           --------------------------------------------------
gi|1473866           --------------------------------------------------
gi|7305365           MGTTASTAQQTVSAGTSLEGLQGGSSSSMDSQHSLGGVQSFRATSLHNSK
gi|1522867           --------------------------------------------------
gi|1283645           --------------------------------------------------

                                60        70        80        90        100
                     ....|....|....|....|....|....|....|....|....|....|
NOV1                 AKSIIPNKVAPVVITYNCKEEFQIHDELLKAHYTLGRLSDNTPEHYLVQG
gi|6331287           --------------------------------------------------
gi|1473866           --------------------------------------------------
gi|7305365           AKSIIPNKVAPVVITYNCKEEFQIHDELLKAHYRMGRLSDATPEHYLVQG
gi|1522867           ----------------------------MSIPKEPEQVMKMRDGSVLGKK
gi|1283645           --------------------------------------------------
```

```
                                110       120       130       140       150
                                ....|....|....|....|....|....|....|....|....|....|
NOV1                            RYFLVRDVTEKMDVLGTVGSCGAPNFRQVQGGLTVFGMGQPSLLGFRRVL
gi|6331287                      -------VTEKMDVLGTVGSCGAPNFRQVQGGLTVFGMGQPSLLGFRRVL
gi|1473866                      ----------MDVLGTVGSCGAPNFRQVQGGLTVFGMGQPSLSGFRRVL
gi|7305365                      RYFLVRDITEKMDILGTLKSCGAPNFRQVRGGLPVFGMGQPSLLGFRRVL
gi|1522867                      TILKSDHFPGCQNKRMTPQIEGAPNYRQADS-LRVHGVAIPTAVGIRNVL
gi|1283645                      ----------------------------------------------------

                                160       170       180       190       200
                                ....|....|....|....|....|....|....|....|....|....|
NOV1                            QKLQKDG---HRECVIFCVREEPVLFLRADEDFVSYTPRDKQNLHENLQG
gi|6331287                      QKLQKDG---HRECVIFCVREEPVLFLRADEDFVSYTPRDKQNLHENLQG
gi|1473866                      QKLQKDG---HRECVIFCVREEPVLFLRADEDFVSYTPRDKQNLHENLQG
gi|7305365                      QKLQTDG---LKECIIFCVREEPVVFLRAEEDFVSYTPRDKESLHENLRD
gi|1522867                      RHIGAHKDGKQVKVLWISLREEPVVYINGRPFVLRDVEKPFTNLEYTGIN
gi|1283645                      ----------------------------------------------------

                                210       220       230       240       250
                                ....|....|....|....|....|....|....|....|....|....|
NOV1                            LGPGVRVESLELAIRKEIHDFAQLSENTYHVYHNTEDLWGEPHAVAIHGE
gi|6331287                      LGPGVRVESLELAIRKEIHDFAQLSENTYHVYHNTEDLWGEPHAVAIHGE
gi|1473866                      LGPGVRVESLELAIRKEIHDFAQLSENTYHVYHNTEDLWGEPHAVAIHGE
gi|7305365                      PSPGVKAENLELAIQKEIHDFAQLRDNVYHVYHNTEDLRGEPHTVAIRGE
gi|1522867                      ---RVRVEQMEARLKEDILMEASRYGNKILVTDELPDGQMVDQWEPVS-T
gi|1283645                      ----------------------------------------------------

                                260       270       280       290       300
                                ....|....|....|....|....|....|....|....|....|....|
NOV1                            DDLHVTEEVYKRPLFLQPTYRYHRLPLPEQGSPLEAQLDAFVSVLRETPS
gi|6331287                      DDLHVTEEVYKRPLFLQPTYRYHRLPLPEQGSPLEAQLDAFVSVLRETPS
gi|1473866                      DDLHVTEEVYKRPLFLQPTYRYHRLPLPEQGSPLEAQLDAFVSVLRETPS
gi|7305365                      DGVCVTEEVFKRPLLLQPTYRYHHLPLPEQGPPLEAQFDAFVSVLRETPS
gi|1522867                      DSLKTLLEVYEELQAEGYLVDYERVPITDEKSPKETDFDLLIRKISQAD-
gi|1283645                      ----------------------------------------------------

                                310       320       330       340       350
                                ....|....|....|....|....|....|....|....|....|....|
NOV1                            LLQLRDAHGPPPALVFSCQMGVGRTNLGMVLGTLILLHRSGTTSQ-----
gi|6331287                      LLQLRDAHGPPPALVFSCQMGVGRTNLGMVLGTLILLHRSGTTSQ-----
gi|1473866                      LLQLRDAHGPPPALVFSCQMGVGRTNLGMVLGTLILLHRSGTTSQ-----
gi|7305365                      LLPLRDNHGPLPALLFSCQSGVGRTNLGMVLGTLVMFHHSRTTSQ-----
gi|1522867                      ----------INTEIIFNCQMGRGRTTTGMVIATLVYFKRTGASDQGFPRN
gi|1283645                      ----------------------------------------------------

                                360       370       380       390       400
                                ....|....|....|....|....|....|....|....|....|....|
NOV1                            ----------PEAAPTQAKPLPMEQFQVIQSFLRMVPQGRRMVEEVD
gi|6331287                      ----------PEAAPTQAKPLPMEQFQVIQSFLRMVPQGRRMVEEVD
gi|1473866                      ----------PEAAPTQAKPLPMEQFQVIQSFLRMVPQGRRMVEEVD
gi|7305365                      ----------LEAASPLAKPLPMEQFQVIQGFICKVPQGKKMVEEVD
gi|1522867                      NSFGRIFKAGENITVNLPNSEEAIRRGEYAVVRSLIRVLEGGVEGKRQVD
gi|1283645                      ----------------------------------------------------

                                410       420       430       440       450
                                ....|....|....|....|....|....|....|....|....|....|
NOV1                            RAITACAELHDLKEVVLENQKKLEGIRPESPAQGSGSRHSVWQRALWSLE
gi|6331287                      RAITACAELHDLKEVVLENQKKLEGIRPESPAQGSGSRHSVWQRALWSLE
gi|1473866                      RAITACAELHDLKEVVLENQKKLEGIRPESPAQGSGSRHSVWQRALWSLE
gi|7305365                      RAISACAELHDLKEEVLKNQRRLESFRPESRGQECGSQQAVQQRALWSLE
gi|1522867                      KVIDKCASMQNLREAIATYRSSILRQPDEKK------REAALSFFVEYLE
gi|1283645                      ----------------------------------------------------
```

```
                                460       470       480       490       500
                     ....|....|....|....|....|....|....|....|....|....|....|
NOV1                 RYFYLILFNYYLHEQYPLAFA-----LSFSRWLCAHPELYRLPVTLSSAG
gi|6331287           RYFYLILFNYYLHEQYPLAFA-----LSFSRWLCAHPELYRLPVTLSSAG
gi|1473866           RYFYLILFNYYLHEQYPLAFA-----LSFSRWLCAHPELYRLPVTLSSAG
gi|7305365           LYFYLLLFNYYLHEQYPLAFA-----LSFSRWLCTHPELYRLLVELNSVG
gi|1522867           RYYFLICFAVYLHSEGAFLQSGSLGHVSFADWMRARPELYSILRRLLRRD
gi|1283645           ----------------------------------------------------

                                510       520       530       540       550
                     ....|....|....|....|....|....|....|....|....|....|....|
NOV1                 PV----------------------APRDLIARGSLREDDLVSPDALST---
gi|6331287           PV----------------------APRDLIARGSLREDDLVSPDALST---
gi|1473866           PV----------------------APRDLIARGSLREDDLVSPDALST---
gi|7305365           PL----------------------VPGDLIAKGSLEADDLVSLDALST---
gi|1522867           PMGALGYAAMKPSLIKIAESTDGRPHEMSVVAALRSGAVLGSQTVLKSDH
gi|1283645           ----------------------------------------------------

                                560       570       580       590       600
                     ....|....|....|....|....|....|....|....|....|....|....|
NOV1                 ----------VREMDVANFRRVPRMPIYGTAQPSAKALGSILAYLTDAKR
gi|6331287           ----------VREMDVANFRRVPRMPIYGTAQPSAKALGSILAYLTDAKR
gi|1473866           ----------VREMDVANFRRVPRMPIYGTAQPSAKALGSILAYLTDAKR
gi|7305365           ----------VREMDVANFRRVPRMPIYGTAQPSAKALGNILAYLSDAKR
gi|1522867           SPGCQILNLPERVEGAPNFREVPGFPVYGVANPTIDGIRSVIERVGSSR-
gi|1283645           ----------------------------------------------------

                                610       620       630       640       650
                     ....|....|....|....|....|....|....|....|....|....|....|
NOV1                 RLRKVVWVSLREEAVLECDGHTYSLRWPGPP----------VAPDQLETLE
gi|6331287           RLRKVVWVSLREEAVLECDGHTYSLRWPGPP----------VAPDQLETLE
gi|1473866           RLRKVVWVSLREEAVLECDGHTYSLRWPGPP----------VAPDQLETLE
gi|7305365           KLRQVVWIFLREEVVLECDGHTHSLWPPGPA----------LAPEHLVALE
gi|1522867           GGRPVFWHNMREEPVIYINGKPFVLREVERPYKNMLEYTGIDRDRVEGME
gi|1283645           ---NFSWVLL----------GRLAGLALPRLP----------AHYQFLLDQG

                                660       670       680       690       700
                     ....|....|....|....|....|....|....|....|....|....|....|
NOV1                 AQLKAHLSEPPPGKEGPLTYRFQT----------------CLTMQEVFSQ
gi|6331287           AQLKAHLSEPPPGKEGPLTYRFQT----------------CLTMQEVFSQ
gi|1473866           AQLKAHLSEPPPGKEGPLTYRFQT----------------CLTMQEVFSQ
gi|7305365           AQLKAHLSAPVPNTKSPTAPRFQK----------------CLTTQEVFSQ
gi|1522867           ARLKEDILREAKRYDGAIMVIHETKDGQIFDLWENVDADSVQTPLEVYKC
gi|1283645           VRHLVSLTE-----RGP------------------------------P

                                710       720       730       740       750
                     ....|....|....|....|....|....|....|....|....|....|....|
NOV1                 HRRACPGLTYHRIPMPDFCAPREEDFDQLLEALRAALSKDPGTGFVFSCL
gi|6331287           HRRACPGLTYHRIPMPDFCAPREEDFDQLLEALRAALSKDPGTGFVFSCL
gi|1473866           HRRACPGLTYHRIPMPDFCAPREEDFDQLLEALRAALSKDPGTGFVFSCL
gi|7305365           HQGACLGLTYCRIPVPDFCAPREEDFDRLLEALRAALTKDPGTGFVFSCL
gi|1522867           LEADGFPIKYARVPITDGKAPKSSDFDTLTSNIASAS---KDTAFVFNCQ
gi|1283645           HSDSCPGLTLHRMRIPDFCPPSPEQIDQFVKIVDEANARGEAVG--VHCA

                                760       770       780       790       800
                     ....|....|....|....|....|....|....|....|....|....|....|
NOV1                 SGQGRTTTAMVVAVLAFWHIQG----------FPEVGEEELVSVPDAK-
gi|6331287           SGQGRTTTAMVVAVLAFWHIQG----------FPEVGEEELVSVPDAK-
gi|1473866           SGQGRTTTAMVVAVLAFWHIQG----------FPEVGEEELVSVPDAK-
gi|7305365           SGQGRTTTAMVVAVLACWHIGG----------CPEVGEEELVSVPDAK-
gi|1522867           MGRGRTTTGTVIACLVKLRINYGRPIKVLYDVLTHEIVDEDSSSGGEETG
gi|1283645           LGFGRTG-----TMLACYLVK----------------ERALAAG-----
```



The NOV1 Clustal W alignment shown in Table 1E was modified to begin at amino residue 1050. The data in Table 1E includes all of the regions overlapping with the NOV1 protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 1F lists the domain description from DOMAIN analysis results against NOV1.

Table 1F Domain Analysis of NOV1

Model	Region of Homology	Score (bits)	E value
#PD396342 PALADIN GENE	19-118	431	4e-43
#PD148197 PALADIN GENE	119-399	1175	e-129
#PD222597 DOMAIN OF UNKNOWN	119-231	119	6e-07
#PD306865 PALADIN GENE	354-454	132	2e-08
#PD024454 PLASMID ORFS	356-445	84	0.007
#PD325716 PALADIN GENE	400-604	800	6e-86
#PD326847 CG18442	458-594	97	2e-04
#PD222597 DOMAIN OF UNKNOWN	505-602	113	3e-06
#PD277963 HYDROLASE KAT04828 FIS CDNA	595-648	97	2e-04
#PD148197 PALADIN GENE	605-678	340	1e-32
#PD306865 PALADIN GENE	680-856	765	7e-82
#PD325716 PALADIN GENE	751-820	86	0.004

Consistent with other known members of the Paladin-like family of proteins, NOV1 has, for example, multiple Paladin gene signature sequences and homology to other members of the Paladin-like Protein Family. NOV 1 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV1 nucleic acids and polypeptides can be used to identify proteins that are members of the Paladin like family of proteins. The NOV1 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV1 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, *e.g.*, cellular activation, cellular metabolism and signal transduction. These molecules can be used to treat, *e.g.*, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD),atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, diabetes, Von Hippel-Lindau (VHL) syndrome, pancreatitis, obesity, hyperthyroidism and hypothyroidism, hypercalceimia, ulcers, cirrhosis, transplantation, inflammatory bowel disease, diverticular disease, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmume disease, allergies, immunodeficiencies, transplantation, graft vesus host, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmume disease, allergies, immunodeficiencies, transplantation, graft versus host disease (GVHD), lymphedema, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's

disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, cancer, trauma, regeneration (*in vitro* and *in vivo*), viral/bacterial/parasitic infections, as well as other diseases, disorders and conditions.

In addition, various NOV1 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV1 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the Paladin-like protein family.

Paladin proteins are a family of protein-tyrosine phosphatases. The protein phosphatases can be divided into 2 large families: the serine/threonine phosphatases, which are metalloproteins, and the protein-tyrosine phosphatases, which proceed via a thiol-phosphate enzyme intermediate. The protein-tyrosine phosphatase family includes the VH1-like dual-specificity phosphatases. These phosphatases dephosphorylate phosphotyrosine- as well as phosphoserine- and phosphothreonine-containing substrates. Members of the dual-specificity phosphatase protein family inactivate mitogen-activated protein (MAP) kinase through dephosphorylation of critical threonine and tyrosine residues. Members of the MAP kinase family play a pivotal role in cellular signal transduction. Using a subtractive screen of mouse gastrulation, Pearce *et al.* (1996) identified a novel mouse gene, paladin, with similarity to the dual specificity protein phosphatase family.

The NOV1 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of cardiac and endocrine physiology. As such, the NOV1 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat muscle and nervous system disorders, *e.g.,* cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD),atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, diabetes, Von Hippel-Lindau (VHL) syndrome, pancreatitis, obesity, hyperthyroidism and hypothyroidism, hypercalceimia, ulcers, cirrhosis, transplantation, inflammatory bowel disease, diverticular disease, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmume disease, allergies, immunodeficiencies, transplantation, graft vesus host, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmume disease, allergies,

immunodeficiencies, transplantation, graft versus host disease (GVHD), lymphedema, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, cancer, trauma, regeneration (*in vitro* and *in vivo*), viral/bacterial/parasitic infections, as well as other diseases, disorders and conditions.

The NOV1 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV1 nucleic acid is expressed in brown adipose, heart, aorta, vein, umbilical vein, adrenal gland/suprarenal gland, pancreas, thyroid, salivary glands, parotid salivary glands, stomach, liver, gall bladder, small intestine, colon, bone marrow, lymphoid tissue, spleen, lymph node, tonsils, thymus, cartilage, muscle, brain, thalamus, hypothalamus, pituitary gland, amygdala, substantia nigra, hippocampus, spinal chord, cervix, mammary gland/breast, ovary, placenta, uterus, vulva, prostate, testis, lung, lung pleura, kidney, retina, dermis.

Additional utilities for NOV1 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV2

A NOV2 polypeptide has been identified as a Plasma Membrane Ring Finger-like protein (also referred to as CG93210-01). The disclosed novel NOV2 nucleic acid (SEQ ID NO:3) of 1205 nucleotides is shown in Table 2A. The novel NOV2 nucleic acid sequences maps to the chromosome 22.

An ORF begins with an ATG initiation codon at nucleotides 17-19 and ends with a ATT codon at nucleotides 1149-1151. A putative untranslated region and/or downstream from the termination codon is underlined in Table 2A, and the start and stop codons are in bold letters.

Table 2A. NOV2 Nucleotide Sequence (SEQ ID NO:3)

CTCGCCGGGTCCGGCC**ATG**GGCCCCGCCGCTCGCCCCGCGCTGAGATCGCCGCCGCCGCCTCCGCCG
CCGCCTCCGTCTCCGCTGCTGCTGCTGCTGCCCCTGCTGCCGCTGTGGCTGGGCCTGGCGGGGCCCG
GGGCCGCGGCGGACGGCAGCGAGCCGGCGGCCGGGGCGGGGCGGGGCGGAGCCCGCGCCGTGCGGGT
GGACGTGAGACTGCCGCGCCAGGACGCTCTGGTCCTGGAGGGCGTCAGGATCGGCTCCGAAGCCGAC
CCGGCGCCCCTGCTGGGCGGTCGTCTGCTGCTGATGGACATCGTGGATGCCGAGCAGGAGGCACCAG
TGGAAGGCTGGATTGCAGTGGCATACGTGGGCAAGGAGCAGGCGGCCCAGTTCCACCAGGAGAATAA
GGGCAGTGGCCCGCAGGCCTATCCCAAGGCCCTGGTCCAGCAGATGCGGCGGGCCCTCTTCCTGGGT
GCCTCTGCCCTGCTTCTTCTCATCCTGAACCACAACGTGGTCCGAGAGCTGGACATATCCCAGCTTC

TGCTCAGGCCAGTGATCGTCCTCCATTATTCCTCCAATGTCACCAAGCTGTTGGATGCATTGCTGCA
GAGGACCCAGGCCACGGCTGAGATCACCAGCGGAGAGTCCCTGTCTGCCAATATCGAGTGGAAGTTG
ACCTTGTGGACCACCTGTGGCCTCTCCAAGGATGGCTATGGAGGATGGCAGGACTTGGTCTGCCTTG
GAGGCAGTCGTGCCCAGGAGCAGAAACCCCTGCAGCAGCTGTGGAACGCCATCCTGCTGGTGGCCAT
GCTCCTGTGCACAGGCCTCGTGGTCCAGGCCCAGCGGCAGGCGTCGCGGCAGAGCCAGCGGGAGCTC
GGAGGCCAGGTGGACCTGTTTAAGCGCCGCGTGGTGCGGAGACTGGCATCCCTCAAGACACGGCGCT
GCCGGCTGAGCAGGGCAGCGCAGGGCCTCCCAGATCCGGGTGCTGAGACCTGTGCGGTGTGCCTGGA
CTACTTCTGCAACAAACAGTGGCTCCGGGTGCTGCCCTGTAAGCACGAGTTTCACCGAGACTGTGTG
GACCCCTGGCTGATGCTCCAGCAGACCTGCCCACTGTGCAAATTCAACGTCCTGGGTGAGCACCGCT
ACTCCGATG**ATT**AGCTGCCCAGCTGGACTCTGCACATGGGGATGGACCCCTCCTGCCTGCACCCCG

The NOV2 protein (SEQ ID NO:4) encoded by SEQ ID NO:3 is 378 amino acid residues in length and is presented using the one-letter amino acid code in Table 2B. Psort analysis predicts the NOV2 protein of the invention to be localized at the plasma membrane with a certainty of 0.6400.

Table 2B. Encoded NOV2 protein sequence (SEQ ID NO:4)

MGPAARPALRSPPPPPPPPSPLLLLLPLLPLWLGLAGPGAAADGSEPAAGAGRGGARAVRVDVR
LPRQDALVLEGVRIGSEADPAPLLGGRLLLMDIVDAEQEAPVEGWIAVAYVGKEQAAQFHQENKG
SGPQAYPKALVQQMRRALFLGASALLLLILNHNVVRELDISQLLLRPVIVLHYSSNVTKLLDALL
QRTQATAEITSGESLSANIEWKLTLWTTCGLSKDGYGGWQDLVCLGGSRAQEQKPLQQLWNAILL
VAMLLCTGLVVQAQRQASRQSQRELGGQVDLFKRRVVRRLASLKTRRCRLSRAAQGLPDPGAETC
AVCLDYFCNKQWLRVLPCKHEFHRDCVDPWLMLQQTCPLCKFNVLGEHRYSDD

A search against the Patp database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 2C.

Table 2C. Patp results for NOV2

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAB42695 Human ORFX ORF2459 polypeptide	+1	1715	3.0e-176
>patp:AAM79288 Human protein SEQ ID NO 1950	+1	612	2.3e-59
>patp:AAM80272 Human protein SEQ ID NO 3918	+1	534	4.2e-51
>patp:AAU28202 Novel human secretory protein	+1	201	5.1e-13
>patp:ABB50251 Human transcription factor TRFX-102	+1	148	5.5e-13

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 287 of 489 bases (58%) identical to a gb:GENBANK-ID:SSI132828|acc:AJ132828.1 mRNA from *Spermatozopsis similis* (mRNA for p210 protein, partial). The full amino acid sequence of the protein of the invention was found to have 341 of 379 amino acid residues (89%) identical to, and 355 of 379 amino acid residues (93%) similar to, the 379 amino acid residue ptnr:SPTREMBL-ACC:Q9DCW1 protein from *Mus musculus* (0610009J22RIK PROTEIN).

NOV2 also has homology to the proteins shown in the BLASTP data in Table 2D.

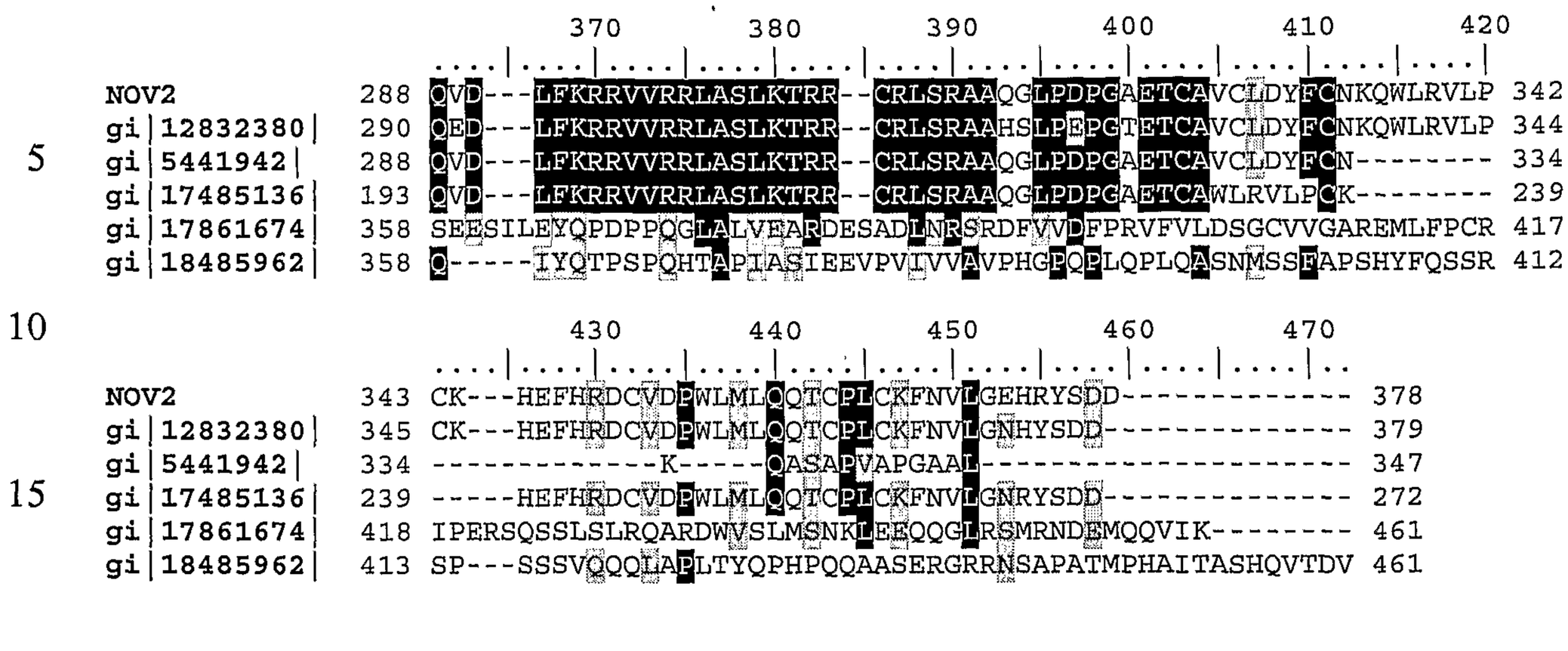
Table 2D. BLAST results for NOV2

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|12832380|dbj|BAB 22082.1| (AK002414)	data source:SPTR, source key:Q9Y6U7, evidence:ISS~homo log to WUGSC:H_DJ130H16. 6 PROTEIN(FRAGMENT) ~putative [*Mus musculus*]	379	340/380 (89%)	354/380 (92%)	e-173
gi|5441942|gb|AAD43 187.1|AC004997_5 (AC004997)	supported by mouse EST AA538043 (NID:g2284036) [*Homo sapiens*]	347	336/336 (100%)	336/336 (100%)	e-148
gi|17485136|ref|XP_ 066294.1| (XM_066294)	similar to data source:SPTR, source key:Q9Y6U7 evidence:ISS~homo log to WUGSC:H_DJ130H16. 6 PROTEIN (FRAGMENT)~putati ve [*Homo sapiens*]	272	271/283 (95%)	272/283 (95%)	e-146
gi|17861674|gb|AAL3 9314.1| (AY069169)	GH20973p [*Drosophila melanogaster*]	461	26/57 (45%)	42/57 (73%)	9e-13
gi|18485962|ref|XP_ 080778.1| (XM_080778)	similar to goliath (*H. sapiens*) [*Drosophila melanogaster*]	461	26/57 (45%)	42/57 (73%)	1e-12

A multiple sequence alignment is given in Table 2E, with the NOV2 protein being shown on line 1 in Table 2E in a ClustalW analysis, and comparing the NOV2 protein with the related protein sequences shown in Table 2D. This BLASTP data is displayed graphically in the ClustalW in Table 2E.

Table 2E. ClustalW Analysis of NOV2

1) > NOV2; SEQ ID NO:4
2) >gi|12832380|/ data source:SPTR, source key:Q9Y6U7, evidence:ISS~homolog to WUGSC:H_DJ130H16.6 PROTEIN(FRAGMENT)~putative [*Mus musculus*]; SEQ ID NO:40
3) >gi|5441942|/ supported by mouse EST AA538043 (NID:g2284036) [*Homo sapiens*]; SEQ ID NO:41
4) >gi|17485136|/ similar to data source:SPTR, source key:Q9Y6U7 evidence:ISS~homolog to WUGSC:H_DJ130H16.6 PROTEIN (FRAGMENT)~putative [*Homo sapiens*]; SEQ ID NO:42
5) >gi|17861674|/ GH20973p [*Drosophila melanogaster*]; SEQ ID NO:43
6) >gi|18485962|/ similar to goliath (H. sapiens) [*Drosophila melanogaster*]; SEQ ID NO:44



The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 2F lists the domain description from DOMAIN analysis results against NOV2.

Table 2F Domain Analysis of NOV2

Model	Region of Homology	Score (bits)	E value
Ring Finger	325-365	49.3	4.0e-07
zf-C3HC4	325-365	34.7	2.2e-09
PHD	324-368	-10.4	1.1

Consistent with other known members of the Membrane Ring Finger-like family of proteins, NOV2 has, for example, a Ring Finger signature sequence and homology to other members of the Plasma Membrane Ring Finger-like Protein Family. NOV 2 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV2 nucleic acids and polypeptides can be used to identify proteins that are members of the Plasma Membrane Ring Finger-like Protein Family. The NOV2 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV1 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, *e.g.,* cellular activation, cellular metabolism and signal transduction. These molecules can be used to treat, *e.g.*, anemia, ataxia-telangiectasia, autoimmume disease, immunodeficiencies, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalceimia, Lesch-Nyhan syndrome, cancer, trauma,

regeneration (*in vitro* and *in vivo*), viral/bacterial/parasitic infections, as well as other diseases, disorders and conditions.

In addition, various NOV2 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV2 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the Plasma Membrane Ring Finger-like Protein Family .

The NOV2 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of immune and renal physiology. As such, the NOV2 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat muscle and nervous system disorders, *e.g.*, anemia, ataxia-telangiectasia, autoimmume disease, immunodeficiencies, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalceimia, Lesch-Nyhan syndrome, cancer, trauma, regeneration (*in vitro* and *in vivo*), viral/bacterial/parasitic infections, as well as other diseases, disorders and conditions.

The NOV2 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV2 nucleic acid is expressed in peripheral blood, and a pool of various mammalian tissues. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CuraGen Acc. No. CG93210-01.The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:SSI132828|acc:AJ132828.1) a closely related *Spermatozopsis similis* mRNA for p210 protein, partial homolog in species *Spermatozopsis similis* :kidney.

Additional utilities for NOV2 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV3

A NOV3 polypeptide has been identified as a Thrombospondin type 1 (tsp_1) domain containing protein (also referred to as CG93275-01). The disclosed novel NOV3 nucleic acid (SEQ ID NO:5) of 799 nucleotides is shown in Table 3A. The novel NOV3 nucleic acid sequences maps to the chromosome 16.

An ORF begins with an ATG initiation codon at nucleotides 51-53 and ends with a TGA codon at nucleotides 744-746. A putative untranslated region and/or downstream from the termination codon is underlined in Table 3A, and the start and stop codons are in bold letters.

Table 3A. NOV3 Nucleotide Sequence (SEQ ID NO:5)

GAATATATTTAGTGTGTTGTTTTTTTTTTTAATGTGGCTACTGAAACCTA**ATG**GGAATGCAAATAGA
ACTTTTTTGTCTTCTCAAGTGTTCCAAGACCTGTGGACGAGGGGTGAGGAAGCGTGAACTCCTCTGC
AAGGGCTCTGCCGCAGAAACCCTCCCCGAGAGCCAGTGTACCAGTCTCCCCAGACCTGAGCTGCAGG
AGGGCTGTGTGCTTGGACGATGCCCCAAGAACAGCCGGCTACAGTGGGTCGCTTCTTCGTGGAGCGA
GTGTTCTGCAACCTGTGGTTTGGGTGTGAGGAAGAGGGAGATGAAGTGCAGCGAGAAGGGCTTCCAG
GGAAAGCTGATAACTTTCCCAGAGCGAAGATGCCGTAATATTAAGAAACCAAATCTGGACTTGGAAG
AGACCTGCAACCGACGGGCTTGCCCAGCCCATCCAGTGTACAACATGGTAGCTGGATGGTATTCATT
GCCGTGGCAGCAGTGCACAGTCACCTGTGGGGGAGGGGTCCAGACCCGGTCAGTCCACTGTGTTCAG
CAAGGCCGGCCTTCCTCAAGTTGTCTGCTCCATCAGAAACCTCCGGTGCTACGAGCCTGTAATACAA
ACTTCTGTCCAGCTCCTGAAAAGAGAGAGGATCCATCCTGCGTAGATTTCTTCAACTGGTGTCACCT
AGTTCCTCAGCATGGTGTCTGCAACCACAAGTTTTACGGAAAACAATGCTGCAAGTCATGCACAAGG
AAGATC**TGA**TCTTGGTGTCCTCCCCAGCCTTAGGGCCAGGGGCTTACCTTTCAACCTCTAGA

The NOV3 protein (SEQ ID NO:6) encoded by SEQ ID NO:5 is 231 amino acid residues in length and is presented using the one-letter amino acid code in Table 3B. Psort analysis predicts the NOV3 protein of the invention to be localized in the cytoplasm with a certainty of 0.4500.

Table 3B. Encoded NOV3 protein sequence (SEQ ID NO:6)

MGMQIELFCLLKCSKTCGRGVRKRELLCKGSAAETLPESQCTSLPRPELQEGCVLGRCPKNSRLQ
WVASSWSECSATCGLGVRKREMKCSEKGFQGKLITFPERRCRNIKKPNLDLEETCNRRACPAHPV
YNMVAGWYSLPWQQCTVTCGGGVQTRSVHCVQQGRPSSSCLLHQKPPVLRACNTNFCPAPEKRED
PSCVDFFNWCHLVPQHGVCNHKFYGKQCCKSCTRKI

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

Table 3C. Patp results for NOV3

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAE09696 Human gene 7 encoding protein HE8CY61	+1	1248	9.2e-127
>patp:AAE09699 Human gene 10 encoding protein HUVHR16	+1	1245	1.9e-126
>patp:AAU72893 Human metalloprotease partial sequence #5	+1	1204	4.2e-122
>patp:AAU72891 Human metalloprotease partial sequence #3	+1	693	3.5e-70
>patp:AAB21253 Human metalloproteinase KIAA0605	+1	327	5.5e-28

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 392 of 396 bases (98%) identical to an EST AA057409 mRNA from human). The full amino acid sequence of the protein of the invention was found to have 74 of 216 amino acid residues (34%) identical to, and 107 of 216 amino acid residues (49%) similar to, the 237 amino acid residue ptnr:SPTREMBL-ACC:Q9HBS6 protein from *Homo sapiens* (HYPOTHETICAL 25.7 KDA PROTEIN).

NOV3 also has homology to the proteins shown in the BLASTP data in Table 3D.

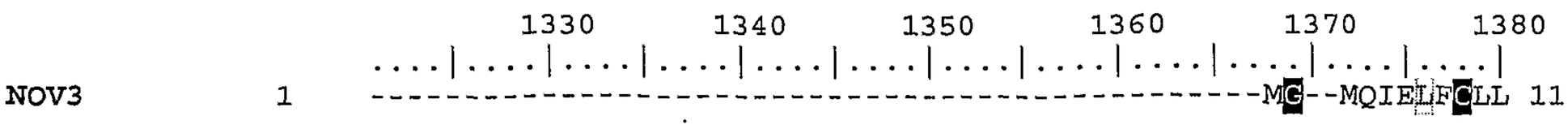
Table 3D. BLAST results for NOV3

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|18598706|ref|XP_091253.1|(XM_091253)	hypothetical protein XP_091253 [*Homo sapiens*]	1123	181/183 (98%)	183/183 (99%)	e-100
gi|19171150|emb|CAC83612.1|(AJ311903)	ADAMTS18 protein [Homo sapiens]	1081	61/62 (98%)	62/62 (99%)	4e-27
gi|7662202|ref|NP_055509.1|(NM_014694)	KIAA0605 gene product [*Homo sapiens*]	951	79/216 (36%)	99/216 (45%)	9e-23
gi|18561227|ref|XP_094442.1|(XM_094442)	hypothetical protein XP_094442 [Homo sapiens]	1365	51/112 (45%)	74/112 (65%)	4e-21
gi|17432918|sp|Q9H324|AT10_	HUMAN ADAMTS-10 precursor (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10)(Fragment)	223	74/223 (33%	104/223 (46%)	5e-20

A multiple sequence alignment is given in Table 3E, with the NOV3 protein being shown on line 1 in Table 3E in a ClustalW analysis, and comparing the NOV3 protein with the related protein sequences shown in Table 3D. This BLASTP data is displayed graphically in the ClustalW in Table 3E.

Table 3E. ClustalW Analysis of NOV3

1) > NOV3; SEQ ID NO:6
2) >gi|18598706|/ hypothetical protein XP_091253 [*Homo sapiens*]; SEQ ID NO:45
3) >gi|19171150|/ ADAMTS18 protein [*Homo sapiens*]; SEQ ID NO:46
4) >gi|7662202|/ KIAA0605 gene product [*Homo sapiens*]; SEQ ID NO:47
5) >gi|18561227|/ hypothetical protein XP_094442 [*Homo sapiens*]; SEQ ID NO:48
6) >gi|17432918|/ AT10_HUMAN ADAMTS-10 precursor (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment); SEQ ID NO:49



```
                gi|18598706|   769  GGVRSAKVLSLEEWIKSETTL-ARKEQQQPSTGWMPGEWSTCSKAC-AGGQQSRKIQCVQ  826
                gi|19171150|   872  ----------------------------------------------------PAYTWSIVQ  881
                gi|7662202|    686  -------------------------------PQWEMSEWSECTAKCGERSVVTRDIRCSE  715
                gi|18561227|   986  QGGGCTPSPPADGLVQNQLRLGRHTQAQCLDQSCSVGNWSACSRTC-VGGAQSRPVQCTR 1044
                gi|17432918|   807  ------------------------------------WTKCSAQC-AGGSQVQAVECRN  828

                                          1390      1400      1410      1420      1430      1440
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
                NOV3             12  K----------------------------------------CSKTCGRGVRKRELLCKG   30
                gi|18598706|    827  KKPFQKEEAVLHSLCPVSTPTQVQACNSHACPPQWSLGPWSQCSKTCGRGVRKRELLCKG  886
                gi|19171150|    882  S----------------------------------------ECSVSCGGGYINVKAICLR  901
                gi|7662202|     716  DEKLCDP---------NTRPVGEKNCTGPPCDRQWTVSDWGPCSGSCGQGRTIRHVYCKT  766
                gi|18561227|   1045  RVHYDSE-PVPASLCPQPAPSSRQACNSQSCPPAWSAGPWAECSHICGKGWRKRAVACKS 1103
                gi|17432918|    829  QLDSSAVAPHYCS-AHSKLPKRQRACNTEPCPPDWVVGNWSLCSRSCDAGVRSRSVVCQR  887

                                          1450      1460      1470      1480      1490      1500
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
                NOV3             31  S----AAETLPESQCTSLPRPELQEG-CVLGRCPKNSRLQWVASSWSECSATCGLGVRKR   85
                gi|18598706|    887  S----AAETLPESQCTSLPRPELQEG-CVLGRCPKNSRLQWVASSWSECSATCGLGVRKR  941
                gi|19171150|    902  D----QNTQVNSSFCSAKTKPVTEPKICNAFSCP----AYWMPGEWSTCSKSCAGGQQSR  953
                gi|7662202|     767  S----DGRVVPESQCQMETKPLAIHP-CGDKNCP----AHWLAQDWERCNTTCGRGVKKR  817
                gi|18561227|   1104  TNPSARAQLLPDAVCTSEPKPRMHEA-CLLQRCHKPKKLQWLVSAWSQCSVTCERGTQKR 1162
                gi|17432918|    888  RVSAAEEKALDDSACP-QPRPPVLEA-CHGPTCP----PEWAALDWSECTPSCGPGLRHR  941

                                          1510      1520      1530      1540      1550      1560
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
                NOV3             86  EMKCSEKGFQGKLITFPERRCRNTKKPNLDLEET-----CNRRACP--AHPVYNMVAGWY  138
                gi|18598706|    942  EMKCSEKGFQGKLITFPERRCRNTKKPNLDLEET-----CNRRACP--AHPVYNMVAGWY  994
                gi|19171150|    954  KIQCVQKKPFQKEEAVLHSLCP-VSTP--TQVQA-----CNSHACP--PQ--------WS  995
                gi|7662202|     818  LVLCMELA-NGKPQTRSGPECGLAKKP---PEEST----CFERPCF----------KWY  858
                gi|18561227|   1163  FLKCAEKYVSGKYRELASKKCSHLPKPSLELERTPHLLLRDPRGAAGLPHPGLREVRAWP 1222
                gi|17432918|    942  VVLCKS---ADHRATLPPAHCSPAAKP--PATMR-----CNLRRCP-------P--ARWV  982

                                          1570      1580      1590      1600      1610      1620
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
                NOV3            139  S--LPWQQC------TVTCGG------GVQTRSVHCVQ----------QGRPSSS---C  170
                gi|18598706|    995  S--LPWQQC------TVTCGG------GVQTRSVHCVQ----------QGRPSSS---C 1026
                gi|19171150|    996  L--GPWSQC------SKTCGR------GVRKRELLCKGSA--------AETLPESQ---C 1030
                gi|7662202|     859  TS--PWSEC------TKTCG-------VGVRMRDVKCYQG----------TDIVRG---C  890
                gi|18561227|   1223  LTRVLEPVCPLIHLFRPTCSGSPCTVPGVAVSYQHNKPIPRRREHPPREHLTPSTSPRAC 1282
                gi|17432918|    983  AG--EWGECP------SAQCG------VGQRQRSVRCTS----------HTGQASH--EC 1015

                                          1630      1640      1650      1660      1670      1680
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
                NOV3            171  LLHQKPPVLRACN---------TNFCPAPEKRE--------------DPSCVDFFNWCH-  206
                gi|18598706|   1027  LLHQKPPVLRACN---------TNFCPAPEKRDGEMQAELDSKLSGFQTISLIWFPYSEG 1077
                gi|19171150|   1031  TSLPRPELQEGCV---------LGRCPKNSRLQ------------------WVASSWSE- 1062
                gi|7662202|    891  DPLVKPVGRQACD---------LQPCPTEPPDD-----------------SCQDQPGTN-  923
                gi|18561227|   1283  AKRKYGQKTDSCLDTLPIPLWAPSLSPSPEPRG----PEICEQ----QGLDLTECPWLLL 1334
                gi|17432918|   1016  TEALRPPTTQQCEAK------CDSPTPGDGPEE--------------CKDVNKVAYCP-  1053

                                          1690      1700      1710      1720
                                    ....|....|....|....|....|....|....|....|.
                NOV3            206  ----LVPQHGVCNHKFYGKQCCKS---CTRKI--------------  231
                gi|18598706|   1078  NERLLLPSFSLHLGGKNGIQYPKKRLPCEQKENMALAAIKMLQSTF 1123
                gi|19171150|   1062  -----VWIRSHCWVRRLRPSWLTQ---------------------- 1081
                gi|7662202|     923  --CALAIKVNLCGHWYYSKACCRS---CRPPHS------------- 951
                gi|18561227|   1335  AIG-LLPCQARDTESRPQGPVPRS--PCGQDIEK---------- 1365
                gi|17432918|   1053  ----LVLKFQFCSRAYFRQMCCKI---CQGH-------------- 1077
```

The NOV3 Clustal W alignment shown in Table 3E was modified to begin at amino residue 1321. The data in Table 3E includes all of the regions overlapping with the NOV3 protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 3F lists the domain description from DOMAIN analysis results against NOV3.

Table 3F Domain Analysis of NOV3

Model	Region of Homology	Score (bits)	E value
tsp_1	12-58	-6.8	4.1
tsp_1	66-125	14.6	0.015
tsp_1	141-187	19.3	0.0041

Consistent with other known members of the Thrombospondin type 1 (tsp_1) family of proteins, NOV3 has, for example, three tsp_1 domain signature sequences and homology to other members of the tsp_1 Domain-containing Protein Family. NOV 3 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV3 nucleic acids and polypeptides can be used to identify proteins that are members of the tsp_1 Domain-containing Protein Family. The NOV3 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV3 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, *e.g.*, cellular activation, cellular metabolism, and signal transduction. These molecules can be used to treat, *e.g.*, Von Hippel-Lindau (VHL) syndrome, diabetes, tuberous sclerosis, fertility, hypogonadism, as well as other diseases, disorders and conditions.

In addition, various NOV3 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV3 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the tsp_1 Domain-containing Protein Family.

Thrombospondin type 1 domain (TSP1, IPR000884) is a repeat found in the thrombospondin protein where it is repeated 3 times. Likewise, the tsp_1 domain is repeated three times in the NOV3 polypeptide. Now a number of proteins involved in the complement pathway (properdin, C6, C7, C8A, C8B, C9) (Patthy,L., *J. Mol. Biol.* 202: 689-696 (1988)) as well as extracellular matrix protein like mindin, F-spondin (Okamoto, *et al.*, *Development* 126: 3637-3648 (1999)), SCO-spondin and even the circumsporozoite surface protein 2 and TRAP proteins of Plasmodium (Wengelnik, *et al.*, *EMBO J.* 18: 5195-5204 (1999); Rogers, *et al.*,

Mol. Biochem. Parasitol. 53: 45-51 (1992)) contain one or more instance of this repeat. It has been involved in cell-cell interraction, inhibition of angiogenesis (Krutzsch, *et al.*, *Circulation* 100: 1423-1431 (1999)), apoptosis [Krutzsch, *et al.*, *Cancer Res.* 57: 1735-1742 (1997)).

The NOV3 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of urogenital physiology. As such, the NOV3 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat reproductive and metabolic disorders, *e.g.,* Von Hippel-Lindau (VHL) syndrome, diabetes, tuberous sclerosis, fertility, hypogonadism, as well as other diseases, disorders and conditions.

The NOV3 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV3 nucleic acid is expressed in eye and testis.

Additional utilities for NOV3 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV4

A NOV4 polypeptide has been identified as a Protocadherin Alpha C2 Short Form-like protein (also referred to as CG93187-01). The disclosed novel NOV4 nucleic acid (SEQ ID NO:7) of 600 nucleotides is shown in Table 4A. The novel NOV4 nucleic acid sequences maps to the chromosome 11.

An ORF begins with an ATG initiation codon at nucleotides 41-43 and ends with a TAG codon at nucleotides 2546-2548. A putative untranslated region and/or downstream from the termination codon is underlined in Table 4A, and the start and stop codons are in bold letters.

Table 4A. NOV4 Nucleotide Sequence (SEQ ID NO:7)

CACCATAAAAGCTCAGAAAATAGACTTTTCCTCTGCCTCT**ATG**GAGGGGCAGGCCAGATCTGGGGAA
GGGATGGGACAGCCTGGCATGAAGAGCCCCAGGCCCCACCTCCTGCTACCATTGCTGCTGCTGCTGC
TGCTGCTGCTGTCTTCGCCTCGCCGTGCACGCGTGCGCCTCCCAGAGGACCAGCCGCCTGGGCCCGC
GGCTGGCACGCTCCTAGCCCGCGACCCGCATCTGGGCGAGGCTGCACGCGTGTCCTATCGGCTGGCA
TCTGGCGGGGACGGCCACTTCCGGCTGCACTCAAGCACTGGAGCGCTGTCCGTGGTGCGGCCGTTGG
ACCGCGAACAACGAGCTGAGCACGTACTGACAGTGGTGGCCTCAGACCGAGCTCCCCGCCCGCGCTC
GGCCACGCAGGTCCTGACCGTCAGTGTCGCTGACGTCAACGACGAGGCGCCTACTTTCCAGCAGCAG
GAGTACAGCGTCCTCTTGCGTGAGAACAACCCTCCTGGCACATCTCTGCTCACCCTGCGAGCAACCG
ACCCCGACGTGGGGGCCAACGGGCAAGTGACTTATGGAGGCGTCTCTAGCGAAAGCTTTTCTCTGGA
TCCTGACACTGGTGTTCTCACGACTCTTCGGGCCCTGGATCGAGAGGAACAGGAGGAGATCAACCTG
ACAGTGTATGCCCAGGACAGGGGCTCACCTCCTCAGTTAACGCATGTCACTGTTCGAGTGGCTGTGG
AGGATGAGAATGACCATGCACCAACCTTTGGGAGTGCCCATCTCTCTCTGGAGGTGCCTGAGGGCCA
GGACCCCCAGACCCTTACCATGCTTCGGGCCTCTGATCCAGATGTGGGAGCCAATGGGCAGTTGCAG
TACCGCATCCTAGATGGGGACCCATCAGGAGCCTTTGTCCTAGACCTTGCTTCTGGAGAGTTTGGCA
CCATGCGGCCACTAGACAGAGAAGTGGAGCCAGCTTTCCAGCTGAGGATAGAGGCCCGGGATGGAGG

CCAGCCAGCTCTCAGTGCCACGCTGCTTTTGACAGTGACAGTGCTGGATGCCAATGACCATGCTCCA
GCCTTTCCTGTGCCTGCCTACTCGGTGGAGGTGCCGGAGGATGTGCCTGCAGGGACCCTGCTGCTGC
AGCTACAGGCTCATGACCCTGATGCTGGAGCTAATGGCCATGTGACCTACTACCTGGGCGCCGGTAC
AGCAGGAGCCTTCCTGCTGGAGCCCAGCTCTGGAGAACTGGTGTTGCTTGAACCTCTAGACTTTGAA
AGCCTGACACAGTACAATCTAACAGTGGCTGCAGCTGACCGTGGGCAGCCACCCCAAAGCTCAGTCG
TGCCAGTCACTGTCACTGTACTAGATGTCAATGACAACCCACCTGTCTTTACCCGAGCATCCTACCG
TGTGACAGTACCTGAGGACACACCTGTTGGAGCTGAGCTGCTGCATGTAGAGGCCTCTGACGCTGAC
CCTGCCCTCATGGCCTCCTCAGGCGACCCATCAGGGCTCTTTGAGCTGGATGAGAGCTCAGGCACCT
TGCGACTGGCCCATGCCCTGGACTGTGAGACCCAGGCTCGACATCAGCTTGTAGTACAGGCTGCTGA
CCCTGCTGGTGCACACTTTGCTTTGGCACCAGTGACAATTGAGGTCCAGGATGTGAATGATCATGGC
CCAGCCTTCCCACTGAACTTACTCAGCACCAGCGTGGCCGAGAATCAGCCTCCAGGCACTCTCGTGA
CCACTCTGCATGCAATCGACGGGGATGCTGGGGCTTTTGGGAGGCTCCGTTACAGCCTGTTGGAGGC
TGGGCCAGGACCTGAGGGCCGTGAGGCATTTGCACTGAACAGCTCAACAGGGGAGTTGCGTGCGCGA
GTGCCCTTTGACTATGAGCACACAGAAAGCTTCCGGCTGCTGGTGGGTGCTGCTGATGCTGGGAATC
TCTCAGCCTCTGTCACTGTGTCGGTGCTAGTGACTGGAGAGGATGAGTATGACCCTGTATTTCTGGC
ACCAGCTTTCCACTTCCAAGTGCCCGAAGGTGCCCGGCGTGGCCACAGCTTGGGTCACGTGCAGGCC
ACAGATGAGGATGGGGGTGCCGATGGCCTGGTTCTGTATTCCCTTGCCACCTCTTCCCCCTATTTTG
GTATTAACCAGACTACAGGAGCCCTGTACCTGCGGGTGGACAGTCGGGCACCAGGCAGCGGAACAGC
CACCTCTGGGGGGTGGGGGCCGGACCCGGCGGGAAGCACCACGGGAGCTGGGGCTCCACCTGGACTCT
TACCAGAGTCACTCCAAGTCCTGTCTCAGGCAGAATACTCAGATCTATTCCAAGCACCTTCCCTGGG
ATCTCAGGCGCATACTGAGAACCAGTGGGACAGGGTTGAGAGAGAGAGCCAACCGAGAATCTCAAAT
GAACCAAACTGAGAAAGATGCCCCTCAGTGGGGCTACAGACCGACACCCCACCATGGGGCAACAGAA
AAACCAAGACCCCCTCCCCAAAGGAATCAAACCAATCGGGAAAAGGAAGGAGGCGTTGGCCGTGCC**T**
AGGATAT

The NOV4 protein (SEQ ID NO:8) encoded by SEQ ID NO:7 is 835 amino acid residues in length and is presented using the one-letter amino acid code in Table 4B. Psort analysis predicts the NOV4 protein of the invention to be localized at the plasma membrane with a certainty of 0.7900.

Table 4B. Encoded NOV4 protein sequence (SEQ ID NO:8)

MEGQARSGEGMGQPGMKSPRPHLLLPLLLLLLLLLSSPRRARVRLPEDQPPGPAAGTLLARDPHL
GEAARVSYRLASGGDGHFRLHSSTGALSVVRPLDREQRAEHVLTVVASDRAPRPRSATQVLTVSV
ADVNDEAPTFQQQEYSVLLRENNPPGTSLLTLRATDPDVGANGQVTYGGVSSESFSLDPDTGVLT
TLRALDREEQEEINLTVYAQDRGSPPQLTHVTVRVAVEDENDHAPTFGSAHLSLEVPEGQDPQTL
TMLRASDPDVGANGQLQYRILDGDPSGAFVLDLASGEFGTMRPLDREVEPAFQLRIEARDGGQPA
LSATLLLTVTVLDANDHAPAFPVPAYSVEVPEDVPAGTLLLQLQAHDPDAGANGHVTYYLGAGTA
GAFLLEPSSGELVLLEPLDFESLTQYNLTVAAADRGQPPQSSVVPVTVTVLDVNDNPPVFTRASY
RVTVPEDTPVGAELLHVEASDADPALMASSGDPSGLFELDESSGTLRLAHALDCETQARHQLVVQ
AADPAGAHFALAPVTIEVQDVNDHGPAFPLNLLSTSVAENQPPGTLVTTLHAIDGDAGAFGRLRY
SLLEAGPGPEGREAFALNSSTGELRARVPFDYEHTESFRLLVGAADAGNLSASVTVSVLVTGEDE
YDPVFLAPAFHFQVPEGARRGHSLGHVQATDEDGGADGLVLYSLATSSPYFGINQTTGALYLRVD
SRAPGSGTATSGGGGRTRREAPRELGLHLDSYQSHSKSCLRQNTQIYSKHLPWDLRRILRTSGTG
LRERANRESQMNQTEKDAPQWGYRPTPHHGATEKPRPPPQRNQTNREKEGGVGRA

A search against the Patp database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 4C.

Table 4C. Patp results for NOV4			
Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAU07054 Human Flamingo protein	+1	968	1.8e-98
>patp:AAU07053 Human Flamingo polypeptide	+1	968	2.0e-98
>patp:ABG21921 Novel human diagnostic protein #21912	+1	642	3.1e-64
>patp:ABG21921 Novel human diagnostic protein #21912	+1	642	3.1e-64

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 273 of 415 bases (65%) identical to a gb:GENBANK-ID:AF061573|acc:AF061573.2 mRNA from *Homo sapiens* (protocadherin (PCDH8) mRNA, complete cds). The full amino acid sequence of the protein of the invention was found to have 273 of 415 amino acid residues (65%) identical to, and 273 of 415 amino acid residues (65%) similar to, the 4076 amino acid residue gb:GENBANK-ID:AF061573|acc:AF061573.2 protein from *Homo sapiens* (protocadherin (PCDH8) mRNA, complete cds).

NOV4 also has homology to the proteins shown in the BLASTP data in Table 4D.

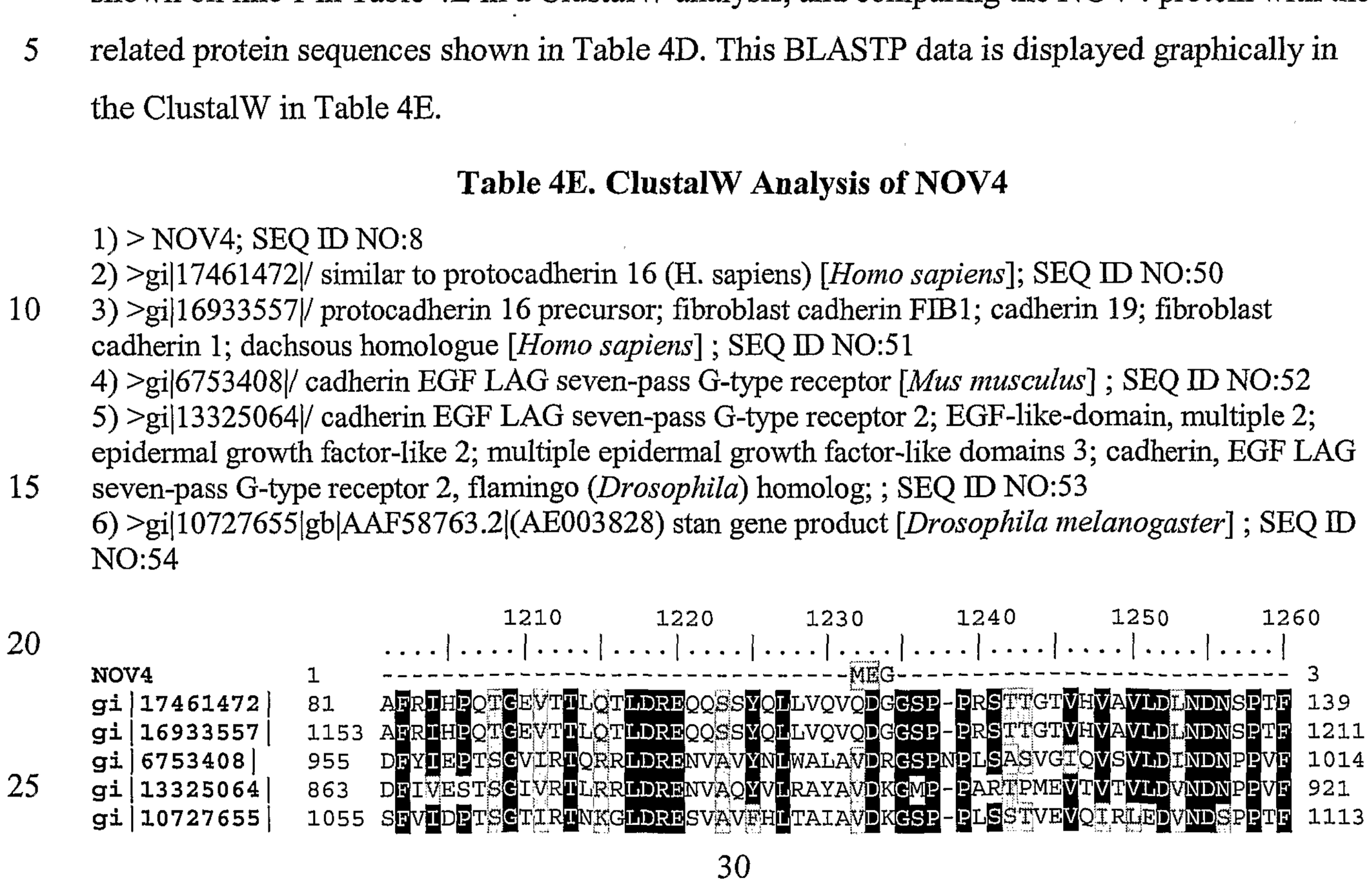
Table 4D. BLAST results for NOV4					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|17461472|ref|XP_052786.2|(XM_052786)	similar to protocadherin 16 [*Homo sapiens*]	1415	459/682 (67%)	503/682 (73%)	0.0

gi|16933557|ref|NP_003728.1|(NM_003737)	protocadherin 16 precursor; fibroblast cadherin FIB1; cadherin 19; fibroblast cadherin 1; dachsous homologue [*Homo sapiens*]		443/676 (65%)	490/676 (71%)	0.0
gi|6753408|ref|NP_034016.1|(NM_009886)	cadherin EGF LAG seven-pass G-type receptor [*Mus musculus*]	3034	48/693 (35%)	358/693 (50%)	1e-98
gi|13325064|ref|NP_001399.1|(NM_001408)	cadherin EGF LAG seven-pass G-type receptor 2; EGF-like-domain, multiple 2; epidermal growth factor-like 2; multiple epidermal growth factor-like domains 3; cadherin, EGF LAG seven-pass G-type receptor 2, flamingo (*Drosophila*)homolog	2923	246/679 (36%)	345/679 (50%)	2e-98
gi|10727655|gb|AAF58763.2| (AE003828)	stan gene product [*Drosophila melanogaster*]	3606	241/700 (34%	361/700 (51%)	3e-98

A multiple sequence alignment is given in Table 4E, with the NOV4 protein being shown on line 1 in Table 4E in a ClustalW analysis, and comparing the NOV4 protein with the related protein sequences shown in Table 4D. This BLASTP data is displayed graphically in the ClustalW in Table 4E.

Table 4E. ClustalW Analysis of NOV4

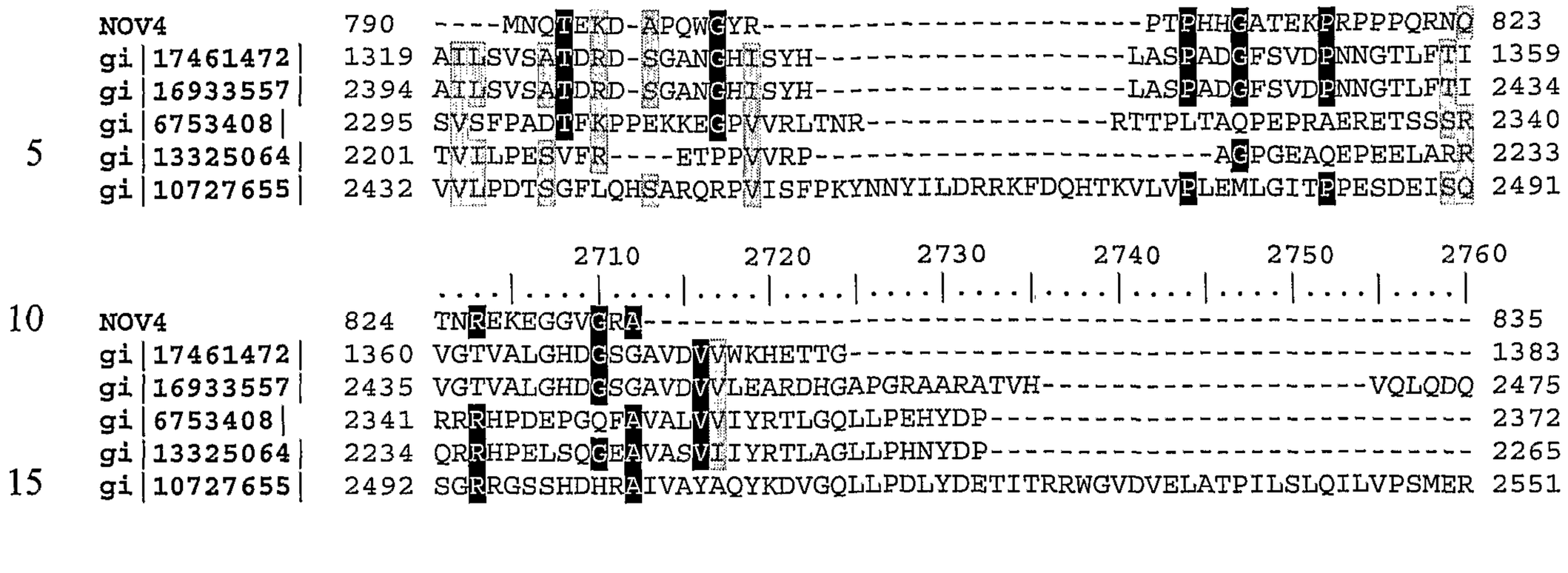
1) > NOV4; SEQ ID NO:8
2) >gi|17461472|/ similar to protocadherin 16 (H. sapiens) [*Homo sapiens*]; SEQ ID NO:50
3) >gi|16933557|/ protocadherin 16 precursor; fibroblast cadherin FIB1; cadherin 19; fibroblast cadherin 1; dachsous homologue [*Homo sapiens*] ; SEQ ID NO:51
4) >gi|6753408|/ cadherin EGF LAG seven-pass G-type receptor [*Mus musculus*] ; SEQ ID NO:52
5) >gi|13325064|/ cadherin EGF LAG seven-pass G-type receptor 2; EGF-like-domain, multiple 2; epidermal growth factor-like 2; multiple epidermal growth factor-like domains 3; cadherin, EGF LAG seven-pass G-type receptor 2, flamingo (*Drosophila*) homolog; ; SEQ ID NO:53
6) >gi|10727655|gb|AAF58763.2|(AE003828) stan gene product [*Drosophila melanogaster*] ; SEQ ID NO:54



```
                                    1270       1280       1290       1300       1310       1320
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                4   QARSGE-GMGQ----------------PGMKSP------------------------- 19
gi|17461472|        140 LQASGAAGGGLPIQVPDRVPPGTLVTTLQAKDPDEGENGTILYTLT-GPGSELFSLHPHS 198
gi|16933557|        1212 LQASGAAGGGLPIQVPDRVPPGTLVTTLQAKDPDEGENGTILYTLT-GPGSELFSLHPHS 1270
gi|6753408|         1015 EKDE------LELFVEENSPVGSVVARIRANDPDEGPNAQIIYQIVEGNVPEVFQLDLLS 1068
gi|13325064|        922  EQDE------FDVFVEENSPIGLAVARVTATDPDEGTNAQIMYQIVEGNIPEVFQLDIFS 975
gi|10727655|        1114 ASDK------ITLYVPENSPVGSVVGEIHAHDPDEGVNAVVHYSIIGGDDSNAFSLVTRP 1167

                                    1330       1340       1350       1360       1370       1380
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                19  ----------------RPHLLLPI-------------LLLLLLLLSS---------- 37
gi|17461472|        199 G----ELLTAAPLIRA-ERPHYVLTLSAHDQGSPPRSASLQLLVQVLPSARLAEPPPDLA 253
gi|16933557|        1271 G----ELLTAAPLIRA-ERPHYVLTLSAHDQGSPPRSASLQLLVQVLPSARLAEPPPDLA 1325
gi|6753408|         1069 G----DLRALVELDFE-VRRDYMLVVQATS---APLVSRATVHIRLLDQN---DNPPELP 1117
gi|13325064|        976  G----ELTALVDLDYE-DRPEYVLVIQATS---APLVSRATVHVRLLDRN---DNPPVLG 1024
gi|10727655|        1168 GSERAQLLTMTELDYESTRKRFELVVRAAS---PPLRNDAHIELLVTDVN---DNAPVLR 1221

                                    1390       1400       1410       1420       1430       1440
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                37  ------------------------------------------------------------ 37
gi|17461472|        254 ERDPAAPVPVVLTVTAAEGLRPGSLLGSVAAPEPAGVGALTYTLVGGADPEGTFALDAAS 313
gi|16933557|        1326 ERDPAAPVPVVLTVTAAEGLRPGSLLGSVAAPEPAGVGALTYTLVGGADPEGTFALDAAS 1385
gi|6753408|         1118 DFQILFNN----YVTNKSNSFPSGVIGRIPAHDPDLSDSLNYTFLQGNELS-LLLLDPAT 1172
gi|13325064|        1025 NFEILFNN----YVTNRSSSFPGGAIGRVPAHDPDISDSLTYSFERGNELS-LVLLNAST 1079
gi|10727655|        1222 DFQVIFN--------NFRDHFPSGEIGRIPAFDADVSDKLHYRILSGNNAN-LLRLNSSS 1272

                                    1450       1460       1470       1480       1490       1500
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                37  ----------------------------------------------------------P-- 38
gi|17461472|        314 GRLYLARPLDFEAGPPWRALTVRAEGPGARARGCCECRCKCRTRMSMRPPLRATRWARCQ 373
gi|16933557|        1386 GRLYLARPLDFEAGPPWRALTVRAEGPGGAGARLLRVQVQVQDENEHAPAFARDPLAIAL 1445
gi|6753408|         1173 GELQLSRDLD-NNRPLEALMEVSVSDGIHSVTALCTLRVTIITDDMLTNSITVRLENMSQ 1231
gi|13325064|        1080 GELKLSRALD-NNRPLEAIMSVLVSDGVHSVTAQCALRVTIITDEMLTHSITLRLEDMSP 1138
gi|10727655|        1273 GGLVLSPQLN-TNVPKFATMEVSVSDGINEAKAIMQLSVRLITEDMLFNSVTVRLNEMTE 1331

                                    1510       1520       1530       1540       1550       1560
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                38  ------------------------------------------------------------ 38
gi|17461472|        374 RTRSPAQRCTLSARRTPTAPAPIATCATACCARSRPCRASPGRAHRGVSSARPGPRD--- 430
gi|16933557|        1446 PE-NPEPGAALYTFRASDADGP-GPNSDVRYRLLRQEPPVPALRLDARTGALSAPRG--- 1500
gi|6753408|         1232 EKFLSPLLSLFVEGVATVLSTTKDDIFVFNIQNDTDVSS-NILNVTFSALLPGGTRG--R 1288
gi|13325064|        1139 ERFLSPLLGLFIQAVAATLATPPDHVVVFNVQRDTDAPGGHILNVSLSVGQPPGPGGGPP 1198
gi|10727655|        1332 EAFLSPLLNFFLDGLAAIIPCPKEHIFVFSIQDDTDVSS-RILNVSFSARRPDVSHE--E 1388

                                    1570       1580       1590       1600       1610       1620
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                38  ----------------------------------------------------RRARVRLPEDQPP 51
gi|17461472|        431 HSRAAAAGGSHRPARQRQPPSCSARFSAYTSRMRMRTRLSS-----PRRARVRLPEDQPP 485
gi|16933557|        1501 LDRETTPALLLLVEATDRPANASRRRAARVSARVFVTDENDNAPVFASPSRVRLPEDQPP 1560
gi|6753408|         1289 FFPSEDLQEQIYLNRTLLTTISAQRVLPFDDNICLREPCEN----YMKCVSVLRFDSSAP 1344
gi|13325064|        1199 FLPSEDLQERLYLNRSLLTAISAQRVLPFDDNICLREPCEN----YMRCVSVLRFDSSAP 1254
gi|10727655|        1389 FYTPQYLQERVYLNRAILARLATVEVLPFDDNLCVREPCLN----FEECLTVLKFGNASE 1444

                                    1630       1640       1650       1660       1670       1680
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                52  GPAAGTLLARDPHLGEAARVSYRLASGG-------DGHFRLHSSTGALSVVRPLDRE--- 101
gi|17461472|        486 GPAAGTLLARDPHLGEAARVSYRLASGG-------DGHFRLHSSTGALSVVRPLDRE--- 535
gi|16933557|        1561 GPAALHVVARDPDLGEAARVSYRLASGG-------DGHFRLHSSTGALSVVRPLDRE--- 1610
gi|6753408|         1345 FISSTTVLFRPIHPITGLRCRCPPGFTG----DYCETEIDLCYSNPCGANGRCRSREGGY 1400
gi|13325064|        1255 FIASSSVLFRPIHPVGGLRCRCPPGFTG----DYCETEVDLCYSRPCGPHGRCRSREGGY 1310
gi|10727655|        1445 FIHSDTVLFRPIYPVNTFACSCPEGFTGSKEHYLCDTEVDLCYSDPCQNGGTCVREGGY 1504

                                    1690       1700       1710       1720       1730       1740
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                101 -------QRAEHVLTVVASDR---APRPRSATQVLTVSVADVNDEAPTFQQQEYSVLLRE 151
gi|17461472|        535 -------QRAEHVLTVVASDR---APRPRSATQVLTVSVADVNDEAPTFQQQEYSVLLRE 585
gi|16933557|        1610 -------QRAEHVLTVVASDH---GSPPRSATQVLTVSVADVNDEAPTFQQQEYSVLLRE 1660
gi|6753408|         1401 TCECFEDFTGEHCQVNVRSGR-CASGVCKNGGTCVNLLIGGFHCVCPPGEYEHPYCEVST 1459
```

```
gi|13325064|  1311 TCLCRDGYTGEHCEVSARSGR-CTPGVCKNGGTCVNLLVGGFKCDCPSGDFEKPYCQVTT 1369
gi|10727655|  1505 TCVCPSTHTGQNCETGVGHLRPCPSETCEGGLSCLSNYPS-----S-QPPPYTATCELRA 1558

                                1750      1760      1770      1780      1790      1800
                      ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                   152 NNPPGTSLLTLRATDPDVGANGQVTYGGVSSESFSLDPDTGVLTTLRALDREEQEEINLT 211
gi|17461472|           586 NNPPGTSLLTLRATDPDVGANGQVTYGGVSSESFSLDPDTGVLTTLRALDREEQEEINLT 645
gi|16933557|          1661 NNPPGTSLLTLRATDPDVGANGQVTYGGVSSESFSLDPDTGVLTTLRALDREEQEEINLT 1720
gi|6753408|           1460 RSFPPQSFVTFRGLRQRFHFTVSLAFATQDRNALLLYNGR-FNEKHDFIALEIVEEQLQL 1518
gi|13325064|          1370 RSFPAHSFITFRGLRQRFHFTLALSFATKERDGLLLYNGR-FNEKHDFVALEVIQEQVQL 1428
gi|10727655|          1559 RAFGRNSFLTFESLKQRHRFNLKLRFATVQENGLLLYNGR-YNELHDFIALEIHEGHVSF 1617

                                1810      1820      1830      1840      1850      1860
                      ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                   212 VYAQDRGSPPQLTHVTVRVAVEDENDHAPTFGSAHLSLEVPEGQDPQTLTMLRASDPDVG 271
gi|17461472|           646 VYAQDRGSPPQLTHVTVRVAVEDENDHAPTFGSAHLSLEVPEGQDPQTLTMLRASDPDVG 705
gi|16933557|          1721 VYAQDRGSPPQLTHVTVRVAVEDENDHAPTFGSAHLSLEVPEGQDPQTLTMLRASDPDVG 1780
gi|6753408|           1519 TFSAGETTTTVTPQVPGGVSDGRWHSVLVQYYNKPNIGHLGLPHGPSGEKVAVVTVDDCD 1578
gi|13325064|          1429 TFSAGESTTTVSPFVPGGVSDGQWHTVQLKYYNKPLLGQTGLPQGPSEQKVAVVTVDGCD 1488
gi|10727655|          1618 SFSLGDHSERISVIQEAKVSDGKWHQVEVVYLNRSVTLVLDNCDTAIALSGQLGDRWSCA 1677

                                1870      1880      1890      1900      1910      1920
                      ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                   272 ANGQLQY----------------------R----ILDGDPSGAFVLDLASGEFG-TMRP 303
gi|17461472|           706 ANGQLQY----------------------R----ILDGDPSGAFVLDLASGEFG-TMRP 737
gi|16933557|          1781 ANGQLQY----------------------R----ILDGDPSGAFVLDLASGEFG-TMRP 1812
gi|6753408|           1579 AAVAVHFGSYVGNYSCAAQGTQSGSKKSLDLTGPLLLGGVPNLPEDFPVHSRQFVGCMRN 1638
gi|13325064|          1489 TGVALRFGSVLGNYSCAAQGTQGGSKKSLDLTGPLLLGGVPDLPESFPVRMRQFVGCMRN 1548
gi|10727655|          1678 NRTTLKLDKRCSLLTETCH-------RFLDLTGPLQVGGLPRIPAHFPVTNRDFVGCISD 1730

                                1930      1940      1950      1960      1970      1980
                      ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                   304 LDREVEPAFQLRIEARDGGQPALSATLLLTVTVLDANDHAPAFPVPAYSVEVPEDVP--- 360
gi|17461472|           738 LDREVEPAFQLRIEARDGGQPALSATLLLTVTVLDANDHAPAFPVPAYSVEVPEDVP--- 794
gi|16933557|          1813 LDREVEPAFQLRIEARDGGQPALSATLLLTVTVLDANDHAPAFPVPAYSVEVPEDVP--- 1869
gi|6753408|           1639 LSIDGRIVDMAAFIANNGTRAGCASQRNFCDGTSCQNGGTCVNRWNTYLCECPLRFGGKN 1698
gi|13325064|          1549 LQVDSRHIDMADFIANNGTVPGCPAKKNVCDSNTCHNGGTCVNQWDAFSCECPLGFGGKS 1608
gi|10727655|          1731 LRIDDRFVDLNSYVADNGTLAGCPQKAPLCQSEPCFNGGTCREGWGTYSCECPEGYAGNS 1790

                                1990      2000      2010      2020      2030      2040
                      ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                   360 -----------AGTLLLQLQAHDPDAGANGHVTYYLGAGTAGAFLLEPSSGELVLLEPLD 409
gi|17461472|           794 -----------AGTLLLQLQAHDPDAGANGHVTYYLGAGTAGAFLLEPSSGELRTAAALD 843
gi|16933557|          1869 -----------AGTLLLQLQAHDPDAGANGHVTYYLGAGTAGAFLLEPSSGELRTAAALD 1918
gi|6753408|           1699 CEQAMPHPQRFTGESVVLWSDLDITISVPWYLGLMFRTRKEDGVLMEATAGTSSRLHLQI 1758
gi|13325064|          1609 CAQEMANPQHFLGSSLVAWHGLSLPISQPWYLSLMFRTRQADGVLLQAITRGRSTITLQL 1668
gi|10727655|          1791 CQDNIPAPWRFSGDGSLSFNPLLRPIQLPWTTSFSLRTRQKEAFLLQIQICQNSSAAVCL 1850

                                2050      2060      2070      2080      2090      2100
                      ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                   410 FESLTQYNLTVAAAD-----RG-QP-PQSSVVPVTVTVLDVNDNP---------PVFTRA 453
gi|17461472|           844 REQCPSYTFSVSAVD-----GA-AAGPLSTTVSVTITVRDVNDHA---------PTFPTS 888
gi|16933557|          1919 REQCPSYTFSVSAVD-----GA-AAGPLSTTVSVTITVRDVNDHA---------PTFPTS 1963
gi|6753408|           1759 LNSYIRFEVSYGPSDVASMQLSKSRITDGGWHHLLIELRSAKEGKDIKYLAVMTLDYGMD 1818
gi|13325064|          1669 REGHVMLSVEGTGLQASSLRLEPGRANDGDWHHAQLALGASGGPG----HAILSFDYGQQ 1724
gi|10727655|          1851 RQGVLYYIFDGEP-----MYLAGAFLSDGEWHRVEIRWQQGSEIH------FS-VDYGQR 1898

                                2110      2120      2130      2140      2150      2160
                      ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                   454 SYRVTVPEDTP------VGAELLHVEASDADP--------ALMASSGD--------PSGL 491
gi|17461472|           889 PLRLRLPRPGPSFSTPTLALATLRAEDRDAGAN-----ASILYRLAGT--------PPPG 935
gi|16933557|          1964 PLRLRLPRPGPSFSTPTLALATLRAEDRDAGAN-----ASILYRLAGT--------PPPG 2010
gi|6753408|           1819 QSTVQIGNQLPGLKMRTIVIGGVTEDKVSVRHG--FRGCMQGVRMGETSTNIATLNMNDA 1876
gi|13325064|          1725 RAEGNLGPRLHGLHLSNITVGGIPGPAGGVARG--FRGCLQGVRVSDTPEGVNSLDPSHG 1782
gi|10727655|          1899 SGSVPMSQKVQGLYVGKIVMGSPDGSIGAVPEASPFEGCIQDVRIGAG----QSVLSRPT 1954

                                2170      2180      2190      2200      2210      2220
                      ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4                   492 FELDESSGTLRLAHAL------DCETQARHQLVVQAADPAG-AHFALAPVTIEVQDVN-- 542
gi|17461472|           936 TTVDSYTGEIRVARSP------VALGPRDRVLFIVATDLGRPARSATGVIIVGLQGEA-- 987
```

```
gi|16933557|   2011 TTVDSYTGEIRVARSP------VALGPRDRVLFIVATDLGRPARSATGVIIVGLQGEA-- 2062
gi|6753408|    1877 LKVRVKDGCDVEDPCASSPCPPHRPCRDTWDSYSCICDRGYFGKKCVDACLLNPCKHVAA 1936
gi|13325064|   1783 ESINVEQGCSLPDPCDSNPCPANSYCSNDWDSYSCSCDPGYYGDNCTNVCDLNPCEHQSV 1842
gi|10727655|   1955 IRENVEDGCESRAQCP-DHCPNHSSCQSSWDLSTCECDSGYVGTDCAPICTVRPCASG-V 2012

                                2230      2240      2250      2260      2270      2280
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4            542  -DHGPAFPLNLLSTSVAE---NQPPGTLVTTLHAIDGDAG-AFGRLRYSLLEAGPGPEGR 597
gi|17461472|    987  -ERGPRFPRASSEATIRE---NAPPGTPIVSPRAVH-AGG-TNGPITYSILSGNE----K 1037
gi|16933557|   2062  -ERGPRFPRASSEATIRE---NAPPGTPIVSPRAVH-AGG-TNGPITYSILSGNE----K 2112
gi|6753408|    1937  CVRSPNTPRGYSCECGPG-HYGQYCENKVDLPCPKGWWGNPVCGPCHCAVSQGFDPDCNK 1995
gi|13325064|   1843  CTRKPSAPHGYTCECPPN-YLGPYCETRIDQPCPRGWWGHPTCGPCNCDVSKGFDPDCNK 1901
gi|10727655|   2013  CRANTSLPRGYDCECNSSSRHGDYCEKELQQPCPGGWWGERVCGPCRCDLAQGYHPDCNK 2072

                                2290      2300      2310      2320      2330      2340
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4            598  EAFALNSSTGELRAR------VPFDYEHTESFRLLVGAADAGNLSASVTVSVLVTGEDEY 651
gi|17461472|   1038  GTFSIQPSTGAITVRSA----EGLDFEVSPRLRLVLQAESGGAFAFTVLTLTLQDANDN- 1092
gi|16933557|   2113  GTFSIQPSTGAITVRSA----EGLDFEVSPRLRLVLQAESGGAFAFTVLTLTLQDANDN- 2167
gi|6753408|    1996  TNGQCQCKENYYKPPAQDACLPCDCFPHGSHSRACDMDTGQCACKPGVIGRQCNRCDNPF 2055
gi|13325064|   1902  TSGECHCKENHYRPPGSPTCLLCDCYPTGSLSRVCDPEDGQCPCKPGVIGRQCDRCDNPF 1961
gi|10727655|   2073  TTGQCYCKTNHYQPPNETACLSCDCYSIGSFSGACNPLTGQCECREGVIGRRCDSCSNPY 2132

                                2350      2360      2370      2380      2390      2400
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4            652  DPVFLAPAFHFQVPEGARRGH----SLGHVQATDEDGGADGLVLYSLATSSP-------- 699
gi|17461472|   1093  APRFLRPHYVAFLPESRPLEG----PLLQVEADDLDQGSGGQISYSLAASQPARG----- 1143
gi|16933557|   2168  APRFLRPHYVAFLPESRPLEG----PLLQVEADDLDQGSGGQISYSLAASQPARG----- 2218
gi|6753408|    2056  AEVTSLGCEVIYNGCPRAFEAGIWWPQTKFGQPAAVPCPKGSVGNAVRHCSGEKG-WLPP 2114
gi|13325064|   1962  AEVTTNGCEVNYDSCPRAIEAGIWWPRTRFGLPAAAPCPKGSFGTAVRHCDEHRG-WLPP 2020
gi|10727655|   2133  AEVTLSGCEVVYDACPRSFAGGVWWPRTPLGGVAIEGCPPPARGKGQRSCDVQSGSWNTP 2192

                                2410      2420      2430      2440      2450      2460
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4            700  YFGINQTTG----------------ALYLRVDSRAP----GSGTAT-------------- 725
gi|17461472|   1144  LFHVDPTTGTITTTAILDREIWAETRLVLMATDRGSPALVGSATLTV------------- 1190
gi|16933557|   2219  LFHVDPTTGTITTTAILDREIWAETRLVLMATDRGSPALVGSATLTV------------- 2265
gi|6753408|    2115  ELFNCTSGSFVDLKALNEKLNRNETRMDGNRSLRLAKALRNATQGNS------------- 2161
gi|13325064|   2021  NLFNCTSITFSELKGFAERLQRNESGLDSGRSQQLALLLRNATQHTA------------- 2067
gi|10727655|   2193  DMYNCTSEPFVELRRQLSQLEKLELELNSFVAIKMAEQLRKACEAVDRRGASKDQKISGN 2252

                                2470      2480      2490      2500      2510      2520
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4            725  ----------------------------------------------SGGG-GRTRREAPR 738
gi|17461472|   1190  ----------------------------------------MVIDTNDNRPTIPQPW 1206
gi|16933557|   2265  ----------------------------------------MVIDTNDNRPTIPQPW 2281
gi|6753408|    2161  ------------------------------------TLFGNDVRTAYQLLARILQH 2181
gi|13325064|   2067  ------------------------------------GYFGSDVKVAYQLATRLLAH 2087
gi|10727655|   2253  GRPNRRYKMESSFLLSNGGNVWSHELEMDYLSDELKFTHDRLYGADLLVTEGLLQELINY 2312

                                2530      2540      2550      2560      2570      2580
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4            739  ELGLHLDSYQS----------------------HS----------------------KS 753
gi|17461472|   1207  ELRVSEDALLGSEIAQVTGNDVDSGPVLWYVLSPSGPQDPFSVGRYGGRVSLTGPLDFEQ 1266
gi|16933557|   2282  ELRVSEDALLGSEIAQVTGNDVDSGPVLWYVLSPSGPQDPFSVGRYGGRVSLTGPLDFEQ 2341
gi|6753408|    2182  ESRQQGFDLAAIREANFHEDVVHTGSALLAPATEASWEQIQRSE--AGAAQLLRHFEAYF 2239
gi|13325064|   2088  ESTQRGFGLSATQDVHFTENLLRVGSALLDTANKRHWELIQQTE--GGTAWLLQHYEAYA 2145
gi|10727655|   2313  ELMQSGLNLSHSQDKYFIKNLVDAASVILDRKYEAEWRRATELIQ-RGPDDLVDAFNKYL 2371

                                2590      2600      2610      2620      2630      2640
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV4            754  CLRQNTQIYSKHLPWDLRRILRTSGTGLRERANRESQ---------------------- 790
gi|17461472|   1267  CDRYQLQLLAHDGPHEGRANLTVLVEDVNDNAPAFSQSLYQVMLLEHTPPGS-------- 1318
gi|16933557|   2342  CDRYQLQLLAHDGPHEGRANLTVLVEDVNDNAPAFSQSLYQVMLLEHTPPGS-------- 2393
gi|6753408|    2240  SNVARNVKRTYLRPFVIVTANMILAVDIFDKLNFTGAQVPRFEDIQEELPRE-----LES 2294
gi|13325064|   2146  SALAQNMRHTYLSPFTIVTPNIVISVVRLDKGNFAGAKLPRYEALRGEQPPD-----LET 2200
gi|10727655|   2372  VVLARSQHDTYTSPFEIVQPNMALGLDIVTTESLFGYEPEQLSEYHRSKYLKPNAFTTES 2431

                                2650      2660      2670      2680      2690      2700
                          ....|....|....|....|....|....|....|....|....|....|....|....|
```



The NOV4 Clustal W alignment shown in Table 4E was modified to begin at amino residue 1201 and end at amino acid residue 2760. The data in Table 1E includes all of the regions overlapping with the NOV4 protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 4F lists the domain description from DOMAIN analysis results against NOV4.

Table 4F Domain Analysis of NOV4

Model	Region of Homology	Score (bits)	E value
cadherin	41-131	97.7	2.4e-25
T2SP_N	16-223	-117.2	1.3
cadherin	145-233	104.1	2.7e-27
cadherin	247-337	78.1	1.8e-19
cadherin	351-441	112.9	6e-30
cadherin	455-539	64.7	2e-15
cadherin	553-646	77.5	2.8e-19
cadherin	660-745	15.4	0.036

Consistent with other known members of the Protocadherin Alpha C2 Short Form Protein-like family of proteins, NOV4 has, for example, seven Cadherin domain signature sequences and homology to other members of the Protocadherin Alpha C2 Short Form Protein-like Protein Family. NOV4 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV4 nucleic acids and polypeptides can be used to identify proteins that are members of the Protocadherin Alpha C2 Short Form Protein-like Protein Family. The NOV4 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV4 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, *e.g.*, cellular

activation, cellular metabolism, and signal transduction. These molecules can be used to treat, *e.g.*, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS, fertility, endometriosis, hypogonadism, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease (GVHD), lymphaedema, as well as other diseases, disorders and conditions.

In addition, various NOV4 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV4 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the Protocadherin Alpha C2 Short Form Protein-like Protein Family.

Cadherins (Takeichi, *Annu. Rev. Biochem.* 59: 237-252 (1990); Takeichi Trends Genet. 3: 213-217 (1987)), first discovered in mouse teratocarcinoma cells (Liaw, *EMBO J.* 9: 2701-2708 (1990)), are a family of animal glycoproteins responsible for calcium-dependent cell-cell adhesion. Cadherins preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand. There are a number of different isoforms distributed in a tissue-specific manner in a wide variety of organisms. Cells containing different cadherins tend to segregate *in vitro*, while those that contain the same cadherins tend to preferentially aggregate together. This observation is linked to the finding that cadherin expression causes morphological changes involving the positional segregation of cells into layers, suggesting they may play an important role in the sorting of different cell types during morphogenesis, histogenesis and regeneration. They may also be involved in the regulation of tight and gap junctions, and in the control of intercellular spacing. Cadherins are evolutionary related to the desmogleins which are component of intercellular desmosome junctions involved in the interaction of plaque proteins.

Structurally, cadherins comprise a number of domains: these include a signal sequence; a propeptide of around 130 residues; an extracellular domain of around 600 residues; a single transmembrane domain; and a well-conserved C-terminal cytoplasmic domain of about 150 residues. The extracellular domain can be subdivided into 5 parts, 4 of which are repeats of about 110 residues, and the fifth contains 4 conserved cysteines. The calcium-binding region of cadherins is thought to be located in the extracellular domain. This indicates that the

sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Maniatis *et al.* has identified 52 novel human cadherin-like genes organized into three closely linked clusters (Wu and Maniatis, *Cell* 97(6):779-90 (1999).) Comparison of the genomic DNA sequences with those of representative cDNAs reveals a striking genomic organization similar to that of immunoglobulin and T cell receptor gene clusters. The N-terminal extracellular and transmembrane domains of each cadherin protein are encoded by a distinct and unusually large exon. These exons are organized in a tandem array. By contrast, the C-terminal cytoplasmic domain of each protein is identical and is encoded by three small exons located downstream from the cluster of N-terminal exons. This unusual organization has interesting implications regarding the molecular code required to establish complex networks of neuronal connections in the brain and the mechanisms of cell-specific cadherin-like gene expression.

The NOV4 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of urogenital, nerve, and endocrine physiology. As such, the NOV4 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat reproductive and nervous system disorders, *e.g.,* Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS, fertility, endometriosis, hypogonadism, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease (GVHD), lymphaedema, as well as other diseases, disorders and conditions.

The NOV4 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV4 nucleic acid is expressed in Heart, Aorta, Umbilical Vein, Thyroid, Colon, Peripheral Blood, Spleen, Lymph node, Bone, Cartilage, Brain, Left cerebellum, Right Cerebellum, Parietal Lobe, Temporal Lobe, Cerebral Medulla/Cerebral white matter, Hippocampus, Cervix, Mammary gland/Breast, Ovary, Placenta, Uterus, Testis, Lung, and Retina.

Additional utilities for NOV4 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV5

A NOV5 polypeptide has been identified as a Nuclear protein-like protein (also referred to as CG95083-01). The disclosed novel NOV5 nucleic acid (SEQ ID NO:9) of 2322 nucleotides is shown in Table 5A.

An ORF begins with an ATG initiation codon at nucleotides 70-72 and ends with a TAA codon at nucleotides 2320-2322. A putative untranslated region and/or downstream from the termination codon is underlined in Table 5A, and the start and stop codons are in bold letters.

Table 5A. NOV5 Nucleotide Sequence (SEQ ID NO:9)

GTGGAAGGACAGTCCAGAGCCCTTGTCATCGCACAGGAACTGCTATCTTCAGAGAAAGCATACGTGG
AG**ATG**CTCCAGCACTTAAATCTGTTCCTGGCAGAGCAGGCTATCAGCAGGAGAGGCCAGGGCTCCAA
AGCCCCAGGGGAAATCTGCCAAGGAGGACTTGTGCTCAGTCCTATCAACCTGTGGGTAACAGACCTT
TTGGTGTTTCAGGATTTCCATGGAGCTGTCATGAGGGCCTTGGATGACATGGACCATGAAGGCAGAG
ACACATTGGCCCGGGAGGAGCTGAGGCAGGGCCTGAGTGAACTCCCAGCCATCCACGACCTTCATCA
AGGCATCCTGGAGGAGCTGGAGGAAAGGCTGTCAAATTGGGAGAGCCAGCAGAAGGTAGCTGACGTC
TTCCTTGCCCGGGAGCAGGGGTTTGATCACCACGCCACTCACATCCTGCAGTTCGACAGGTACCTAG
GTCTGCTCAGTGAGAATTGCCTCCACTCTCCCCGGCTGGCAGCTGCTGTCCGTGAATTTGAGCAGAG
TGTACAAGGAGGCAGCCAGACTGCGAAGCATCGGCTGCTGCGGGTGGTTCAACGCCTCTTCCAGTAC
CAAGTGCTCCTCACAGACTATTTAAACAACCTTTGTCCGGACTCCGCCGAGTACGACAACACACAGG
GTGCACTGAGCCTCATCTCCAAAGTCACAGACCGTGCCAACGACAGCATGGAGCAAGGGGAAAACCT
GCAGAAGCTGGTCCACATTGAGCACAGCGTCCGGGGCCAAGGGGATCTCCTCCAGCCAGGAAGGGAG
TTTCTGAAGGAAGGGACGCTGATGAAAGTAACAGGGAAAAACAGACGGCCCCGGCACCTATTTCTGA
TGAACGATGTGCTCCTGTACACCTATCCCCAGAAGGATGGGAAGTACCGGCTGAAGAACACATTGGC
TGTGGCCAACATGAAGGCTCTTTACCATGGGGAAGGGGAAGGAGGAAGCACCTTTCTCAGCATGGAG
GTTTGTTCCCTTTTGGAACCAAAGGCTCCACCGAGGAGCCTGTTAGAAAAAGGCATGGGAGACGTGG
TCACTGGCAGGTACTTGTCCAACATGACAGTGCACCTGGGGTTGCCCGGGCTGGGCCCTGAGCATGA
CGCTCTGCAGCCTTCCCAGCGGTGGGTCAGCCGCCCTGTGATGGAGAAAGTGCCCTACGCTCTAAAG
ATTGAGACTTCCGAGTCCTGCCTGATGCTGTCTGCGAGGCTGCAGGTCAGGAAGTCCAAGGTCAAGG
CACTGACTGATTCGGTGTCTGCAGCCCTGGGAGTTAGGGGAATATCATTATTCCAGTGTAAGAAGAA
ACAGACCCAAGGACAGCTAATGGACCAGTGGTCTGCTCGTAAACCTAGTCTGGCAGGTGATCTCTTC
TTTGCTGGTGGTTCTGGGCAGTGTGAGAGGTGCAGGCTCAAGGGGCATCTGAGTGAGAACCTCATCC
ATGCCGAGATGGAGGCCCATGCCCGCAGCTCCTGTGCAGAGAGGGACGAGTGGTATGGCTGTCTGAG
CAGAGCCCTCCCTGAGGACTACAAGGCCCAGGCGCTGGCTGCATTCCACCATAGCGTGGAGATACGA
GAGAGGCTGGGGGTTAGCCTTGGGGAGAGGCCCCCCACCCTGGTGCCTGTCACACACGTCATGATGT
GCATGAACTGCGGCTGCGACTTCTCCCTCACCCTGCGGCGTCATCACTGTCACGCCTGTGGCAAGCA
GATCGTGTGCCGGAACTGTTCGCGGAACAAGTACCCGCTGAAGTACCTGAAGGACAGGATGGCCAAG
GTCTGCGACGGCTGCTTCGGGGAGCTGAAGAAGCGGGGCAGGGCTGTCCCGGGCCTGATGAGAGTTA
CAGAGCGGCCTGTGAGCATGAGCTTCCCGCTGTCTTCACCCCGCTTCTCGGGCAGTGCCTTTTCATC
CGTCTTCCAGAGCATTAACCCCTCGACCTTCAAGAAGCAGAAGAAAGTCCCTTCAGCCCTGACAGAG
GTAGCTGCCTCTGGAGAGGGCTCTGCCATCAGTGGCTATCTCAGCCGGTGTAAGAGGGGCAAGCGGC
ACTGGAAGAAGCTCTGGTTTGTCATCAAAGGCAAAGTTCTCTACACCTACATGGCCAGTGAGGACAA
AGTGGCCTTGGAGAGTATGCCTCTGCTAGGCTTCACCATTGCTCCAGAAAAGGAAGAGGGCAGCAGT
GAAGTAGGACCTATTTTTCACCTTTACCACAAGAAAACCCTATTTTATAGCTTCAAAGCAGAAGATA
CCAATTCATGGATCGAGGCCATGGAAGATGCGAGTGTGTTA**TAG**

Variant sequences of NOV5 are included in Example 3, Table 19. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA.

The NOV5 protein (SEQ ID NO:10) encoded by SEQ ID NO:9 is 750 amino acid residues in length and is presented using the one-letter amino acid code in Table 5B. Psort analysis predicts the NOV5 protein of the invention to be localized in the nucleus with a certainty of 0.3000.

Table 5B. Encoded NOV5 protein sequence (SEQ ID NO:10)

```
MLQHLNLFLAEQAISRRGQGSKAPGEICQGGLVLSPINLWVTDLLVFQDFHGAVMRALDDMDHEG
RDTLAREELRQGLSELPAIHDLHQGILEELEERLSNWESQQKVADVFLAREQGFDHHATHILQFD
RYLGLLSENCLHSPRLAAAVREFEQSVQGGSQTAKHRLLRVVQRLFQYQVLLTDYLNNLCPDSAE
YDNTQGALSLISKVTDRANDSMEQGENLQKLVHIEHSVRGQGDLLQPGREFLKEGTLMKVTGKNR
RPRHLFLMNDVLLYTYPQKDGKYRLKNTLAVANMKALYHGEGEGGSTFLSMEVCSLLEPKAPPRS
LLEKGMGDVVTGRYLSNMTVHLGLPGLGPEHDALQPSQRWVSRPVMEKVPYALKIETSESCLMLS
ARLQVRKSKVKALTDSVSAALGVRGISLFQCKKKQTQGQLMDQWSARKPSLAGDLFFAGGSGQCE
RCRLKGHLSENLIHAEMEAHARSSCAERDEWYGCLSRALPEDYKAQALAAFHHSVEIRERLGVSL
GERPPTLVPVTHVMMCMNCGCDFSLTLRRHHCHACGKQIVCRNCSRNKYPLKYLKDRMAKVCDGC
FGELKKRGRAVPGLMRVTERPVSMSFPLSSPRFSGSAFSSVFQSINPSTFKKQKKVPSALTEVAA
SGEGSAISGYLSRCKRGKRHWKKLWFVIKGKVLYTYMASEDKVALESMPLLGFTIAPEKEEGSSE
VGPIFHLYHKKTLFYSFKAEDTNSWIEAMEDASVL
```

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

Table 5C. Patp results for NOV5

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAB93568 Human protein sequence SEQ ID NO:12972	+1	577	1.7e-95
>patp:AAY51248 Rat actin-binding protein frabin	+1	312	1.9e-41
>patp:AAU21630 Novel human neoplastic disease polypeptide	+1	256	1.6e-38
>patp:AAU27818 Human full-length polypeptide #143	+1	300	2.6e-29
>patp:ABG00573 Novel human diagnostic protein #564	+1	261	1.8e-26

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 443 of 754 bases (58%) identical to a gb:GENBANK-ID:AB037783|acc:AB037783.1 mRNA from *Homo sapiens* (mRNA for KIAA1362 protein, partial cds). The full amino acid sequence of the protein of the invention was found to have 114 of 263 amino acid residues (43%) identical to, and 173 of 263 amino acid residues (65%) similar to, the 699 amino acid residue ptnr:SPTREMBL-ACC:Q9P2I5 protein from *Homo sapiens* (KIAA1362 PROTEIN).

NOV5 also has homology to the proteins shown in the BLASTP data in Table 5D.

Table 5D. BLAST results for NOV5

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|8922921|ref|NP_060821.1|(NM_018351)	hypothetical protein FLJ11183 [*Homo sapiens*]	432	135/284 (47%)	169/284 (58%)	5e-57
gi|16716345|ref|NP_444302.1|(NM_053072)	ethanol decreased 4 [*Mus musculus*]	431	131/284 (46%)	171/284 (60%)	2e-55
gi|7243105|dbj|BAA92600.1|(AB037783)	KIAA1362 protein [*Homo sapiens*]	699	111/251 (44%)	166/251 (65%)	2e-54
gi|13648298|ref|XP_012133.2|(XM_012133)	hypothetical protein FLJ11183 [*Homo sapiens*]	204	115/222 (51%)	141/222 (62%),	1e-49
gi|15426438|gb|AAH13319.1|AAH13319 (BC013319)	Similar to hypothetical protein FLJ11183 [*Homo sapiens*]	376	103/221 (46%)	129/221 (57%)	4e-40

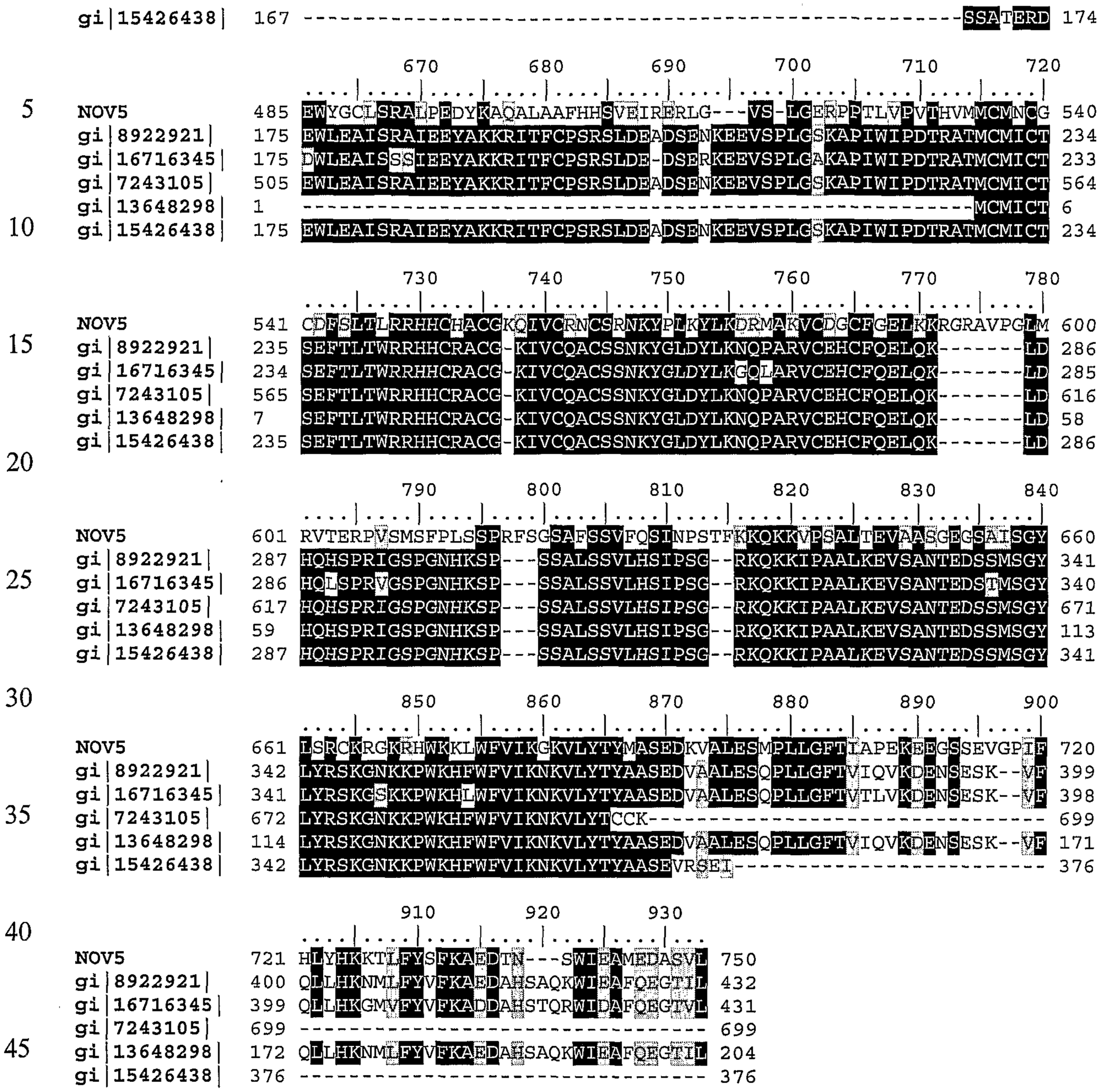
A multiple sequence alignment is given in Table 5E, with the NOV5 protein being shown on line 1 in Table 5E in a ClustalW analysis, and comparing the NOV5 protein with the related protein sequences shown in Table 5D. This BLASTP data is displayed graphically in the ClustalW in Table 5E.

Table 5E. ClustalW Analysis of NOV5

1) > NOV5; SEQ ID NO:10
2) >gi|8922921|/ hypothetical protein FLJ11183 [*Homo sapiens*]; SEQ ID NO:55
3) >gi|16716345|/ ethanol decreased 4 [*Mus musculus*]; SEQ ID NO:56
4) >gi|7243105|/ KIAA1362 protein [*Homo sapiens*]; SEQ ID NO:57
5) >gi|13648298|/ hypothetical protein FLJ11183 [*Homo sapiens*]; SEQ ID NO:58
6) >gi|15426438|/ Similar to hypothetical protein FLJ11183 [*Homo sapiens*]; SEQ ID NO:59

```
                                10        20        30        40        50        60
                       ....|....|....|....|....|....|....|....|....|....|....|....|
NOV5                1  ------------------------------------------------------------ 1
gi|8922921|         1  ------------------------------------------------------------ 1
gi|16716345|        1  ------------------------------------------------------------ 1
gi|7243105|         1  GIESDWQGLLVGEEKRSKPIKAYSTENYSLESQKKRKKSRGQTSAANGLRAESLDDQMLS 60
gi|13648298|        1  ------------------------------------------------------------ 1
gi|15426438|        1  ------------------------------------------------------------ 1

                                70        80        90       100       110       120
                       ....|....|....|....|....|....|....|....|....|....|....|....|
NOV5                1  ----------------------------------MLQHLNLFLAEQAISRRGQGS------ 21
gi|8922921|         1  ------------------------------------------------------------ 1
gi|16716345|        1  ------------------------------------------------------------ 1
gi|7243105|        61  RESSSQAPYKSVTSLCAPEYENIRHYEEIPEYENLPFIMAIRKTQELEWQNSSSMEDADA 120
gi|13648298|        1  ------------------------------------------------------------ 1
gi|15426438|        1  ------------------------------------------------------------ 1

                               130       140       150       160       170       180
                       ....|....|....|....|....|....|....|....|....|....|....|....|
NOV5               21  -----KAPGEICQGGLVLSP---------------------------------------- 36
gi|8922921|         1  ------------------------------------------------------------ 1
gi|16716345|        1  ------------------------------------------------------------ 1
gi|7243105|       121  NVYEVEEPYEAPDGQLQLGPRHQHSSSGASQEEQNDLGLGDLPSDEEEIINSSDEDDVSS 180
gi|13648298|        1  ------------------------------------------------------------ 1
gi|15426438|        1  ------------------------------------------------------------ 1
```

```
                                  190       200       210       220       230       240
                         ....|....|....|....|....|....|....|....|....|....|....|....|
NOV5                 36  ----------------------INLWVTDLLVFQDFHGAVMRAL-----DDMDHE---- 64
gi|8922921|          1   ------------------------------------------------------------ 1
gi|16716345|         1   ------------------------------------------------------------ 1
gi|7243105|          181 ESSKGEPDPLEDKQDEDNGMKSKVHHIAKEIMSSEKVFVDVLKLLHIDFRDAVAHASRQL 240
gi|13648298|         1   ------------------------------------------------------------ 1
gi|15426438|         1   ------------------------------------------------------------ 1

                                  250       260       270       280       290       300
                         ....|....|....|....|....|....|....|....|....|....|....|....|
NOV5                 65  GRDTLAREELRQGLSELPAIHDLHQGILEELEERLSNWESQQKVADVFLAREQGFDHHAT 124
gi|8922921|          1   ------------------------------------------------------------ 1
gi|16716345|         1   ------------------------------------------------------------ 1
gi|7243105|          241 GKPVIEDRILNQILYYLPQLYELNRDLLKELEERMLHWTEQQRIADIFVKKGPYLKMYST 300
gi|13648298|         1   ------------------------------------------------------------ 1
gi|15426438|         1   ------------------------------------------------------------ 1

                                  310       320       330       340       350       360
                         ....|....|....|....|....|....|....|....|....|....|....|....|
NOV5                 125 HILQFDRYLGLLSENCLHSPRLAAAVREFEQSVQGGSQTAKHRLLRVVQRLFQYQVLLTD 184
gi|8922921|          1   ----------------------------MSPRCANLALKHYLLKPVQRIPQYRLLLTD 30
gi|16716345|         1   ----------------------------MSPRCANLALKHYLLKPVQRIPQYRLLLTD 30
gi|7243105|          301 YIKEFDKNIALLDEQCKKNPGFAAVVREFEMSPCCANLALKHYLLKPVQRIPQYRLLLTD 360
gi|13648298|         1   ------------------------------------------------------------ 1
gi|15426438|         1   ----------------------------MSPRCANLALKHYLLKPVQRIPQYRLLLTD 30

                                  370       380       390       400       410       420
                         ....|....|....|....|....|....|....|....|....|....|....|....|
NOV5                 185 YLNNLCPDSAEYDNTQGALSLISKVTDRANDSMEQGENLQKLVHIEHSVRGQGDLLQPGR 244
gi|8922921|          31  YLKNLIEDAGDYRDTQDALAVVIEVANHANDTMKQGDNFQKLMQIQYSLNGHHEIVQPGR 90
gi|16716345|         31  YLKNLLEDSVDHRDTQDALAVVIEVANHANDTMKQGDNFQKLMQIQYSLSGHHEIVQPGR 90
gi|7243105|          361 YLKNLIEDAGDYRDTQDALAVVIEVANHANDTMKQGDNFQKLMQIQYSLNGHHEIVQPGR 420
gi|13648298|         1   ------------------------------------------------------------ 1
gi|15426438|         31  YLKNLIEDAGDYRDTQDALAVVVEVANHANDTMKQGDNFQKLMQIQYSLNGHHEIVQPGR 90

                                  430       440       450       460       470       480
                         ....|....|....|....|....|....|....|....|....|....|....|....|
NOV5                 245 EFLKEGTLMKVTGKNRRPRHLFLMNDVLLYTYPQKDGKYRLKNTLAVANMKALYHGEGEG 304
gi|8922921|          91  VFLKEGILMKLSRKVMQPRMFFLFNDALLYTTPVQSGMYKLNNMLSLAGMK--------- 141
gi|16716345|         91  VFLKEGTLMKLSRKVMQPRMFFLFNDALLYTTPMQSGMYKLNNMLSLAGMK--------- 141
gi|7243105|          421 VFLKEGILMKLSRKVMQPRMFFLFNDALLYTTPVQSGMYKLNNMLSLAGMK--------- 471
gi|13648298|         1   ------------------------------------------------------------ 1
gi|15426438|         91  VFLKEGILMKLSRKVMQPRMFFLFNDALLYTTPVQSGMYKLNNMLSLAGMK--------- 141

                                  490       500       510       520       530       540
                         ....|....|....|....|....|....|....|....|....|....|....|....|
NOV5                 305 GSTFLSMEVCSLLEPKAPPRSLLEKGMGDVVTGRYLSNMTVHLGLPGLGPEHDALQPSQR 364
gi|8922921|          141 ------------------------------------------------------------ 141
gi|16716345|         141 ------------------------------------------------------------ 141
gi|7243105|          471 ------------------------------------------------------------ 471
gi|13648298|         1   ------------------------------------------------------------ 1
gi|15426438|         141 ------------------------------------------------------------ 141

                                  550       560       570       580       590       600
                         ....|....|....|....|....|....|....|....|....|....|....|....|
NOV5                 365 WVSRPVMEKVPYALKIETSESCLMLSARLQVRKSKVKALTDSVSAALGVRGISLFQCKKK 424
gi|8922921|          141 -VRKPTQEAYQNELKIESVERSFILSA---------------------------------- 167
gi|16716345|         141 -VRKPTQEAYQNELKIESVERSFILSA---------------------------------- 167
gi|7243105|          471 -VRKPTQEAYQNELKIESVERSFILSA---------------------------------- 497
gi|13648298|         1   ------------------------------------------------------------ 1
gi|15426438|         141 -VRKPTQEAYQNELKIESVERSFILSA---------------------------------- 167

                                  610       620       630       640       650       660
                         ....|....|....|....|....|....|....|....|....|....|....|....|
NOV5                 425 QTQGQLMDQWSARKPSLAGDLFFAGGSGQCERCRLKGHLSENLIHAEMEAHARSSCAERD 484
gi|8922921|          167 ----------------------------------------------SSATERD 174
gi|16716345|         167 ----------------------------------------------SSAAERD 174
gi|7243105|          497 ----------------------------------------------SSATERD 504
gi|13648298|         1   ------------------------------------------------------------ 1
```



The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 5F lists the domain description from DOMAIN analysis results against NOV5.

Table 5F Domain Analysis of NOV5

Model	Region of Homology	Score (bits)	E value
RhoGEF	33-215	-1.9	1.2e-05
FYVE Ring Finger	525-591	55.6	6.4e-14
Plekstrin (PH)	657-748	49.3	8.0e-7

Consistent with other known members of the Nuclear Protein-like family of proteins, NOV5 has, for example, an RhoGEF signature sequence and a FYVE Zinc Finger signature sequence, aw well as homology to other members of the Nuclear Protein-like Protein Family. NOV5 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV5 nucleic acids and polypeptides can be used to identify proteins that are members of the Nuclear Protein-like Protein Family. The NOV5 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV5 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, *e.g.,* cellular activation, cellular replication, and signal transduction. These molecules can be used to treat, *e.g.*, Cardiovascular diseases, 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, Diabetes,Von Hippel-Lindau (VHL) syndrome , Pancreatitis, Obesity, Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection as well as other diseases, disorders and conditions.

In addition, various NOV5 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV5 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the Nuclear Protein-like Protein Family.

The NOV5 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of cardiac and nerve physiology. As such, the NOV5 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat cardiovascular and nervous system disorders, *e.g.,* Cardiovascular diseases, 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, Diabetes,Von Hippel-Lindau (VHL)

syndrome , Pancreatitis, Obesity, Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection as well as other diseases, disorders and conditions.

The NOV5 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV5 nucleic acid is expressed in Brown adipose, Vein, Umbilical Vein, Adrenal Gland/Suprarenal gland, Gall Bladder, Small Intestine, Colon, Lymphoid tissue, Spleen, Lymph node, Thymus, Brain, Temporal Lobe, Basal Ganglia/Cerebral nuclei, Substantia Nigra, Spinal Chord, Cervix, Ovary, Uterus, Testis, Lung, Lung Pleura, Larynx, Urinary Bladder, Kidney.

Additional utilities for NOV5 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV6

A NOV6 polypeptide has been identified as a Secretory Protein-like protein (also referred to as CG94989-01). The disclosed novel NOV6 nucleic acid (SEQ ID NO:11) of 2372 nucleotides is shown in Table 6A. The novel NOV6 nucleic acid sequences maps to the chromosome 17.

An ORF begins with an ATG initiation codon at nucleotides 99-101 and ends with a TAA codon at nucleotides 1710-1712. A putative untranslated region and/or downstream from the termination codon is underlined in Table 6A, and the start and stop codons are in bold letters.

Table 6A. NOV6 Nucleotide Sequence (SEQ ID NO:11)

CCGGCAAGGATGACGCCTCCGGAGGCCCTGGCCTCACTCCCACCTGGGCGCTAGGAGCCATCCCGGG
GCTCCAGCCAGGAGCCCTGCTGCCCAGGGGC**ATG**GCCAAACCTTTCTTCCGACTCCAGAAGTTTCTC
CGCCGAACACAGTTCCTGCTGTTCTTCCTCACGGCTGCCTACCTGATGACCGGCAGCCTGCTGCTGC
TGCAGCGGGTCCGCGTGGCTCTCCCACAGGGCCCCCGGGCACCCGGCCCCCTGCAGACCTTGCCAGT
GGCCGCCGTGGCGCTGGGCGTGGGCTTGCTGGACAGCAGAGCCCTGCACGACCCTCGAGTCAGCCCA
GAGCTGCTGCTGGGTGTGGACATGCTGCAGAGCCCCCTGACCCGGCCCCGGCCCGGCCCCCGCTGGC
TCCGGAGCCGCAACTCGGAGCTGCGTCAGTTGCGTCGCCGCTGGTTCCACCACTTCATGAGTNGACT
CCCAGGGACCGCCCGCCCTGGGCCCCGAGGCTGCCAGGCCCGCCATCCACAGCCGAGGTCCTATGTC
TACGCCGGCTTGGAGGCCGGGGCGGAGTGTTACTGCGGGAACCGGCTGCCAGCGGTGAGCGTGGGGC
TGGAAGAGTGTAACCATGAGTGCAAAGGCGAGAAGGGCTCTGTGTGCGGGGCTGTGGACCGGCTCTC
CGTGTACCGTGTGGACGAGCTGCAGCCGGGCTCCAGGAAGCGGCGGACCGCCACCTACCGCGGATGC
TTCCGACTGCCAGAGAACATCACACATGCCTTCCCCAGCTCCCTGATACAGGCCAATGTGACCGTGG
GGACTTGCTCGGGCTTTTGTTCCCAGAAAGAGTTCCCCTTGGCCATTCTCAGGGGCTGGGAATGCTA
CTGTGCTTACCCTACCCCCCGGTTCAACCTGCGGGATGCCATGGACAGCTCAGTATGTGGCCAGGAC
CCTGAGGCACAGAGGCTGGCAGAATACTGTGAGGTCTACCAGACACCTGTGCAAGACACTCGTTGTA
CAGACAGGAGGTTCCTGCCTAACAAATCCAAAGTGTTTGTGGCTTTGTCAAGCTTCCCAGGAGCCGG
GAACACGTGGGCACGGCACCTCATTGAGCATGCCACTGGCTTCTATACAGGGAGCTACTACTTTGAT

GGAACCCTCTACAACAAAGGGTTCAAGGGCGAAAAGGACCACTGGCGGAGCCGACGCACCATCTGTG
TCAAAACCCACGAGAGTGGCAGGAGGGAGATTGAGATGTTTGATTCAGCCATCCTGCTAATCCGGAA
CCCATACAGGTCCCTGGTGGCAGAATTCAACAGAAAATGTGCCGGGCACCTGGGATATGCAGCTGAC
CGCAACTGGAAGAGCAAAGAGTGGCCGGACTTTGTCAACAGCTACGCCTCGTGGTGGTCCTCGCACG
TCCTGGACTGGCTCAAGTACGGGAAGCGGCTGCTGGTGGTGCACTACGAGGAGCTGCGGCGCAGCCT
GGTGCCCACGTTACGGGAGATGGTGGCCTTCCTCAACGTGTCTGTGAGCGAGGAGCGGCTGCTCTGC
GTGGAGAACAACAAGGAGGGCAGCTTCCGGCGGCGCGGCCGGCGCTCCCACGACCCTGAGCCCTTCA
CCCCGGAGATGAAAGACTTGATCAATGGCTACATCCGGACGGTGGACCAAGCCCTGCGTGACCACAA
CTGGACGGGGCTGCCCAGGGAGTATGTGCCCAGA**TGA**TAGGCCTGGCCCACGCCGCCGCCCCCGCTG
AGTGACGCAATCGCACCACGGGGCTGCGCTCCCCACTCTGATGCTCAGGCCCGTGGCCTCACTGGGA
CGAACGGTGGGTGGGGGGCTCACCCTGGTGCTGCCTCCCGCACAAGGAGACCTGGACACAACAGACA
CACATCACAAGGCGAACACAAATGGACACACATACCTGGCCACGAACCCACACCTCCTCAGACACTC
AGACACCACTCCAGGCTCATAGCCCCGTCTTGATGCAGAGAAGCCACCCACGTGGGGTGTGCCAGGC
ACCCCCAGCTACAAATGCAGCCACGCACAGACGTAACACACAGGTGCCAGGCCGTGTGCTCCTGGAG
GCTGGCTGGCTGTCTCTCTCACACAGATACACGTGCGCTCCCTGGGATCCGGGAGGCCCTGGGCTTC
CTGTGTGTAGCCCTGGCATAGACTTGCTCGTCAGGGTGTTTGACTCTGGGATGCTGGGCCGGGCAGA
CATTTATGCTCTGAGCAGCAAGGACCATTGGGATGGAGGTGGGCACAAAGACTGCTGCTTCCAGGGT
GTGCGGCCCTGGCCGTGTGTCTGACATCCCATAAATGTGTGTGTGGTGTGACTACGGGCACCACAAA
CTCCGCAAAAAAAAAAAAAAAAAAAA

The NOV6 protein (SEQ ID NO:12) encoded by SEQ ID NO:11 is 537 amino acid residues in length and is presented using the one-letter amino acid code in Table 6B. Psort analysis predicts the NOV6 protein of the invention to be localized outside the cell with a certainty of 0.6997.

Table 6B. Encoded NOV6 protein sequence (SEQ ID NO:12)

MAKPFFRLQKFLRRTQFLLFFLTAAYLMTGSLLLLQRVRVALPQGPRAPGPLQTLPVAAVALGVG
LLDSRALHDPRVSPELLLGVDMLQSPLTRPRPGPRWLRSRNSELRQLRRRWFHHFMSXLPGTARP
GPRGCQARHPQPRSYVYAGLEAGAECYCGNRLPAVSVGLEECNHECKGEKGSVCGAVDRLSVYRV
DELQPGSRKRRTATYRGCFRLPENITHAFPSSLIQANVTVGTCSGFCSQKEFPLAILRGWECYCA
YPTPRFNLRDAMDSSVCGQDPEAQRLAEYCEVYQTPVQDTRCTDRRFLPNKSKVFVALSSFPGAG
NTWARHLIEHATGFYTGSYYFDGTLYNKGFKGEKDHWRSRRTICVKTHESGRREIEMFDSAILLI
RNPYRSLVAEFNRKCAGHLGYAADRNWKSKEWPDFVNSYASWWSSHVLDWLKYGKRLLVVHYEEL
RRSLVPTLREMVAFLNVSVSEERLLCVENNKEGSFRRRGRRSHDPEPFTPEMKDLINGYIRTVDQ
ALRDHNWTGLPREYVPR

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

Table 6C. Patp results for NOV6

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:ABB15485 *Human nervous system related polypeptide*	+1	92	0.0036
>patp:AAU50001 Propionibacterium acnes immunogenic protein	+1	82	0.042
>patp:AAU50001 Propionibacterium acnes immunogenic protein	+1	82	0.042
>patp:AAU18674 Renal and cardiovascular-associated protein	+1	79	0.085
>patp:AAB95341 Human protein sequence SEQ ID NO:17621	+1	99	0.17

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 2188 of 2189 bases (99%) identical to a gb:GENBANK-ID:AK000243|acc:AK000243.1 mRNA from *Homo sapiens* (cDNA FLJ20236 fis, clone COLF5810, highly similar to AB011095 *Homo sapiens* mRNA for KIAA0523 protein). The full amino acid sequence of the protein of the invention was found to have 395 of 395 amino acid residues (100%) identical to, and 395 of 395 amino acid residues (100%) similar to, the 468 amino acid residue ptnr:SPTREMBL-ACC:O60276 protein from *Homo sapiens* (KIAA0523 PROTEIN)(Fig. 3B).

NOV6 also has homology to the proteins shown in the BLASTP data in Table 6D.

Table 6D. BLAST results for NOV6

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|14602977|gb|AAH09975.1|AAH09975 (BC009975)	Similar to KIAA0789 gene product [*Homo sapiens*]	575	523/575 (90%)	524/575 (90%)	0.0
gi|3043570|dbj|BAA25449.1| (AB011095)	KIAA0523 protein [*Homo sapiens*]	468	417/468 (89%)	417/468 (89%)	0.0
gi|18489296|ref|XP_082751.1| (XM_082751)	CG9164 [*Drosophila melanogaster*]	317	76/206 (36%)	15/206 (54%)	3e-28
gi|16944644|emb|CAD11404.1| (AL513445)	hypothetical protein [*Neurospora crassa*]	2117	43/131 (32%)	62/131 (46%)	5e-08
gi|11359357|pir||T43257	beta-1,3 exoglucanase (EC 3.2.1.-) precursor - fungus (*Trichoderma harzianum*)	1032	40/128 (31%)	55/128 (42%)	2e-05

A multiple sequence alignment is given in Table 6E, with the NOV6 protein being shown on line 1 in Table 6E in a ClustalW analysis, and comparing the NOV6 protein with the related protein sequences shown in Table 6D. This BLASTP data is displayed graphically in the ClustalW in Table 6E.

Table 6E. ClustalW Analysis of NOV6

1) > NOV6; SEQ ID NO:12
2) >gi|14602977|/ Similar to KIAA0789 gene product [*Homo sapiens*]; SEQ ID NO:60
3) >gi|3043570|/ KIAA0523 protein [*Homo sapiens*]; SEQ ID NO:61
4) >gi|18489296|/ CG9164 [*Drosophila melanogaster*]; SEQ ID NO:62
5) >gi|16944644|/ hypothetical protein [*Neurospora crassa*]; SEQ ID NO:63
6) >gi|11359357|/ beta-1,3 exoglucanase (EC 3.2.1.-) precursor - fungus [*Trichoderma harzianum*]; SEQ ID NO:64

850 860 870 880 890 900

```
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                    1   ------------------------MAKPFFRLQKFLRRTQFLLFFLTAAYLMTG----SL 32
gi|14602977|            1   ------------------------MAKPFFRLQKFLRRTQFLLFFLTAAYLMTG----SL 32
gi|3043570|             1   ---------------------------------------------------------- 1
gi|18489296|            1   ---------------------------------------------------------- 1
gi|16944644|          841   GTSPGLTTTEVTIRFTNKGGSDLMIDKSKPPMGSVLGAQNPSSDLFEGMVIKPGKSESAT 900
gi|11359357|          235   TFNGGLIG--------------AAIGNQQYTMRNLVFNN--CAQPLSAASIGSG----FT 274

                                         910       920       930       940       950       960
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                   33   LLLQRVRVALPQGPRAPGPLQTLPVAAVALGVGLLDSR-ALHDPRVSPELLLGVDMLQSP 91
gi|14602977|           33   LLLQRVRVALPQGPRAPGPLQTLPVAAVALGVGLLDSR-ALHDPRVSPELLLGVDMLQSP 91
gi|3043570|            1   ---------------------------------------------------------- 1
gi|18489296|           1   ---------------------------------------------------------- 1
gi|16944644|         901   LFFTPGAAPLNADPIVYSGAWTLNVNDLTFGVHVVNFIGTLRATKVGPTLPDGSARFKYL 960
gi|11359357|         275   RAISINNCGLGIDMTAAESITLIDSSISGTPVGIKTSFRRNQSPATSNSLIVENLSLNNV 334

                                              970       980       990      1000      1010      1020
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                   92   LTRPRPGPRWLRSRNSELR----------------------------------------QL 112
gi|14602977|           92   LTRPRPGPRWLRSRNSELR----------------------------------------QL 112
gi|3043570|            1   ----------------ELR----------------------------------------QL 5
gi|18489296|           1   ---------------------------------------------------------- 1
gi|16944644|         961   GCYRDSSANRLETTQAQFPSDNDNGKCQQYAITNKAAFAGTQYTYECWVGRSIPPASLKV 1020
gi|11359357|         335   PVAIQSSSGSTILAGGTTT------------------------------------------I 354

                                                   1030      1040      1050      1060      1070      1080
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                  113   RRRWFHHFMSX------------------------------LPGTARPG--PR------- 133
gi|14602977|          113   RRRWFHHFMSDSQGP--------------------------PALGPEAARPAIHSRG---TYI 146
gi|3043570|             6   RRRWFHHFMSDSQGP--------------------------PALGPEAARPAIHSRG---TYI 39
gi|18489296|           -1   ---------------------------------------------------------- 1
gi|16944644|         1021   DDYLCNIYICPGDKSQFCGGVGSYMMMWYDTTGYFPENGTLAPAFRPPASKAVVGDWEYA 1080
gi|11359357|          355   AAWGQGHQYTPNGPTTFQG----------------------SITPNSRPSSLLSGSN--YYT 392

                                                        1090      1100      1110      1120      1130      1140
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                  134   GCQA---------R---------------HPQP--RSYVYAGLEAGAECYCGNRLPAVSV 167
gi|14602977|          147   GCFSDDGHERTLKGAVFYDLRKMTV-SHCQDACAERSYVYAGLEAGAECYCGNRLPAVSV 205
gi|3043570|            40   GCFSDDGHERTLKGAVFYDLRKMTV-SHCQDACAERSYVYAGLEAGAECYCGNRLPAVSV 98
gi|18489296|           1   ---------------------------------MALQGWRFFGVSATIIIYIG-----G-- 21
gi|16944644|        1081   GCRTDNSASPATRALNDRIVGQSSTNTIESCAQACAGFSFFGVEWGVECYCGNNLNPGST 1140
gi|11359357|         393   RSKPQYETLPVSSFRSVRSAGATGN-AVTDDTAALQSVLNSATACGQIVYFDAGIYRITS 451

                                                             1150      1160      1170      1180      1190      1200
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                  168   GLEE--CNHECKGEKGSVCGAVDRLSVYRVDELQPG----------------------S 202
gi|14602977|          206   GLEE--CNYECKGEKGSVCGAVDRLSVYRVDELQPG----------------------S 240
gi|3043570|            99   GLEE--CNHECKGEKGSVCGAVDRLSVYRVDELQPG----------------------S 133
gi|18489296|           21   ------------------------VLELSMNNIPG----------------------S 33
gi|16944644|         1141   VADEKTCNYVCGGDPTELCGGSGRISVYKQKGTVIGNPSTGVSSSSGTASGTASATASAS 1200
gi|11359357|          452   TLS------IPPG--AKIVG--EEYPIIMSSGSFFND----------------------Q 479

                                                                  1210      1220      1230      1240      1250      1260
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                  203   RKRRTATYRGCFRLPENI---------THAFPSSLIQAN------VTVGTCSGFCS-QKE 246
gi|14602977|          241   RKRRTATYRGCFRLPENI---------THAFPSSLIQAN------VTVGTCSGFCS-QKE 284
gi|3043570|           134   RKRRTATYRGCFRLPENI---------THAFPSSLIQAN------VTVGTCSGFCS-QKE 177
gi|18489296|           34   HPKRPRIER--F----------------AEFPS----------------------FHS-PR- 53
gi|16944644|         1201   STSSSAIASGTPGNPQSIGQYSSLGCYSDAVASRSLQGKNTQSNVMSLDDCATYCAGYKY 1260
gi|11359357|          480   SNPKPVVQVGTPGQTG-----------QVEWSDMIVSTQG--------TQAGAVLIEWNL 520

                                                                       1270      1280      1290      1300      1310      1320
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                  247   FPLAILRGWECYCAYPTPRFNLRDAMDSSVCGQDPE---------------------- 282
gi|14602977|          285   FPLAILRGWECYCAYPTPRFNLRDAMDSSVCGQDPE---------------------- 320
gi|3043570|           178   FPLAILRGWECYCAYPTPRFNLRDAMDSSVCGQDPE---------------------- 213
gi|18489296|           53   ----FP--------MPSRKMTIRWCRDLKYINRD------------------------- 75
gi|16944644|         1261   FGTEYSAECFCGNDLLNGAAPVTDGRCNMLCNGNQQQICGGSNGLSMYQLNPNGTSSSVT 1320
gi|11359357|          521   ATSGTPSGMWDVHTRIGGFKGSNLQVAQCPVTAS------------------------- 554
```

```
                                    1330       1340       1350       1360       1370       1380
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                 282  ----------AQRLAEYCEVYQTPVQDTRCTDRRFLPNK-------------------- 311
gi|14602977|         320  ----------AQRLAEYCEVYQTPVQDTRCTDRRFLPNK-------------------- 349
gi|3043570|          213  ----------AQRLAEYCEVYQTPVQDTRCTDRRFLPNK-------------------- 242
gi|18489296|         75   ----------LPIYADYKSDFYTALPSDVSAALQSLP---------------------- 102
gi|16944644|         1321 ASGSATQSATASGTASGTASSSSTALATTTSSAVPTMPAVSVKCPDNNNGTYLSLNGKTF 1380
gi|11359357|         554  ----------STTVNTACIGAYMSMHITASASNLYMEN-------------------- 582

                                    1390       1400       1410       1420       1430       1440
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                 311  ----------------------------SKVFVALSSFPGAG---------------- 325
gi|14602977|         349  ----------------------------SKVFVALSSFPGAG---------------- 363
gi|3043570|          242  ----------------------------SKVFVALSSFPGAG---------------- 256
gi|18489296|         102  ----------------------------ALTALASFPGSG---------------- 114
gi|16944644|         1381 LLECFTDHEGGDLALAYVDSYALCAEKCSTTDLCMAFAYVPGGTGIQAPCYMKKSVGRGF 1440
gi|11359357|         582  ----------------------------NWLWTADHDIDDSS----------N--- 597

                                    1450       1460       1470       1480       1490       1500
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                 325  ---NTWARHLIEHATGFYT----------------------GSYYFDGTLYNKGFKGEK 359
gi|14602977|         363  ---NTWARHLIEHATGFYT----------------------GSYYFDGTLYNKGFKGEK 397
gi|3043570|          256  ---NTWARHLIEHATGFYT----------------------GSYYFDGTLYNKGFKGEK 290
gi|18489296|         114  ---NTWLRYLLQQATGILT----------------------GSIYKDYGLLKTGFPAEN 148
gi|16944644|         1441 NNQAVWGAQIIAVASASTTTSAAATMTSAASAMGSNSVVGPSGSVVTGSTATGSATDSTT 1500
gi|11359357|         598  TQITIFSGRGLYVESTAGT--------------------FWFVGTAVEHHTLYQYQFANTQ 638

                                    1510       1520       1530       1540       1550       1560
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                 360  DHWRSRRTICVKTHESGRREIEMFDSAILLIRNPYRSLVAEFNRKCAGHLGYAADRNWKS 419
gi|14602977|         398  DHWRSRRTICVKTHESGRREIEMFDSAILLIRNPYRSLVAEFNRKCAGHLGYAADRNWKS 457
gi|3043570|          291  DHWRSRRTICVKTHESGRREIEMFDSAILLIRNPYRSLVAEFNRKCAGHLGYAADRNWKS 350
gi|18489296|         148  --VCNSSVLLVKTHEWGSKAWAPFSKAILLVRDPEKAIIAEFNRQSGGHIGFASPDRYKR 206
gi|16944644|         1501 TGSTATGSTAAGSSVASATGSDAISAALPAVSGTAATSIVASGSGSGAASGAAPTAASTV 1560
gi|11359357|         639  NIYAGVIQTETPYYQPNPDAPTPFNVNTALNDPNFATSCSGSSGRCAEAWGLRIVSSQN- 697

                                    1570       1580       1590       1600       1610       1620
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                 420  KE---WPDFVNSYASWWSSHVLDWLK-YGKRLLVVHYEELRRS---------------- 458
gi|14602977|         458  KE---WPDFVNSYASWWSSHVLDWLK-YGKRLLVVHYEELRRS---------------- 496
gi|3043570|          351  KE---WPDFVNSYASWWSSHVLDWLK-YGKRLLVVHYEELRRS---------------- 389
gi|18489296|         207  TKGKYWQQFVSNKLKGWEMMNLSWARNFTGSIKVVFYDDLVHH---------------- 249
gi|16944644|         1561 ASGSASASVAPSASVAPSDIVVPSAS-VITSASVVSSDSIAPSDSIAPSENTTPSASVAP 1619
gi|11359357|         697  ---------ILIYAAGLYSFFENNDGN--TGCDVALGPENCQN---------------- 729

                                    1630       1640       1650       1660       1670       1680
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                 458  --------LVPTLREMVAFLNVSVSEERLLCVENNKEGSFRRRG---------------- 494
gi|14602977|         496  --------LVPTLREMVAFLNVSVSEERLLCVENNKEGSFRRRG---------------- 532
gi|3043570|          389  --------LVPTLREMVAFLNVSVSEERLLCVENNKEGSFRRRG---------------- 425
gi|18489296|         249  --------TERELRSILDFLQFPINEQLMRCAIMRKEGIFRRK---------------- 284
gi|16944644|         1620 SNSVVPSDSVAPSASIIPSASVASSASVVPSTSIAHSGSAAASESVAPAESIAPSASVSS 1679
gi|11359357|         729  --------NIFDLEGTLTNINVYNLGTVGVVNQITQNGNVLATS------S---------- 766

                                    1690       1700       1710       1720       1730       1740
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                 494  ------------------------------------------------------------ 494
gi|14602977|         532  ------------------------------------------------------------ 532
gi|3043570|          425  ------------------------------------------------------------ 425
gi|18489296|         284  ------------------------------------------------------------ 284
gi|16944644|         1680 GSNTGVAPTNSASVTPTNSASVATTISVSVAPTASDAPTTSITLSVAPGSSSSTTAPAVV 1739
gi|11359357|         766  ------------------------------------------------------------ 766

                                    1750       1760       1770       1780       1790       1800
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                 494  ------------------------RRSHDPEPFTPEMKDLINGYIRTVDQALRDHNWTG 529
gi|14602977|         532  ------------------------RRSHDPEPFTPEMKDLINGYIRTVDQALRDHNWTG 567
gi|3043570|          425  ------------------------RRSHDPEPFTPEMKDLINGYIRTVDQALRDHNWTG 460
gi|18489296|         284  ------------------------KRLLSFDPYTESMRAEVQNRRRIVYGLLGRQEP-- 317
```

```
gi|16944644|   1740 STTLSAPTVTSVPPAAGTSTTTAAAVTSTDPAPVTTTTTTATSSTTNTTPVLFTTSTTTT 1799
gi|11359357|    766 -------------------------SNVNAFADVIALFRLASGSGGVTPPPSSTTKAQST 801

                          1810      1820      1830      1840      1850      1860
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV6                530 LPREYVPR---------------------------------------------- 537
gi|14602977|        568 LPREYVPR---------------------------------------------- 575
gi|3043570|         461 LPREYVPR---------------------------------------------- 468
gi|18489296|        317 ---------------------------------------------------- 317
gi|16944644|       1800 TAAAVATTAPTQATTTIATAATTSTTNIASASPTIPAVVNWDYQGCASDSNTAAPTARAL 1859
gi|11359357|        802 TFSTIITSSPPKQTG-----------------------WNFLGCYSDNVNGRTLANQV 836
```

The NOV6 Clustal W alignment shown in Table 6E was modified to begin at amino residue 841 and end at amino acid residue 1860. The data in Table 6E includes all of the regions overlapping with the NOV6 protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 6F lists the domain description from DOMAIN analysis results against NOV6.

Table 6F Domain Analysis of NOV6

Model	Region of Homology	Score (bits)	E value
Disintegrin	151-159	5.8	0.73
WSC domain	120-186	36.8	5e-07
Peptidase family M1	346-354	-0.2	8.4
Sulfotransferase proteins	288-518	-143.3	0.14

Consistent with other known members of the Secretory Protein-like family of proteins, NOV6 has, for example, has homology to other members of the Secretory Protein-like Protein Family. NOV6 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV6 nucleic acids and polypeptides can be used to identify proteins that are members of the Secretory Protein-like Protein Family. The NOV6 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV6 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, *e.g.,* cellular activation and signal transduction. These molecules can be used to treat, *e.g.*, Cardiovascular diseases, Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus as well as other diseases, disorders and conditions.

In addition, various NOV6 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV6 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the Secretory Protein-like Protein Family.

The NOV6 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of cardiac physiology. As such, the NOV6 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat cardiac and vascular system disorders, *e.g.,* Cardiovascular diseases, Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus as well as other diseases, disorders and conditions

The NOV6 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV6 nucleic acid is expressed in Aorta.

Additional utilities for NOV6 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV7

A NOV7 polypeptide has been identified as a Transmission Blocking Target Antigen S230 Precursor-like protein (also referred to as CG94978-01). The disclosed novel NOV7 nucleic acid (SEQ ID NO:13) of 1629 nucleotides is shown in Table 7A. The novel NOV7 nucleic acid sequences maps to the chromosome 1.

An ORF begins with an ATG initiation codon at nucleotides 1-3 and ends with a TGA codon at nucleotides 1627-1629. A putative untranslated region and/or downstream from the termination codon is underlined in Table 7A, and the start and stop codons are in bold letters.

Table 7A. NOV7 Nucleotide Sequence (SEQ ID NO:13)

ATGGCGGTGCCCGGCGAGGCGGAGGAGGAGGCGACAGTTTACCTGGTAGTGAGCGGTATCCCCTCCG
TGTTGCGCTCGGCCCATTTACGGAGCTATTTTAGCCAGTTCCGAGAAGAGCGCGGCGGTGGCTTCCT
CTGTTTCCACTACCGGCATCGGCCTGAGCGGGCCCCTCCGCAGGCCGCTCCTAACTCTGCCCTAATT
CCTACCGACCCAGCCGCTGAGGGCCAGCTTCTCTCTCAGACTTCGGCCACCGATGTCCGGCCTCTCT
CCACTCGAGACTCTACTCCAATCCAGACCCGCACCTGCTGCTGCGTCATCTCGGTAAGGGGGTTGGC
TCAAGCTCAGAGGCTTATTCGCATGTACTCGGGCCGCCGGTGGCTGGATTCTCACGGGACTTGGCTA
CCGGGTCGCTGTCTCATCCGCAGACTTCGGCTACCTACGGAGGCATCAGGTCTGGGCTCCTTTCCCT
TCAAGACCCGGAAGGAACTGCAGAGTTGGAAGGCAGAGAATGAAGCCTTCACCCTGGCTGACCTGAA
GCAACTGCCGGAGCTGAACCCACCAGTGCTGATGCCCAGAGGGAATGTGGGGACTCCCCTGCGGGTC
TTTTTGGAGTTGATCCGGGCCTGCCGCCTACCCCCTCGGATCATCACCCAGCTGCAGCTCCAGTTCC

CCAAGACAGGTTCCTCCCGGCGCTACGGCAATGTGCCTTTTGAGTATGAGGACTCAGAGACTGTGGA
GCAGGAAGAGCTTGTGTATACAGCAGAGGGTGAAGAAATACCCCAAGGAACCTACCTGGCAGATATA
CCAGCCAGCCCCTGTGGAGAGCCTGAGGAAGAAGTGGGGAAGGAAGAGGAAGAAGAGTCTCACTCAG
ATGAGCTGTTCGGGTGTGCTGTGGTCATCCTCCCTGCGCACCTACAGCCGCAGACCGCCGGTGGGGG
GCGGGGGATGCCGGGCTGCCGCATCAGCGCCTGCGGCCCGGGGGCCCAGGAGGGGACGGCAGAGCAG
AGGTCGCCGCCGCCGCCCTGGGATCCCATGCCGTCCTCTCAGCCCCCGCCCCCAACTCCGACCTTGA
CTCCTACCCCGACCCCGGGTCAGTCCCCGCCGCTGCCGGACGCAGCTGGGGCTTCAGCAGGCGCGGC
CGAGGACCAGGAGCTGCAGCGCTGGCGCCAGGGCGCTAGCGGGATCGCGGGGCTCGCCGGCCCCGGA
GGGGGCTCTGGCGCGGCTGCGGGGGCGGGGGGCCGCGCGCTGGAGCTGGCCGAAGCACGGCGGCGGC
TGCTGGAGGTGGAGGGCCGCCGGCGCCTGGTGTCGGAGCTGGAGAGCCGCGTGCTGCAGCTGCACCG
CGTTTTCTTGGCGGCCGAGCTGCGCCTGGCGCACCGCGCGGAGAGCCTGAGCCGCCTGAGCGGCGGC
GTGGCGCAGGCCGAGCTCTACCTGGCGGCTCACGGGTCGCGCCTCAAGAAGGGCCCGCGCCGCGGCC
GCCGCGGCCGACCCCCCGCGCTGCTGGCCTCGGCGCTGGGCCTGGGCGGCTGCGTGCCCTGGGGTGC
CGGGCGACTGCGGCGCGGCCACGGCCCCGAGCCCGACTCGCCCTTCCGCCGCAGCCCGCCCCGCGGC
CCCGCCTCCCCGCAGCGC**TGA**

The NOV7 protein (SEQ ID NO:14) encoded by SEQ ID NO:13 is 542 amino acid residues in length and is presented using the one-letter amino acid code in Table 7B. Psort analysis predicts the NOV7 protein of the invention to be localized in the cytoplasm with a certainty of *0.4500*.

Table 7B. Encoded NOV7 protein sequence (SEQ ID NO:14)

MAVPGEAEEEATVYLVVSGIPSVLRSAHLRSYFSQFREERGGGFLCFHYRHRPERAPPQAAPNSA
LIPTDPAAEGQLLSQTSATDVRPLSTRDSTPIQTRTCCCVISVRGLAQAQRLIRMYSGRRWLDSH
GTWLPGRCLIRRLRLPTEASGLGSFPFKTRKELQSWKAENEAFTLADLKQLPELNPPVLMPRGNV
GTPLRVFLELIRACRLPPRIITQLQLQFPKTGSSRRYGNVPFEYEDSETVEQEELVYTAEGEEIP
QGTYLADIPASPCGEPEEEVGKEEEEESHSDELFGCAVVILPAHLQPQTAGGGRGMPGCRISACG
PGAQEGTAEQRSPPPPWDPMPSSQPPPPTPTPLTPTPTPGQSPPLPDAAGASAGAAEDQELQRWRQ
GASGIAGLAGPGGGSGAAAGAGGRALELAEARRRLLEVEGRRRLVSELESRVLQLHRVFLAAELR
LAHRAESLSRLSGGVAQAELYLAAHGSRLKKGPRRGRRGRPPALLASALGLGGCVPWGAGRLRRG
HGPEPDSPFRRSPPRGPASPQR

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

Table 7C. Patp results for NOV7

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAU33166 Novel human secreted protein #3657	+1	1533	5.8e-157
>patp:AAE04880 Human protease protein-7 (PRTS-7)	+1	1533	5.8e-157
>patp:AAB94023 Human protein sequence SEQ ID NO:14157	+1	1519	1.8e-155
>patp:AAU33124 Novel human secreted protein #3615	+1	390	7.7e-36
>patp:AAG02700 Human secreted protein, SEQ ID NO: 6781	+1	268	3.5e-22

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 874 of 876 bases (99%) identical to a

gb:GENBANK-ID:AK022517|acc:AK022517.1 mRNA from *Homo sapiens* (cDNA FLJ12455 fis, clone NT2RM1000563, weakly similar to TRANSMISSION-BLOCKING TARGET ANTIGEN S230 PRECURSOR). The full amino acid sequence of the protein of the invention was found to have 290 of 292 amino acid residues (99%) identical to, and 290 of 292 amino acid residues (99%) similar to, the 525 amino acid residue ptnr:SPTREMBL-ACC:Q9H9Z3 protein from *Homo sapiens* (CDNA FLJ12455 FIS, CLONE NT2RM1000563, WEAKLY SIMILAR TO TRANSMISSION- BLOCKING TARGET ANTIGEN S230 PRECURSOR).

NOV7 also has homology to the proteins shown in the BLASTP data in Table 7D.

Table 7D. BLAST results for NOV7

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|18545154|ref|XP_084046.1|(XM_084046)	hypothetical protein FLJ12455 [*Homo sapiens*]	525	274/274 (100%)	274/274 (100%)	e-147
gi|11545793|ref|NP_071361.1|(NM_022078)	hypothetical protein FLJ12455 [*Homo sapiens*]	525	272/274 (99%)	272/274 (99%)	e-145
gi|18545156|ref|XP_086159.1|(XM_086159)	similar to hypothetical protein FLJ12455 [*Homo sapiens*]	107	83/84 (98%)	84/84 (99%)	6e-37
gi|18545158|ref|XP_097448.1|(XM_097448)	hypothetical protein XP_097448 [*Homo sapiens*]	141	135/137 (98%)	136/137 (98%)	2e-33
gi|17562286|ref|NP_505420.1|(NM_073019)	K07B1.7b.p [*Caenorhabditis elegans*]	487	76/237 (32%)	119/237 (50%)	3e-30

A multiple sequence alignment is given in Table 7E, with the NOV7 protein being shown on line 1 in Table 7E in a ClustalW analysis, and comparing the NOV7 protein with the select related protein sequences shown in Table 7D. This BLASTP data is displayed graphically in the ClustalW in Table 7E.

Table 7E. ClustalW Analysis of NOV7

1) > NOV7; SEQ ID NO:14
2) > gi|18545154|/ hypothetical protein FLJ12455 [*Homo sapiens*]; SEQ ID NO:65
3) > gi|11545793|/ hypothetical protein FLJ12455 [*Homo sapiens*]; SEQ ID NO:66

```
                                10        20        30        40        50        60
                                |....|....|....|....|....|....|....|....|....|....|....|....|
NOV7                       1    MAVPGEAEEEATVYLVVSGIPSVLRSAHLRSYFSQFREERGGGFLCFHYRHRPERAPPQA  60
gi|18545154|               1    MAVPGEAEEEATVYLVVSGIPSVLRSAHLRSYFSQFREERGGGFLCFHYRHRPERAPPQA  60
gi|11545793|               1    MAVPGEAEEEATVYLVVSGIPSVLRSAHLRSYFSQFREERGGGFLCFHYRHRPERAPPQA  60

                                70        80        90        100       110       120
                                |....|....|....|....|....|....|....|....|....|....|....|....|
NOV7                       61   APNSALIPTDPAAEGQLLSQTSATDVRPLSTRDSTPIQTRTCCCVISVRGLAQAQRLIRM 120
```

```
gi|18545154|   61  APNSALIPTDPAAEGQLLSQTSATDVRPLSTRDSTPIQTRTCCCVISVRGLAQAQRLIRM 120
gi|11545793|   61  APNSALIPTDPAAEGQLLSQTSATDVRPLSTRDSTPIQTRTCCCVISVRGLAQAQRLIRM 120

                        130       140       150       160       170       180
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV7           121 YSGRRWLDSHGTWLPGRCLIRRLRLPTEASGLGSFPFKTRKELQSWKAENEAFTLADLKQ 180
gi|18545154|   121 YSGRRWLDSHGTWLPGRCLIRRLRLPTEASGLGSFPFKTRKELQSWKAENEAFTLADLKQ 180
gi|11545793|   121 YSGRRWLDSHGTWLPGRCLIRRLRLPTEASGLGPFPFKTRKELQSWKAENEAFTLADLKQ 180

                        190       200       210       220       230       240
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV7           181 LPELNPPVLMPRGNVGTPLRVFLELIRACRLPPRIITQLQLQFPKTGSSRRYGNVPFEYE 240
gi|18545154|   181 LPELNPPVLMPRGNVGTPLRVFLELIRACRLPPRIITQLQLQFPKTGSSRRYGNVPFEYE 240
gi|11545793|   181 LPELNPPVLMPRGNVGTPLRVFLELIRACRLPPRIITQLQLQFPKTGSSRRYGNVPFEYE 240

                        250       260       270       280       290       300
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV7           241 DSETVEQEELVYTAEGEEIPQGTYLADIPASPCGEPEEEVGKEEEEESHSDELFG-CAVV 299
gi|18545154|   241 DSETVEQEELVYTAEGEEIPQGTYLADIPASPCGEPEEEVGKEEEEESHSDEDDDRGEEW 300
gi|11545793|   241 DSETVEQEELVCTAEGEEIPQGTYLADIPASPCGEPEEEVGKEEEEESHSDEDDDRGEEW 300

                        310       320       330       340       350       360
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV7           300 ILPAHLQPQTAGGGRGMPGCRISACGPGAQEGTAEQRSPPPPWDPMPSSQPPPPTPTLTP 359
gi|18545154|   301 ERHEALHEDVTGQERTTEQLFEEEIELKWEKGGSGLVFYT---DAQFWQEEEGDFDEQTA 357
gi|11545793|   301 ERHEALHEDVTGQERTTEQLFEEEIELKWEKGGSGLVFYT---DAQFWQEEEGDFDEQTA 357

                        370       380       390       400       410       420
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV7           360 TPTPGQSPPLPDAAGASAGAAEDQELQRWRQGASGIAGLAGPGGGSGAAAGAGGRALELA 419
gi|18545154|   358 DDWDVDMSVYYDRDGGDK-DARDSVQMRLEQRLR-----DGQEDGSVIERQVGTFERHTK 411
gi|11545793|   358 DDWDVDMSVYYDRDGGDK-DARDSVQMRLEQRLR-----DGQEDGSVIERQVGTFERHTK 411

                        430       440       450       460       470       480
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV7           420 EARRRLLEVEGRRRLVSELESRVLQLHRVFLAAELRLAHRAESLSRLSGGVAQAELYLAA 479
gi|18545154|   412 GIGRKVMERQG----WAEGQGLGCRCSGVPEALDSDGQH-----PRCKRGLGYHGEKLQP 462
gi|11545793|   412 GIGRKVMERQG----WAEGQGLGCRCSGVPEALDSDGQH-----PRCKRGLGYHGEKLQP 462

                        490       500       510       520       530       540
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV7           480 HGSRLKKGPRRGRRGRPPAILASALGLGGCVPWGAGRLRRGHGPEPDSPFRRSPPRGPAS 539
gi|18545154|   463 FG-QLKR-PRRNGLGLISTIYDEPLPQDQTESLLRRQPPTSMKFRTDMAFVRGSSCASDS 520
gi|11545793|   463 FG-QLKR-PRRNGLGLISTIYDEPLPQDQTESLLRRQPPTSMKFRTDMAFVRGSSCASDS 520

                   ....|
NOV7           540 PQR-- 542
gi|18545154|   521 PSLPD 525
gi|11545793|   521 PSLPD 525
```

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 7F lists the domain description from DOMAIN analysis results against NOV7.

Table 7F Domain Analysis of NOV7

Model	Region of Homology	Score (bits)	E value
#PD343750 BLOCKING NT2RM1000563	14-54	222	4e-19

TRANSMISSION-FIS			
#PD229850 BLOCKING NT2RM1000563 TRANSMISSION-FIS	55-90	179	4e-14
#PD138963 BLOCKING NT2RM1000563 TRANSMISSION-FIS	91-258	872	2e-94

Consistent with other known members of the Transmission Blocking Target Antigen S230 Precursor-like family of proteins, NOV7 has, for example, three Blocking NT2RM1000563 Transmission-FIS Antigen Weakly Precursor Peptidase A2 signature sequences and homology to other members of the Transmission Blocking Target Antigen S230 Precursor-like Protein Family. NOV7 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV7 nucleic acids and polypeptides can be used to identify proteins that are members of the Transmission Blocking Target Antigen S230 Precursor-like Protein Family. The NOV7 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV7 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, *e.g.*, cellular activation and signal transduction. These molecules can be used to treat, *e.g.*, Cardiovascular diseases, 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, Diabetes,Von Hippel-Lindau (VHL) syndrome , Pancreatitis, Obesity, Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection

In addition, various NOV7 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV7 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the Transmission Blocking Target Antigen S230 Precursor-like Protein Family.

The NOV7 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of cardiac and nerve physiology. As such, the NOV7 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat cardiovascular and nervous system disorders, *e.g.*, Cardiovascular diseases, 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, Diabetes,Von Hippel-Lindau (VHL) syndrome, Pancreatitis, Obesity, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection

The NOV7 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV7 nucleic acid is expressed in Adipose, Heart, Aorta, Coronary Artery, Umbilical Vein, Pancreas, Liver, Gall Bladder, Colon, Bone Marrow, Thymus, Bone, Cartilage, Synovium/Synovial membrane, Skeletal Muscle, Brain, Left cerebellum, Right Cerebellum, Thalamus, Hypothalamus, Pituitary Gland, Frontal Lobe, Parietal Lobe, Cerebral Medulla/Cerebral white matter, Basal Ganglia/Cerebral nuclei, Substantia Nigra, Hippocampus, Cervix, Mammary gland/Breast, Uterus, Oviduct/Uterine Tube/Fallopian tube, Prostate, Testis, Lung, Bronchus, Larynx, Kidney, Retina, Skin, Epidermis.

Additional utilities for NOV7 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV8

A NOV8 polypeptide has been identified as a Nuclear Protein-like protein (also referred to as CG94713-01). The disclosed novel NOV8 nucleic acid (SEQ ID NO:15) of 3807 nucleotides is shown in Table 8A. The novel NOV8 nucleic acid sequences maps to the chromosome 1.

An ORF begins with an ATG initiation codon at nucleotides 16-18 and ends with a TGA codon at nucleotides 3793-3795. A putative untranslated region and/or downstream

from the termination codon is underlined in Table 8A, and the start and stop codons are in bold letters.

Table 8A. NOV8 Nucleotide Sequence (SEQ ID NO:15)

ATGATAAAATAGAAG**ATG**AATTGCAAACCTTCTTTACCAGTGATAAAGATGGAAATTACACATGCAT
ACAACCCGAAATCACCACCTACACAAAACTCTTCAGCCAGCAGTGTGAACTGGAATTCTGCCAACCC
AGATGACATGGTGGTTGATTATGAAACTGACCCTGCTGTAGTTACTGGTGAAAATATTTCTTTAAGC
CTTCAGGGTGTTGAAGTATTTGGTCATGAAAAGTCTTCTAGTGATTTCATTAGTAAGCAGGTGTTAG
ATATGCATAAAGATTCTATTTGTCAGTGTCCTGCACTTGTAGGTACTGAGAAGCCCAAATATCTGCA
ACACAGTTGTCATTCCCTAGAAGCAGTTGAGGGCCAGAGTGTTGAGCCATCTTTGCCTTTTGTGTGG
AAGCCTAATGACAATTTGAACTGTGCAGGCTACTGTGATGCCTTGGAGCTGAACCAAACATTTGACA
TGACAGTGGATAAAGTTAACTGCACCTTTATATCACATCATGCCATCGGAAAGAGTCAGTCCTTCCA
TACTGCTGGAAGCCTGCCACCAACTGGTAGGAGAAGTGGAAGTACATCTTCTTTATCCTATTCCACT
TGGACATCTTCCCATTCTGATAAGACGCATGCAAGAGAAACTACTTATGATAGAGAAAGCTTTGAAA
ACCCTCAAGTCACACCATCAGAAGCCCAAGACATGACTTACACAGCATTTTCTGATGTGGTGATGCA
AAGTGAGGTTTTTGTTTCAGATATTGGAAATCAGTGTGCATGTTCTTCAGGAAAGGTCACCAGTGAG
TACACAGATGGATCACAACAAAGACTAGTTGGAGAAAAAGAGACACAAGCACTAACACCAGTTTCTG
ATGGCATGGAAGTCCCCAATGATTCTGCATTACAAGAGTTCTTTTGTTTATCCCATGATGAATCCAA
TAGCGAACCACATTCACAGAGCTCATACAGGCACAAGGAAATGGGCCAAAATCTGAGAGAGACAGTG
TCCTATTGTCTTATTGATGATGAATGCCCTTTAATGGTGCCAGCTTTTGATAAGAGCGAAGCTCAAG
TGCTGAACCCAGAGCATAAAGTCACTGAGACTGAAGACACACAAATGGTCTCCAAAGGAAAGGATTT
GGGAACCCAAAATCATACCTCAGAATTGATTCTAAGTAGCCCGCCAGGACAAAAGGTGGGCTCGTCA
TTTGGACTGACTTGGGATGCAAATGATATGGTCATTAGCACAGACAAAACGATGTGCATGTCAACAC
CAGTCCTAGAACCCACAAAAGTAACCTTTTCTGTTTCACCGATTGAAGCGACGGAGAAATGTAAGAA
AGTGGAGAAGGGTAATCGAGGGCTTAAAAACATACCAGACTCGAAGGAGGCACCTGTGAACCTGTGT
AAACCCAGTTTAGGAAAATCAACAATCAAAACGAATACCCCAATAGGCTGCAAAGTTAGAAAAACTG
AAATTATAAGTTACCCAAGACCAAACTTCAAGAATGTCAAAGCAAAAGTTATGTCTAGAGCAGTGTT
GCAGCCCAAAGATGCTGCTTTATCAAAGGTCACGCCCAGACCTCAGCAGACCAGTGCCTCATCACCC
TCATCAGTGAATTCAAGACAACAAACAGTCTTGAGCAGAACACCGAGATCTGACTTGAATGCAGACA
AAAAAGCAGAAATTCTAATTAACAAGACACATAAGCAGCAGTTTAATAAACTCATTACTAGCCAGGC
TGTGCATGTTACAACTCATTCTAAAAATGCTTCACACAGGGTTCCAAGAACAACATCTGCCGTGAAA
TCGAATCAGGAAGATGTTGACAAAGCCAGTTCTTCTAACTCAGCATGCGAGACCGGGTCCGTTTCTG
CGTTGTTTCAGAAGATCAAAGGCATACTCCCTGTTAAAATGGAAAGTGCAGAATGTTTGGAAATGAC
CTATGTTCCCAACATTGATAGGATTAGCCCTGAAAAGAAGGGTGAAAAAGAAAATGGGACATCTATG
GAAAAACAAGAGCTGAAACAAGAGATTATGAATGAGACTTTTGAATATGGTTCTCTGTTTTTGGGCT
CTGCTTCAAAAACAACGACCACCTCAGGTAGGAATATATCCAAGCCTGACTCCTGCGGTTTGAGGCA
AATAGCTGCTCCAAAAGCCAAAGTGGGGCCCCCTGTTTCCTGTTTGAGGCGGAACAGTGACAATAGA
AATCCCAGTGCTGATCGAGCCGTATCTCCTCAGAGGATCAGGCGTGTGTCCAGTTCTGGAAAGCCTA
CATCCTTGAAAACTGCACAGTCGTCATGGGTGAATTTGCCTAGACCACTTCCTAAATCCAAAGCATC
TTTGAAAAGTCCTGCGCTGCGGAGGACAGGAAGCACCCCCTCAATAGCCAGCACCCACAGTGAGCTG
AGCACTTACAGCAACAATTCTGGTAATGCCGCTGTCATCAAATATGAGGAGAAACCTCCAAAACCAG
CATTTCAGAATGGTTCCTCAGGATCCTTTTATTTGAAGCCTTTGGTATCCAGGGCTCATGTTCACTT
GATGAAAACTCCTCCAAAAGGTCCTTCGAGAAAAAATTTATTTACAGCTCTTAATGCAGTTGAAAAG
AGCAGGCAAAAGAATCCTCGAAGCTTATGTATCCAGCCACAGACAGCTCCCGATGCGCTGCCCCCTG
AGAAAACACTTGAATTGACGCAATATAAAACAAAATGTGAAAACCAAAGTGGATTTATCCTGCAGCT
CAAGCAGCTTCTTGCCTGTGGTAATACCAAGTTTGAGGCATTGACAGTTGTGATTCAGCACCTGCTG
TCTGAGCGGGAGGAAGCACTGAAACAACACAAAACCCTATCTCAAGAACTTGTTAACCTCCGGGGAG
AGCTAGTCACTGCTTCAACCACCTGTGAGAAATTAGAAAAAGCCAGGAATGAGTTACAAACAGTGTA
TGAAGCATTCGTCCAGCAGCACCAGGCTGAAAAAACAGAACGAGAGAATCGGCTTAAAGAGTTTTAC
ACCAGGGAGTATGAAAAGCTTCGGGACACTTACATTGAAGAAGCAGAGAAGTACAAAATGCAATTGC
AAGAGCAGTTTGACAACTTAAATGCTGCGCATGAAACCTCTAAGTTGGAAATTGAAGCTAGCCACTC
AGAGAAACTTGAATTGCTAAAGAAGGCCTATGAAGCCTCCCTTTCAGAAATTAAGAAAGGCCATGAA
ATAGAAAAGAAATCGCTTGAAGATTTACTTTCTGAGAAGCAGGAATCGCTAGAGAAGCAAATCAATG
ATCTGAAGAGTGAAAATGATGCTTTAAATGAAAAATTGAAATCAGAAGAACAAAAAAGAAGAGCAAG
AGAAAAAGCAAATTTGAAAAATCCTCAGATCATGTATCTAGAACAGGAGTTAGAAAGCCTGAAAGCT
GTGTTAGAGATCAAGAATGAGAAACTGCATCAACAGGACATCAAGTTAATGAAAATGGAGAAACTGG
TGGACAACAACACAGCATTGGTTGACAAATTGAAGCGTTTCCAGCAGGAGAATGAAGAATTGAAAGC
TCGGATGGACAAGCACATGGCAATCTCAAGGCAGCTTTCCACGGAGCAGGCTGTTCTGCAAGAGTCG
CTGGAGAAGGAGTCGAAAGTCAACAAGCGACTCTCTATGGAAAACGAGGAGCTTCTGTGGAAACTGC

ACAATGGGGACCTGTGTAGCCCCAAGAGATCCCCCACATCCTCCGCCATCCCTTTGCAGTCACCAAG
GAATTCGGGCTCCTTCCCTAGCCCCAGCATTTCACCCAGA**TGA**CACCTCCCCAAA

Variant sequences of NOV8 are included in Example 3, Table 20. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA.

The NOV8 protein (SEQ ID NO:16) encoded by SEQ ID NO:15 is 1259 amino acid residues in length and is presented using the one-letter amino acid code in Table 8B. Psort analysis predicts the NOV8 protein of the invention to be localized in the nucleus with a certainty of 0.7600.

Table 8B. Encoded NOV8 protein sequence (SEQ ID NO:16)

MNCKPSLPVIKMEITHAYNPKSPPTQNSSASSVNWNSANPDDMVVDYETDPAVVTGENISLSLQG
VEVFGHEKSSSDFISKQVLDMHKDSICQCPALVGTEKPKYLQHSCHSLEAVEGQSVEPSLPFVWK
PNDNLNCAGYCDALELNQTFDMTVDKVNCTFISHHAIGKSQSFHTAGSLPPTGRRSGSTSSLSYS
TWTSSHSDKTHARETTYDRESFENPQVTPSEAQDMTYTAFSDVVMQSEVFVSDIGNQCACSSGKV
TSEYTDGSQQRLVGEKETQALTPVSDGMEVPNDSALQEFFCLSHDESNSEPHSQSSYRHKEMGQN
LRETVSYCLIDDECPLMVPAFDKSEAQVLNPEHKVTETEDTQMVSKGKDLGTQNHTSELILSSPP
GQKVGSSFGLTWDANDMVISTDKTMCMSTPVLEPTKVTFSVSPIEATEKCKKVEKGNRGLKNIPD
SKEAPVNLCKPSLGKSTIKTNTPIGCKVRKTEIISYPRPNFKNVKAKVMSRAVLQPKDAALSKVT
PRPQQTSASSPSSVNSRQQTVLSRTPRSDLNADKKAEILINKTHKQQFNKLITSQAVHVTTHSKN
ASHRVPRTTSAVKSNQEDVDKASSSNSACETGSVSALFQKIKGILPVKMESAECLEMTYVPNIDR
ISPEKKGEKENGTSMEKQELKQEIMNETFEYGSLFLGSASKTTTTSGRNISKPDSCGLRQIAAPK
AKVGPPVSCLRRNSDNRNPSADRAVSPQRIRRVSSSGKPTSLKTAQSSWVNLPRPLPKSKASLKS
PALRRTGSTPSIASTHSELSTYSNNSGNAAVIKYEEKPPKPAFQNGSSGSFYLKPLVSRAHVHLM
KTPPKGPSRKNLFTALNAVEKSRQKNPRSLCIQPQTAPDALPPEKTLELTQYKTKCENQSGFILQ
LKQLLACGNTKFEALTVVIQHLLSEREEALKQHKTLSQELVNLRGELVTASTTCEKLEKARNELQ
TVYEAFVQQHQAEKTERENRLKEFYTREYEKLRDTYIEEAEKYKMQLQEQFDNLNAAHETSKLEI
EASHSEKLELLKKAYEASLSEIKKGHEIEKKSLEDLLSEKQESLEKQINDLKSENDALNEKLKSE
EQKRRAREKANLKNPQIMYLEQELESLKAVLEIKNEKLHQQDIKLMKMEKLVDNNTALVDKLKRF
QQENEELKARMDKHMAISRQLSTEQAVLQESLEKESKVNKRLSMENEELLWKLHNGDLCSPKRSP
TSSAIPLQSPRNSGSFPSPSISPR

A search against the Patp database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 8C.

Table 8C. Patp results for NOV8

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAG63542 Amino acid sequence of a human ATIP isoform	+1	6389	0.0
>patp:AAG63529 Amino acid sequence of a human ATIP isoform	+1	6233	0.0
>patp:AAG63541 Amino acid sequence of a human ATIP isoform	+1	3928	0.0
>patp:AAG63537 Amino acid sequence of a ATIP isoform	+1	3279	0.0
>patp:AAG63530 Amino acid sequence of a human ATIP isoform	+1	2954	1.5e-307

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 3751 of 3751 bases (100%) identical to a gb:GENBANK-ID:AB033114|acc:AB033114.1 mRNA from *Homo sapiens* (mRNA for KIAA1288 protein, partial cds). The full amino acid sequence of the protein of the invention was found to have 1245 of 1245 amino acid residues (100%) identical to, and 1245 of 1245 amino acid residues (100%) similar to, the 1245 amino acid residue ptnr:SPTREMBL-ACC:Q9ULD2 protein from *Homo sapiens* (KIAA1288 PROTEIN).

NOV8 also has homology to the proteins shown in the BLASTP data in Table 8D.

Table 8D. BLAST results for NOV8

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|6331407|dbj|BAA86602.1|(AB033114)	KIAA1288 protein [*Homo sapiens*]	1245	1245/1245 (100%)	1245/1245 (100%)	0.0
gi|17865632|ref|NP_065800.1|(NM_020749)	AT2 receptor-interacting protein 1 [*Homo sapiens*]	436	404/436 (92%)	409/436 (93%)	0.0
gi|10436722|dbj|BAB14894.1| (AK024357)	unnamed protein product [*Homo sapiens*]	240	239/240 (99%)	239/240 (99%)	e-107
gi|3882269|dbj|BAA34494.1|(AB018317)	KIAA0774 protein [*Homo sapiens*]	1163	135/366 (36%)	224/366 (60%)	9e-49
gi|17475630|ref|XP_029364.3| (XM_029364)	KIAA0774 protein [*Homo sapiens*]	901	135/366 (36%)	224/366 (60%)	3e-48

A multiple sequence alignment is given in Table 8E, with the NOV8 protein being shown on line 1 in Table 8E in a ClustalW analysis, and comparing the NOV8 protein with the related protein sequences shown in Table 8D. This BLASTP data is displayed graphically in the ClustalW in Table 8E.

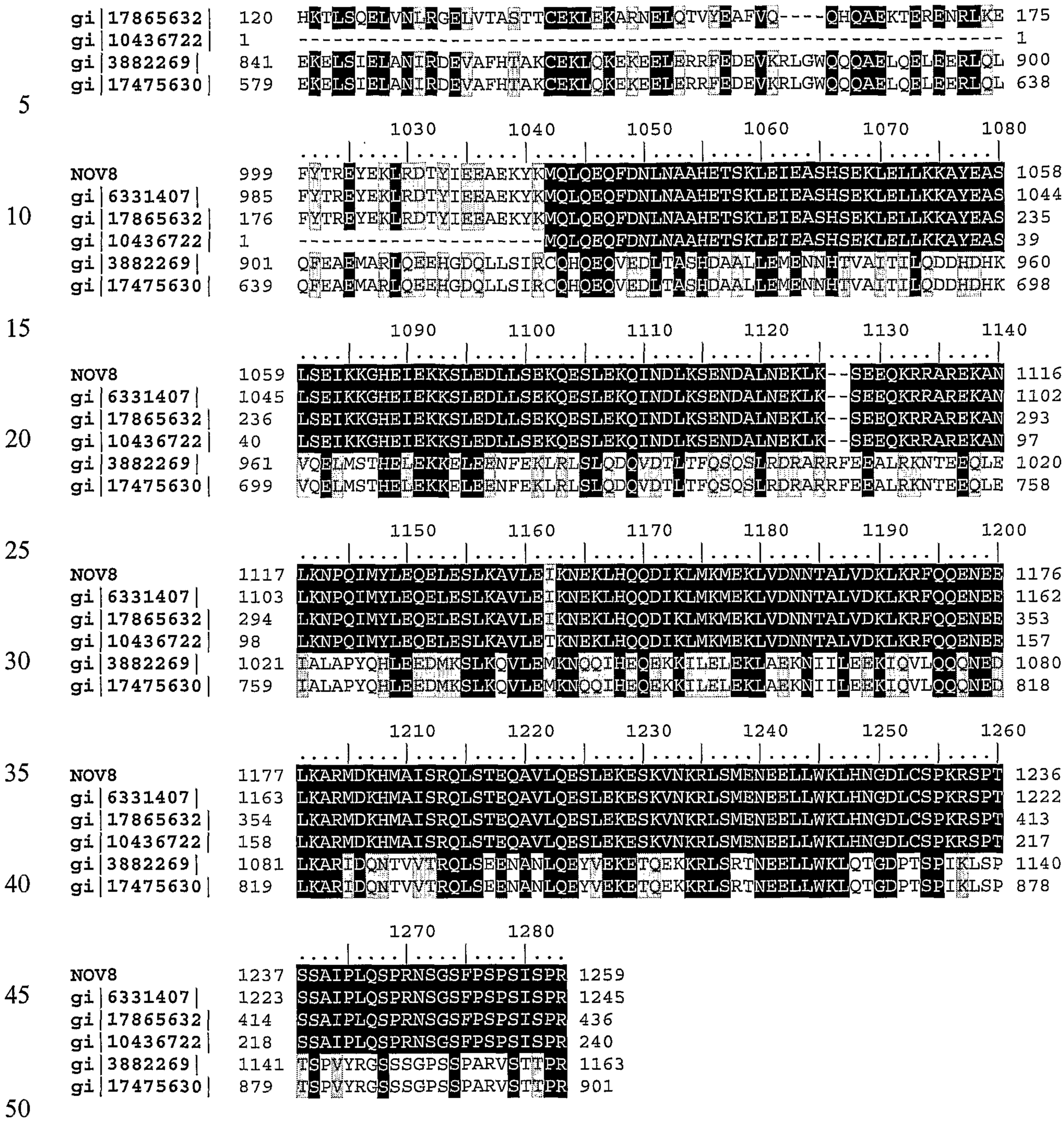
Table 8E. ClustalW Analysis of NOV8

1) > NOV8; SEQ ID NO:16
2) > gi|6331407|/ KIAA1288 protein [*Homo sapiens*]; SEQ ID NO:67
3) > gi|17865632|/ AT2 receptor-interacting protein 1 [*Homo sapiens*]; SEQ ID NO:68
4) > gi|10436722|/ unnamed protein product [*Homo sapiens*]; SEQ ID NO:69
5) > gi|3882269|/ KIAA0774 protein [*Homo sapiens*]; SEQ ID NO:70
6) > gi|17475630|/ KIAA0774 protein [*Homo sapiens*]; SEQ ID NO:71

```
                                    10        20        30        40        50        60
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8                1    -MNCKPSLPVIKMEITHAYNPKSPPTQNSSASSVNWNSANPDDMVVDYETDPAVVTG--- 56
gi|6331407|         1    ---------------THAYNPKSPPTQNSSASSVNWNSANPDDMVVDYETDPAVVTG--- 42
gi|17865632|        1    ------------------------------------------------------------ 1
gi|10436722|        1    ------------------------------------------------------------ 1
gi|3882269|         1    RGQIPGGGEGPQKTLPDHAVPAAFPATDSTSEGKSVRHPKPSTSESKQSTPSETQTVGAH 60
gi|17475630|        1    ------------------------------------------------------------ 1
```

```
                                    70        80        90        100       110       120
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8                 56     ------ENISLSLQGVEVFGHEKSSSDFISKQVLDMHKDSICQCPALVGTEKPKYLQHSC 110
gi|6331407|          42     ------ENISLSLQGVEVFGHEKSSSDFISKQVLDMHKDSICQCPALVGTEKPKYLQHSC 96
gi|17865632|         1      ------------------------------------------------------------ 1
gi|10436722|         1      ------------------------------------------------------------ 1
gi|3882269|          61     VLQVCSEHTSHSAHPEPALNLTLASKEIPSKLEAQLGQGKGEAKLDLKYVPPRRVEQEGK 120
gi|17475630|         1      ------------------------------------------------------------ 1

                                    130       140       150       160       170       180
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8                 111    HSLEAVEGQSVEPSLPFVWKPNDN----LNCAGYCDALELNQTFDMTVDKVNCTFISHHA 166
gi|6331407|          97     HSLEAVEGQSVEPSLPFVWKPNDN----LNCAGYCDALELNQTFDMTVDKVNCTFISHHA 152
gi|17865632|         1      ------------------------------------------------------------ 1
gi|10436722|         1      ------------------------------------------------------------ 1
gi|3882269|          121    AAQEGYLGCHKEENLSALEGRDPCGEAHPEATDALGHLLNSDLHHLGVGRGNCEEKRGVN 180
gi|17475630|         1      ------------------------------------------------------------ 1

                                    190       200       210       220       230       240
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8                 167    IGKSQSFHT---AGSLPPTGRRSGSTSSLSYSTWTSSHSDKTHARETTYDRESFENPQVT 223
gi|6331407|          153    IGKSQSFHT---AGSLPPTGRRSGSTSSLSYSTWTSSHSDKTHARETTYDRESFENPQVT 209
gi|17865632|         1      ------------------------------------------------------------ 1
gi|10436722|         1      ------------------------------------------------------------ 1
gi|3882269|          181    PGEQDSLHTTPKQGSASLGGADNQPTGKISPCAGEKLGERTSSSFSPGDSHVAFIPNNLT 240
gi|17475630|         1      ------------------------------------------------------------ 1

                                    250       260       270       280       290       300
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8                 224    PSEAQDMTYTAFSDVVMQSEVFVSDIGNQCACSSGKVTSEYTDGSQQRLVGEKETQALTP 283
gi|6331407|          210    PSEAQDMTYTAFSDVVMQSEVFVSDIGNQCACSSGKVTSEYTDGSQQRLVGEKETQALTP 269
gi|17865632|         1      ------------------------------------------------------------ 1
gi|10436722|         1      ------------------------------------------------------------ 1
gi|3882269|          241    DSKPLDVIEEERRLGSGNKDSVMVLVFNPSVGENKTEVPEPLDPQSGRSEARESKEVTTS 300
gi|17475630|         1      ----------------------MVLVFNPSVGENKTEVPEPLDPQSGRSEARESKEVTTS 38

                                    310       320       330       340       350       360
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8                 284    VSDGMEVPNDSALQEFFCLSHDESNSEPHSQSSYRHKEMGQNLRETVSYCLIDDECPLMV 343
gi|6331407|          270    VSDGMEVPNDSALQEFFCLSHDESNSEPHSQSSYRHKEMGQNLRETVSYCLIDDECPLMV 329
gi|17865632|         1      ------------------------------------------------------------ 1
gi|10436722|         1      ------------------------------------------------------------ 1
gi|3882269|          301    VAENRNLLENADKIESTSARADSVLNIP--APLHPETTVNMTYQPTTPSSSFQDVS---V 355
gi|17475630|         39     VAENRNLLENADKIESTSARADSVLNIP--APLHPETTVNMTYQPTTPSSSFQDVS---V 93

                                    370       380       390       400       410       420
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8                 344    PAFDKSEAQVLNPEHKVTETEDTQMVSKGKDLGTQNHTSELILSSPPGQKVGSSFGLTWD 403
gi|6331407|          330    PAFDKSEAQVLNPEHKVTETEDTQMVSKGKDLGTQNHTSELILSSPPGQKVGSSFGLTWD 389
gi|17865632|         1      ------------------------------------------------------------ 1
gi|10436722|         1      ------------------------------------------------------------ 1
gi|3882269|          356    FGMDAGSPLVVPPPTDSARLLNTSPKVPDKNTCPSGIPKPVFTHSKDTPSSQEGMENYQV 415
gi|17475630|         94     FGMDAGSPLVVPPPTDSARLLNTSPKVPDKNTCPSGIPKPVFTHSKDTPSSQEGMENYQV 153

                                    430       440       450       460       470       480
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8                 404    ANDMVISTDKTMCMSTPVLEPTKVTFSVSPIEATEKCKKVEKGNRGLKNIPDSKEAPVNL 463
gi|6331407|          390    ANDMVISTDKTMCMSTPVLEPTKVTFSVSPIEATEKCKKVEKGNRGLKNIPDSKEAPVNL 449
gi|17865632|         1      ------------------------------------------------------------ 1
gi|10436722|         1      ------------------------------------------------------------ 1
gi|3882269|          416    EKTEERTETKPIIMPKPKHVRPKIITYIRRNPQALGQVDASLVPVGLPYAPPTCTMPLPH 475
gi|17475630|         154    EKTEERTETKPIIMPKPKHVRPKIITYIRRNPQALGQVDASLVPVGLPYAPPTCTMPLPH 213

                                    490       500       510       520       530       540
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8                 464    CKPSLGKSTIKINTPIGCKVRKTEIISYPRPNFKNVKAKVMSRAVLQPKDAALSKVTPRP 523
gi|6331407|          450    CKPSLGKSTIKINTPIGCKVRKTEIISYPRPNFKNVKAKVMSRAVLQPKDAALSKVTPRP 509
gi|17865632|         1      ------------------------------------------------------------ 1
gi|10436722|         1      ------------------------------------------------------------ 1
```

```
gi|3882269|    476  EEKAAGGDLKPSANLYEKFKPDLQKPRVFSSGLMVSGIKPPGHPFSQMSEKFLQEVTDHP  535
gi|17475630|   214  EEKAAGGDLKPSANLYEKFKPDLQKPRVFSSGLMVSGIKPPGHPFSQMSEKFLQEVTDHP  273

                          550       560       570       580       590       600
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8               524  QQTSASSPSSVNSRQQTVLSRTPRSDLNADKKAEILINKTHKQQFNKLITSQAVHVTTHS  583
gi|6331407|        510  QQTSASSPSSVNSRQQTVLSRTPRSDLNADKKAEILINKTHKQQFNKLITSQAVHVTTHS  569
gi|17865632|       1    ------------------------------------------------------------  1
gi|10436722|       1    ------------------------------------------------------------  1
gi|3882269|        536  GKEEFCSPPYAHYEVPPTFYR--SAMLLKPQLGLGAMSRLPSAKSRILIASQ--RSSASA  591
gi|17475630|       274  GKEEFCSPPYAHYEVPPTFYR--SAMLLKPQLGLGAMSRLPSAKSRILIASQ--RSSASA  329

                          610       620       630       640       650       660
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8               584  KNASHRVPRTTSAVKSNQEDVDKASSSNSACETGSVSALFQK-IKGILPVKMESAECLEM  642
gi|6331407|        570  KNASHRVPRTTSAVKSNQEDVDKASSSNSACETGSVSALFQK-IKGILPVKMESAECLEM  628
gi|17865632|       1    ------------------------------------------------------------  1
gi|10436722|       1    ------------------------------------------------------------  1
gi|3882269|        592  IHPPGPITTATSLYSSDPSDLKKASSSNAAKSNLPKSGLRPPGYSRLPAAKLAAFGFVRS  651
gi|17475630|       330  IHPPGPITTATSLYSSDPSDLKKASSSNAAKSNLPKSGLRPPGYSRLPAAKLAAFGFVRS  389

                          670       680       690       700       710       720
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8               643  TYVPNIDRISPEKKGEKENGTSMEKQELKQEIMNETFEYGSLFLGSASKTTTTSGRNISK  702
gi|6331407|        629  TYVPNIDRISPEKKGEKENGTSMEKQELKQEIMNETFEYGSLFLGSASKTTTTSGRNISK  688
gi|17865632|       1    ------------------------------------------------------------  1
gi|10436722|       1    ------------------------------------------------------------  1
gi|3882269|        652  SSVSSVSSTQSGDSAQPEQGRPATRSTFGNEEQ------PVLKASLPSKDTPKGAGRVAP  705
gi|17475630|       390  SSVSSVSSTQSGDSAQPEQGRPATRSTFGNEEQ------PVLKASLPSKDTPKGAGRVAP  443

                          730       740       750       760       770       780
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8               703  PDSCGLRQIAAPKAKVGPPVSCLRRNSDNRNPSADRAVSPQRIRRVSSSGKPTSLKTAQS  762
gi|6331407|        689  PDSCGLRQIAAPKAKVGPPVSCLRRNSDNRNPSADRAVSPQRIRRVSSSGKPTSLKTAQS  748
gi|17865632|       1    -------MLLSPKFS----------------------LSTIHIR-LTAKGLLRNLR----  26
gi|10436722|       1    ------------------------------------------------------------  1
gi|3882269|        706  PASS--SVTAPRRSLLPAPKS----------------TSTPAGTKKDAPKDQDTNKPAVSS  748
gi|17475630|       444  PASS--SVTAPRRSLLPAPKS----------------TSTPAGTKKDAQKDQDTNKPAVSS  486

                          790       800       810       820       830       840
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8               763  SWVNLPRPLPKSKASLKSPALRRTGSTPSIASTHSELSTYSNNSGNAAVIKYEEKPPKPA  822
gi|6331407|        749  SWVNLPRPLPKSKASLKSPALRRTGSTPSIASTHSELSTYSNNSGNAAVIKYEEKPPKPA  808
gi|17865632|       26   ----LP------------SGFRR--------ST--------------------------  35
gi|10436722|       1    ------------------------------------------------------------  1
gi|3882269|        748  ----PKR-VAASTTKLHSPGYPK------------------------------------  766
gi|17475630|       486  ----PKR-VAASTTKLHSPGYPK------------------------------------  504

                          850       860       870       880       890       900
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8               823  FQNGSSGSFYLKPLVSRAHVHLMKTPPKGPSRKNLFTALNAVEKSRQKNPRSLCIQPQTA  882
gi|6331407|        809  FQNGSSGSFYLKPLVSRAHVHLMKTPPKGPSRKNLFTALNAVEKSRQKNPRSLCIQPQTA  868
gi|17865632|       35   ----------------------------------VVFHTVEKSRQKNPRSLCIQPQTA  59
gi|10436722|       1    ------------------------------------------------------------  1
gi|3882269|        766  ----------------------------------------QRTAAARN-GFPPKPD  781
gi|17475630|       504  ----------------------------------------QRTAAARN-GFPPKPD  519

                          910       920       930       940       950       960
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8               883  PDALPPEKTLELTQYKTKCENQSGFILQLKQLLACGNTKFEALTVVIQHLLSEREEALKQ  942
gi|6331407|        869  PDALPPEKTLELTQYKTKCENQSGFILQLKQLLACGNTKFEALTVVIQHLLSEREEALKQ  928
gi|17865632|       60   PDALPPEKTLELTQYKTKCENQSGFILQLKQLLACGNTKFEALTVVIQHLLSEREEALKQ  119
gi|10436722|       1    ------------------------------------------------------------  1
gi|3882269|        782  PQAREAERQLVLR-LKERCEQQTRQLGVAQGELKRAICGFDALAVATQHFFRKNESALVK  840
gi|17475630|       520  PQAREAERQLVLR-LKERCEQQTRQLGVAQGELKRAICGFDALAVATQHFFRKNESALVK  578

                          970       980       990      1000      1010      1020
                   ....|....|....|....|....|....|....|....|....|....|....|....|
NOV8               943  HKTLSQELVNLRGELVTASTTCEKLEKARNELQTVYEAFVQ----QHQAEKTERENRLKE  998
gi|6331407|        929  HKTLSQELVNLRGELVTASTTCEKLEKARNELQTVYEAFVQ----QHQAEKTERENRLKE  984
```



The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 8F lists the domain description from DOMAIN analysis results against NOV8.

Table 8F Domain Analysis of NOV8

Model	Region of Homology	Score (bits)	E value
RNA polymerase omega subunit	1008-1094	-12.8	5.7
Intermediate filament	967-1193	48.9	2.0e-06

Consistent with other known members of the Nuclear Protein-like family of proteins, NOV8 has, for example, an RNA polymerase omega subunit signature sequence and homology to other members of the Nuclear Protein-like Protein Family. NOV8 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV8 nucleic acids and polypeptides can be used to identify proteins that are members of the Nuclear Protein-like Protein Family. The NOV8 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV8 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, *e.g.,* cellular activation, cellular replication, and signal transduction. These molecules can be used to treat, *e.g.*, Cardiovascular diseases, 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, Diabetes,Von Hippel-Lindau (VHL) syndrome , Pancreatitis, Obesity, Hyperparathyroidism, Hypoparathyroidism as well as other diseases, disorders and conditions.

In addition, various NOV8 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV8 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the Nuclear Protein-like Protein Family.

The NOV8 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of cardiac or endocrine physiology. As such, the NOV8 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat infection, cardiovascular system, immune system, and nervous system disorders, *e.g.,* Cardiovascular diseases, 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, Diabetes,Von Hippel-Lindau (VHL) syndrome , Pancreatitis, Obesity, Hyperparathyroidism, Hypoparathyroidism as well as other diseases, disorders and conditions.

The NOV8 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV8 nucleic acid is expressed in Heart, Aorta, Coronary Artery, Vein, Umbilical Vein, Adrenal Gland/Suprarenal gland, Pancreas, Islets of Langerhans, Parathyroid Gland, Thyroid, Pineal Gland, Tongue, Salivary Glands, Stomach, Liver, Small Intestine, Colon, Ascending Colon, Lymphoid tissue, Spleen, Brain, Thalamus, Hypothalamus, Temporal Lobe, Amygdala, Cerebral Medulla/Cerebral white matter, Basal Ganglia/Cerebral nuclei, Substantia Nigra, Hippocampus, Spinal Chord, Cervix, Mammary gland/Breast, Ovary, Placenta, Uterus, Prostate, Testis, Lung, Nasoepithelium, Larynx, Urinary Bladder, Kidney, Kidney Cortex, Retina, Skin, Foreskin, Epidermis, Dermis.

Additional utilities for NOV8 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV9

A NOV9 polypeptide has been identified as a Hemicentin precursor-like protein (also referred to as CG94702-01). The disclosed novel NOV9 nucleic acid (SEQ ID NO:17) of 11796 nucleotides is shown in Table 9A. The novel NOV9 nucleic acid sequences maps to the chromosome 9.

An ORF begins with an ATG initiation codon at nucleotides 1-3 and ends with a TAA codon at nucleotides 11794-11796. A putative untranslated region and/or downstream from the termination codon is underlined in Table 9A, and the start and stop codons are in bold letters.

Table 9A. NOV9 Nucleotide Sequence (SEQ ID NO:17)

ATGTCTGCCTTATTTGCAGCTGTGTACCAGATGCTAAAACCACGCCTGGTCCATAACAGCCCACATC
CGGTGACCTATCAAATTGAGGCAAGTTTAAAGCCAGAGCAGCCTGGTGTCACGCTGGTGTCCATCCC
AGTCTTCCTGGCACCTTCCTGGCACAAAGCCTCAGAGCTGATCCCGACCCAGTCCTTCCGAGCACAG
GGGGCAGGGAAGCAGCTCCTCGGCTCTCCTTGCCCCCAAGTGCCCCCCAGCATCCGGGAGGACGGGC
GCAAGGCCAACGTGTCGGGTATGGCCGGGCAGTCCCTGACGCTGGAGTGTGACGCGAACGGCTTTCC
AGTCCCTGAGATCGTGTGGCTGAAGGACGCGCAGCTGATTCCTAAGGTGGGCGGCCACCGCCTCCTG
GACGAGGGCCAGTCCCTCCACTTCCCCAGGATCCAGGAGGGTGATTCTGGGCTCTACTCCTGCCGGG
CAGAGAACCAGGCTGGCACCGCCCAGAGGGACTTCCATCTCCTTGTGCTCACCCCTCCTTCCGTGCT
TGGAGCCGGGGCCGCTCAGGAGGTGCTAGGATTGGCCGGTGCAGACGTGGAGCTGCAGTGTTGGACC
TCAGGGGTCCCCACGCCCCAGGTGGAGTGGACCAAGGACAGGCAGCCTGTCCTTCCGGGAGGCCCTC
ACCTGCAGGTCCAGGAGGATGGCCAGGTTCTCAGGATCACCGGCAGTCACGTGGGGGATGAGGGACG
ATACCAGTGCGTGGCCTTCAGCCCAGCTGGTCAGCAGGCCAGGGACTTCCAGCTCCGAGTTCATGCG
CCCCCCACTATCTGGGGCTCCAACGAGACAGGCGAGGTGGCCGTCATGGAGGACCACCTAGTGCAGC
TCCTGTGTGAGGCTCGAGGAGTGCCCACCCCAAACATCACCTGGTTCAAGGACGGGGCCCTGCTCCC
CACCAGCACCAAGGTGGTCTACACTAGGGGCGGTCGGCAGTTGCAGCTGGGGAGGGCCCAGAGCTCC
GATGCCGGCGTCTACACCTGCAAGGCCAGCAATGCTGTGGGGGCCGCAGAGAAGGCCACCAGGCTGG
ATGTTTATGTCCCACCTACCATCGAGGGCGCCGGTGGAAGACCATACGTGGTGAAGGCTGTGGCTGG
GAGGCCTGTGGCGCTGGAGTGCGTGGCCAGAGGCCACCCGTCCCCCACCCTCTCCTGGCACCACGAG
GGGCTGCCCGTGGCAGAGAGCAACGAGTCGCGGCTGGAGACAGACGGGAGTGTGCTGAGGCTGGAGA

GCCCGGGGGAGGCATCCAGTGGCCTGTACAGCTGTGTGGCCAGCAGTCCTGCCGGGGAAGCCGTCCT
GCAGTACTCCGTGGAGGTTCAGGTGCCCCCACAGCTCCTGGTGGCTGAAGGCTTGGGACAGGTGACC
ACCATCGTGGGACAGCCCCTGGAACTTCCCTGCCAGGCCTCAGGCTCCCCAGTACCCACTATCCAGT
GGCTGCAGAATGGCCGCCCAGCCGAGGAGCTGGCTGGGGTGCAGGTGGCCTCGCAGGGGACCACACT
GCACATTGACCATGTGGAGCTGGACCACTCAGGCCTCTTCGCCTGCCAGGCCACCAATGAGGCGGGC
ACTGCCGGGGCCGAGGTGGAGGTGTCTGTGCATGAGTTCCCATCGGTCAGTATCATTGGGGGTGAGA
ACATCACAGCTCCTTTCCTGCAGCCTGTGACCCTCCAGTGCATAGGGGATGGGGTGCCCACCCCAAG
CCTCCGTTGGTGGAAGGATGGTGTAGCCCTGGCAGCCTTTGGGGGGAACCTACAGATTGAGAAGGTG
GACCTGAGGGACGAGGGCATCTACACTTGTGCTGCTACCAACCTGGCTGGGGAGAGCAAGAGGGAAG
TGGCGCTGAAAGTTTTGGTGCCCCCCAACATCGAGCCAGGCCCAGTCAACAAGGCAGTGCTGGAAAA
TGCCTCAGTGACCTTGGAGTGTCTGGCTTCGGGCGTGCCCCCTCCTGATGTCTCCTGGTTCAAGGGC
CACCAACCTGTCTCTTCATGGATGGGAGTGACAGTATCAGTGGATGGGAGAGTTCTCCGCATTGAGC
AAGCCCAGCTTTCTGATGCTGGGAGCTACCGCTGTGTGGCATCCAATGTGGCAGGTAGCACAGAGCT
GCGGTATGGCCTACGGGTCAATGTGCCCCCTCGAATCACACTGCCACCCAGCCTGCCAGGCCCTGTG
TTGGTCAACACCCCTGTCCGGCTGACCTGCAATGCCACCGGTGCCCCCAGCCCCACACTGATGTGGC
TGAAGGATGGAAACCCTGTGTCCCCTGCAGGGACCCCTGGCCTGCAGGTCTTCCCTGGGGGCCGGGT
CCTCACCTTGGCTAGTGCCCGGGCCTCCGACTCTGGGAGGTACTCCTGCGTGGCTGTGAGCGCGGTG
GGCGAGGACCGCCAGGATGTTGTCCTGCAAGTCCACATGCCCCCGAGTATCCTTGGAGAAGAGCTGA
ATGTGTCCGTTGTGGCCAATGAGTCAGTGGCCCTGGAGTGCCAGAGCCACGCCATGCCCCCTCCTGT
GCTGAGCTGGTGGAAGGACGGGCGGCCCCTGGAACCACGGCCTGGAGTCCACCTCTCCGCAGACAAA
GCCTTGCTGCAGGTGGACAGAGCCGATGTGTGGGATGCGGGCCATTACACCTGTGAGGCACTGAACC
AGGCCGGCCACTCAGAGAAACACTACAATCTGAACGTCTGGGGTCAACCCCTCCCCGGGGAGGGGGC
TGGCCTCCAGCACGTGTCGGCTGTGGGGAGGCTGTTGTACCTGGGACAGGCCCAGCTGGCTCAGGAA
GGAACATACACCTGTGAATGCAGCAACGTGGTGGGGAACAGCAGCCAGGACCTGCAGCTGGAGGTGC
ACGTTCCCCCTCAGATTGCCGGTCCCCGGGAGCCTCCCACACAAGTCTCTGTGGTCCAGGATGGAGT
GGCCACTCTGGAGTGCAACGCCACAGGGAAACCCCCTCCGACAGTGACATGGGAGCGGGACGGCCAG
CCCGTGGGGGCTGAACTGGGCCTGCAGCTGCAGAACCAGGGTCAGAGCCTGCATGTGGAGCGGGCCC
AGGCTGCCCACACTGGACGCTACAGCTGTGTGGCCGAGAACCTGGCTGGGAGGGCAGAGAGGAAGTT
TGAGCTCTCCGTACTGGTGCCCCCAGAGCTCATTGGAGACTTGGACCCGCTGACCAACATCACTGCT
GCCTTGCACAGCCCCTTAACTCTGCTCTGTGAAGCCATGGGGATCCCACCTCCAGCCATCCGCTGGT
TCCGAGGGGAGGAGCCTGTCAGCCCCGGGGAGGACACCTACCTGCTGGCAGGTGGCTGGATGCTGAA
GATGACTCAGACACAGGAGCAAGACAGTGGCCTCTACTCATGCCTGGCAAGCAACGAGGCTGGGGAG
GCACGGAGGAACTTCAGTGTGGAGGTGCTGGTTCCTCCCAGTATTGAGAACGAGGACTTGGAGGAGG
TGATCAAGGTCCTTGATGGACAGACTGCCCATCTTATGTGCAACGTCACAGGCCACCCACAGCCCAA
GCTCACATGGTTCAAAGATGGCCGGCCTCTGGCTAGGGGAGATGCTCACCACATCTCCCCAGACGGA
GTCCTCCTGCAGGTCCTCCAGGCAAACCTGTCCAGTGCTGGCCACTACTCCTGCATTGCAGCCAACG
CTGTTGGGGAGAAGACCAAACACTTCCAGCTCAGTGTCCTGTTGGCTCCCACCATCCTGGGAGGGGC
CGAGGACAGTGCAGATGAGGAGGTGACCGTGACTGTCAACAACCCCATCTCTCTGATCTGCGAGGCC
CTGGCCTTCCCTTCCCCCAACATCACCTGGATGAAGGACGGGGCCCCGTTTGAGGCCTCCAGGAACA
TCCAGCTGCTCCCAGGTACCCACGGGCTGCAGATCCTGAATGCCCAGAAGGAAGATGCTGGCCAGTA
CACCTGCGTGGTCACCAATGAGCTCGGGGAGGCCGTGAAAAACTACCATGTGGAAGTGCTCATCCCC
CCTTCCATCTCCAAAGACGACCCCTTGGCGGAGGTCGGCGTGAAGGAGGTGAAGACCAAGGTCAACA
GCACCTTGACCTTGGAGTGTGAGAGCTGGGCTGTGCCCCCGCCCACCATCCGCTGGTACAAGGATGG
ACAGCCCGTGACCCCCAGCTCGCGGCTGCAGGTCCTGGGTGAAGGGCGACTGCTCCAGATCCAGCCC
ACACAGGTCTCAGACTCGGGGCGGTACCTGTGTGTGGCCACCAATGTGGCTGGCGAGGACGACCAGG
ACTTCAACGTGCTCATCCAGGTGCCCCCCATGTTCCAGAAGGTGGGTGATTTCAGTGCAGCCTTCGA
GATCCTGTCCCGGGAGGAGGAGGCCCGGGGCGGAGTCACGGAATACAGGGAGATCGTGGAGAACAAC
CCAGCCTACCTGTACTGCGACACCAACGCGATCCCACCCCCGGACCTCACCTGGTACAGAGAGGATC
AGCCCCTCTCGGCCGGGGATGAGGTGTCTGTGCTGCAAGGAGGCCGGGTCCTGCAGATCCCCCTGGT
GCGGGCAGAGAACGCCGGGAGGTACTCGTGCAAGGCCTCCAACGAGGTGGGCGAGGACTGGCTGCAC
TACGAGCTGCTGGTGCTGACCCCACCTGTGATCCTGGGTGACACAGAGGAGCTGGTGGAAGAGGTGA
CAGTCAATGCCAGCAGCACCGTCAGCCTGCAGTGCCCGGCCCTGGGAAACCCCGTGCCCACCATCTC
ATGGCTCCAGAATGGGCTGCCTTTCTCCCCGAGCCCACGGCTGCAGGTCCTGGAGGACGGGCAAGTC
TTGCAGGTTTCCACGGCAGAGGTGGCCGACGCCGCCAGCTACATGTGTGTGGCCGAGAACCAGGCGG
GCTCCGCTGAGAAGCTCTTCACCCTCAGGGTTCAAGGCCTGGACTTGGAGCAGGTCACTGCCATCCT
CAACAGCAGCGTCTCCCTCCCTTGCGACGTCCACGCTCACCCAAACCCCGAGGTCACGTGGTACAAG
GACAGCCAGGCCCTCTCCCTGGGTGAAGAGGTCTTCCTCCTGCCTGGCACCCACACGCTGCAGCTGG
GGAGAGCACGGCTGTCGGACTCCGGGATGTACACATGCGAAGCCCTCAATGCTGCCGGCCGAGACCA
GAAGCTGGTGCAGCTCAGTGTTCTGGTTCCCCCGGCCTTCAGGCAGGCTCCCAGAGGTCCCCAGGAT
GCGGTCCTGGTGAGGGTCGGGGACAAAGCTGTCCTGAGCTGCGAGACAGATGCGCTCCCTGAGCCAA
CTGTGACCTGGTACAAGGATGGGCAGCCCCTGGTCCTGGCACAGCGGACCCAGGCTCTGCGGGGTGG
GCAGAGGCTGGAGATCCAGGAAGCCCAGGTATCGGATAAAGGTTTATACAGCTGTAAAGTCAGCAAC

GTGGCTGGGGAGGCCGTGCGGACCTTCACCCTCACCGTCCAGGTGCCCCCAACATTTGAGAACCCCA
AGACAGAGACAGTGAGCCAGGTGGCTGGGAGCCCCCTGGTCCTGACCTGTGATGTGTCCGGGGTCCC
TGCACCCACGGTCACTTGGCTGAAGGACAGGATGCCTGTGGAGAGCAGCGCGGTGCACGGTGTGGTC
TCCCGGGGGGGCCGCCTCCAGCTGAGCCGCCTGCAACCGGCCCAGGCGGGCACCTACACGTGCGTGG
CTGAGAACACCCAGGCTGAGGCCCGCAAGGACTTCGTGGTAGCAGTGCTGGTGGCCCCCCGGATCCG
GAGCTCGGGCGTGGCGCGGGAGCACCATGTCTTGGAAGGGCAGGAGGTGCGGCTGGACTGTGAGGCC
GATGGGCAGCCGCCGCCGGACGTGGCCTGGCTGAAGGACGGCAGCCCGCTGGGCCAGGACATGGGCC
CCCACCTCCGGTTCTACCTGGACGGCGGCTCCCTGGTGCTAAAAGGCCTGAGGGCCTCGGACGCGGG
TGCCTACACCTGCGTGGCCCACAACCCAGCCGGGGAGGACGCCAGGCTGCACACGGTGAATGTGCTG
GTTCCTCCCACCATCAAGCAGGGAGCAGACGGCTCGGGGACCCTGGTGAGCAGGCCTGGGGAGCTGG
TGACCATGGTGTGCCCTGTGCGGGGCTCCCCGCCCATCCACGTGAGCTGGCTCAAGGACGGCCTGCC
CCTCCCGCTCTCCCAGCGCACCCTCCTCCACGGCTCTGGCCACACCCTCAGGATTTCCAAGGTGCAA
TTGGCAGACGCTGGCATCTTCACCTGTGTGGCCGCAAGCCCAGCTGGCGTGGCGGACAGGAACTTCA
CCTTGCAGGTGCAGGTGCCCCCTGTCCTGGAGCCGGTGGAGTTCCAGAATGACGTGGTGGTGGTTCG
TGGCTCCCTGGTGGAACTCCCGTGCGAGGCCCGGGGCGTTCCCCTGCCTCTCGTGTCGTGGATGAAG
GATGGGGAACCCTTGTTGTCCCAGAGCCTCGAGCAGGGGCCCAGCCTGCAGCTGGAGGCAGTGGGAG
CTGGTGACTCGGGGACCTACTCCTGTGTGGCCGTGAGCGAGGCGGGGGAAGCCAGGAGGCATTTCCA
GCTGACCGTCATGGAGCCCCCTCACATTGAGGACTCAGGCCAGCCTACAGAGCTGTCGCTGACCCCC
GGCGCCCCCATGGAGCTCCTCTGTGATGCCCAGGGCACCCCCCAGCCCAACATCACCTGGCATAAGG
ACGGGCAGGCCCTGACCAGGCTGGAGAACAACAGCAGAGCCACACGGGTGCTCCGGGTGGAGAATGT
GCAGGTTAGGGATGCTGGGCTGTACACTTGTCTGGCTGAAAGCCCTGCAGGTGCAATTGAGAAGAGC
TTCCGGGTCAGGGTTCAAGCCCCTCCAAACATTGTTGGGCCCCGAGGCCCCCGCTTTGTGGTCGGCC
TGGCCCCAGGGCAGCTGGTCCTGGAGTGTTCGGTGGAGGCAGAGCCAGCGCCCAAGATCACGTGGCA
CCGAGACGGCATTGTGCTGCAGGAGGACGCCCACACACAATTCCCGGAGCGGGGCAGGTTCCTCCAG
CTGCAGGCCCTGAGCACGGCTGACAGCGGCGACTACAGCTGCACAGCCCGCAACGCCGCAGGCAGCA
CTAGTGTCGCCTTCCGCGTGGAGATCCACACGGTGCCCACCATCCGGTCAGGACCACCTGCAGTGAA
CGTCTCAGTGAACCAGACAGCCCTGCTGCCTTGCCAGGCCGACGGCGTGCCCGCACCCCTCGTGAGC
TGGCGGAAGGACAGGGTCCCCCTGGATCCCAGGAGCCCCAGGGCAACCCCCATCCATTCTAGGTTTG
AAATTCTGCCTGAGGGTTCCCTGAGAATCCAGCCAGTCCTTGCCCAGGACGCCGGCCACTACCTCTG
CCTGGCATCCAACTCTGCTGGCTCCGATCGTCAAGGCCGTGACCTACGGGTCTTGGAGCCTCCAGCC
ATCGCCCCCAGCCCCTCCAACCTGACCCTGACCGCCCACACCCCAGCCTTGCTGCCCTGCGAGGCCA
GCGGCTCCCCTAAGCCCCTGGTGGTCTGGTGGAAGGACGGACAGAAGCTGGACTTCCGCCTGCAGCA
GGGCGCCTACCGGCTCCTGCCCTCCAACGCCCTGCTCCTCACGGCCCCCGGCCCCCAGGACTCAGCC
CAGTTTGAATGCGTGGTGAGCAATGAGGTGGGCGAGGCCCACAGGCTCTACCAGGTGACCGTCCATG
TGCCTCCCACCATTGCCGATGACCAGACAGACTTCACCGTGACCATGATGGCACCTGTGGTCCTCAC
ATGTCACAGCACGGGTATACCAGCTCCGACCGTGTCCTGGAGCAAGGCAGGCGCCCAGCTAGGAGCT
CGGGGGAGTGGCTATCGTGTCTCACCATCGGGCGCCCTGGAGATCGGGCAGGCCCTCCCCATCCACG
CAGGCCGCTACACCTGCTCAGCCCGCAACTCTGCCGGCGTAGCCCACAAGCACGTCTTCCTCACTGT
GCAAGCCTCCCCGGTGGTGAAGCCGCTGCCCAGCGTGGTTCGGGCAGTGGCAGAGGAGGAGGTGCTG
CTGCCCTGCGAGGCCTCAGGCATCCCCCGGCCGACCATCACCTGGCAGAAGGAAGGGCTCAACGTCG
CTACTGGAGTGAGTACCCAGGTCCTACCAGGCGGACAGCTGCGGATTGCCCATGCCAGCCCAGAGGA
TGCTGGAAACTATCTCTGCATCGCTAAGAACAGTGCGGGCAGTGCCATGGGGAAGACGCGGCTGGTG
GTGCAAGTCCCACCAGTGATCGAGAATGGCCTCCCAGACCTGTCCACCACCGAAGGCTCCCACGCCT
TCTTGCCTTGCAAGGCGAGGGGCAGTCCTGAGCCCAACATCACCTGGGACAAAGATGGCCAGCCTGT
GTCGGGCGCCGAGGGGAAGTTCACCATCCAGCCTTCTGGGGAGTTGCTGGTGAAGAACTTGGAGGGC
CAGGACGCAGGCACCTATACCTGTACCGCTGAGAACGCCGTGGGCCGGGCCCGCCGCCGCGTGCACC
TCACCATCCTGGTACTGCCTGTGTTCACCACCCTGCCTGGGGACCGCAGCCTGCGCCTTGGGGACAG
GCTGTGGCTTCGCTGTGCAGCCCGGGGCAGCCCCACCCCTCGCATTGGCTGGACTGTCAACGACCGG
CCAGTCACAGAAGGGGTGTCTGAGCAGGATGGAGGCAGCACGCTGCAGCGGGCCGCTGTCTCCAGAG
AAGACAGCGGGACCTATGTCTGCTGGGCGGAGAACAGAGTGGGCCGCACGCAGGCGGTCAGCTTCGT
CCACGTGAAGGAGGCTCCTGTCCTACAAGGGGAGGCTTTCTCCTACCTGGTGGAACCTGTAGGAGGC
AGCATTCAGCTAGACTGTGTGGTGCGTGGAGACCCAGTGCCGGACATCCACTGGATCAAAGATGGCC
TTCCACTGCGGGGCAGCCACCTCCGGCACCAGCTGCAGAATGGCTCGCTGACCATCCGCAGGACTGA
GGCAAGGCGGGGCCTGGCACCTTGGAGGGACGATGCGGGACGGTACCAGTGCCTGGCAGAGAATGAG
ATGGGCGTGGCGAAGAAAGTGGTGATCCTCGTCCTGCAGACCAGGATGGTGCCAGCAGAGCCCCACT
TGAAGCGCCAACTCCCACCGATCCCCAGCAATAATGAGGCACCCTCCCTGTTCCCGGGTGTCCATGG
AGGCCACGTGGGGAACCCGGACTTCCACTCTCATCTAGCAGAAGTTCTCGCCGTTCAGTTGCTGGCT
GGGTCCCTGCTCTTCTCAGCCAGGGCCATGCCGCAGGCCAGCACAGCAGCCATTTCCCTTTTGGCTC
CTACCAGTTTTGCCCCTTTTCCTGATGATATTTCTCAGGGCATACTTTCATCCTCTACTGCACATCA
AGGCAGCCCCCAGGGGTGGCAAAAGCTGCTGTTTTTCACAGCCATCCCTAATAAAACCACTGTGATG
GTCACGGTGGAGCCCCAGGACATGACAGTGAGATCTGGGGATGACGTGGCCCTGCGGTGCCAGGCCA
CTGGAGAGCCCACACCCACCATTGAATGGCTACAGGCGGGTCAACCCTTGCGGGCCAGCCGGCGGCT

CCGGACCCTGCCCGATGGGAGCCTGTGGCTGGAGAACGTGGAGACTGGGGATGCAGGCACCTACGAC
TGCGTCGCTCACAACCTCCTGGGCTCTGCCACAGCCCGGGCGTTCCTGGTCTGTGCCAGCCACGCCA
TCGTGGGCTCCCGGCATTTCAGAGACCCACAGGTCTTCTGTGAGTTTGTGGTCCCGCCTCCTCATTT
TACAGGGGAGCCCCAGGGGAGCTGGGGCAGCATGACTGGGGTGATAAATGGCCGGAAATTTGGCGTG
GCCACACTCAACACCAGCGTGATGCAGGAGGCACACTCCGGGGTCAGCAGCATCCACAGCAGCATCC
GCCATGTCCCAGCAAACGTGGGGCCTCTGATGCGGGTGCTCGTGGTCACCATCGCCCCCATCTACTG
GGCCCTGGCCAGAGAGAGTGGGGAAGCCCTGAATGGCCACTCTCTGACTGGGGGCAGGTTCCGGCAG
GAGTCACACGTGGAGTTTGCTACAGGGGAGCTGCTCACGATGACCCAGGTGGCCCGGGGTCTGGATC
CCGATGGCCTCCTGCTCCTCGACGTGGTGGTCAATGGCGTTGTCCCCGAGAGCCTGGCTGACGCAGA
TCTTCAAGTGCAGGACTTTGAGGAGCACTACGTGCAAACAGGGCCTGGCCAGCTGTTCGTGGGCTCC
ACACAGCGCTTCTTCCAGGGCGGCCTCCCCTCGTTCCTACGCTGCAACCACAGCATCCAGTACAACG
CGGCCCGGGGCCCCCAGCCCCAGCTGGTGCAGCACCTGCGGGCCTCAGCTATCAGCTCGGCCTTTGA
TCCAGAGGCCGAGGCCCTGCGCTTCCAGCTCGCTACAGCCCTGCAGGCGGAGGAGAACGAGGTCGGC
TGCCCCGAGGGCTTTGAGCTGGACTCCCAGGGAGCGTTTTGTGTGGACAGGGACGAGTGCTCAGGAG
GCCCTAGCCCCTGCTCCCATGCCTGCCTTAATGCACCCGGCCGCTTCTCCTGCACCTGCCCCACTGG
CTTCGCCCTGGCCTGGGATGACAGGAACTGCAGAGATGTGGACGAGTGTGCGTGGGATGCTCACCTC
TGCCGAGAGGGACAGCGCTGTGTGAACCTGCTCGGGTCCTACCGCTGCCTCCCCGACTGTGGGCCTG
GCTTCCGGGTGGCTGATGGGGCCGGCTGTGAAGATGTGGACGAATGCCTGGAGGGGTTGGACGACTG
TCACTACAACCAGCTCTGCGAGAACACCCCAGGCGGTCACCGCTGCAGCTGCCCCAGGGGTTACCGG
ATGCAGGGCCCCAGCCTGCCCTGCCTAGATGTCAATGAGTGCCTGCAGCTGCCCAAGGCCTGCGCCT
ACCAGTGCCACAACCTCCAGGGCAGCTACCGCTGCCTGTGCCCCCCAGGCCAGACCCTCCTTCGCGA
CGGCAAGGCCTGCACCTCACTGGAGCGGAATGGACAAAATGTGACCACCGTCAGCCACCGAGGCCCT
CTATTGCCCTGGCTGCGGCCCTGGGCCTCGATCCCCGGTACCTCCTACCACGCCTGGGTCTCTCTCC
GTCCGGGTCCCATGGCCCTGAGCAGTGTGGGCCGGGCCTGGTGCCCTCCTGGTTTCATCAGGCAGAA
CGGAGTCTGCACAGACCTTGACGAGTGCCGCGTGAGGAACCTGTGTCAGCACGCCTGCCGCAACACT
GAGGGCAGCTACCAGTGCCTGTGCCCCGCCGGCTACCGTCTGCTCCCCAGCGGGAAGAACTGCCAGG
ACATCAACGAGTGCGAGGAGGAGAGCATCGAGTGTGGACCCGGCCAGATGTGCTTCAACACCCGTGG
CAGCTACCAGTGTGTGGACACACCCTGTCCTGCCACCTACCGGCAGGGCCCCAGCCCTGGGACGTGC
TTCCGGCGCTGCTCGCAGGACTGCGGCACGGGCGGCCCCTCTACGCTGCAGTACCGGCTGCTGCCGC
TGCCCCTGGGCGTGCGCGCCCACCACGACGTGGCCCGCCTCACCGCCTTCTCCGAGGTCGGCGTCCC
CGCCAACCGCACCGAGCTCAGCATGCTGGAGCCCGACCCCCGCAGCCCCTTCGCGCTGCGTCCGCTG
CGCGCGGGCCTTGGCGCGGTCTACACCCGTCGCGCGCTCACCCGCGCCGGCCTCTACCGGCTCACCG
TGCGTGCTGCGGCACCGCGCCACCAAAGCGTCTTCGTCTTGCTCATCGCCGTGTCCCCCTACCCCTA
C**TAA**

Variant sequences of NOV9 are included in Example 3, Table 21. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA.

The NOV9 protein (SEQ ID NO:18) encoded by SEQ ID NO:17 is 3931 amino acid residues in length and is presented using the one-letter amino acid code in Table 9B. Psort analysis predicts the NOV9 protein of the invention to be localized at the plasma membrane with a certainty of 0.7300.

Table 9B. Encoded NOV9 protein sequence (SEQ ID NO:18)

MSALFAAVYQMLKPRLVHNSPHPVTYQIEASLKPEQPGVTLVSIPVFLAPSWHKASELIPTQSFR
AQGAGKQLLGSPCPQVPPSIREDGRKANVSGMAGQSLTLECDANGFPVPEIVWLKDAQLIPKVGG
HRLLDEGQSLHFPRIQEGDSGLYSCRAENQAGTAQRDFHLLVLTPPSVLGAGAAQEVLGLAGADV
ELQCWTSGVPTPQVEWTKDRQPVLPGGPHLQVQEDGQVLRITGSHVGDEGRYQCVAFSPAGQQAR
DFQLRVHAPPTIWGSNETGEVAVMEDHLVQLLCEARGVPTPNITWFKDGALLPTSTKVVYTRGGR
QLQLGRAQSSDAGVYTCKASNAVGAAEKATRLDVYVPPTIEGAGGRPYVVKAVAGRPVALECVAR
GHPSPTLSWHHEGLPVAESNESRLETDGSVLRLESPGEASSGLYSCVASSPAGEAVLQYSVEVQV

PPQLLVAEGLGQVTTIVGQPLELPCQASGSPVPTIQWLQNGRPAEELAGVQVASQGTTLHIDHVE
LDHSGLFACQATNEAGTAGAEVEVSVHEFPSVSIIGGENITAPFLQPVTLQCIGDGVPTPSLRWW
KDGVALAAFGGNLQIEKVDLRDEGIYTCAATNLAGESKREVALKVLVPPNIEPGPVNKAVLENAS
VTLECLASGVPPPDVSWFKGHQPVSSWMGVTVSVDGRVLRIEQAQLSDAGSYRCVASNVAGSTEL
RYGLRVNVPPRITLPPSLPGPVLVNTPVRLTCNATGAPSPTLMWLKDGNPVSPAGTPGLQVFPGG
RVLTLASARASDSGRYSCVAVSAVGEDRQDVVLQVHMPPSILGEELNVSVVANESVALECQSHAM
PPPVLSWWKDGRPLEPRPGVHLSADKALLQVDRADVWDAGHYTCEALNQAGHSEKHYNLNVWGQP
LPGEGAGLQHVSAVGRLLYLGQAQLAQEGTYTCECSNVVGNSSQDLQLEVHVPPQIAGPREPPTQ
VSVVQDGVATLECNATGKPPPTVTWERDGQPVGAELGLQLQNQGQSLHVERAQAAHTGRYSCVAE
NLAGRAERKFELSVLVPPELIGDLDPLTNITAALHSPLTLLCEAMGIPPPAIRWFRGEEPVSPGE
DTYLLAGGWMLKMTQTQEQDSGLYSCLASNEAGEARRNFSVEVLVPPSIENEDLEEVIKVLDGQT
AHLMCNVTGHPQPKLTWFKDGRPLARGDAHHISPDGVLLQVLQANLSSAGHYSCIAANAVGEKTK
HFQLSVLLAPTILGGAEDSADEEVTVTVNNPISLICEALAFPSPNITWMKDGAPFEASRNIQLLP
GTHGLQILNAQKEDAGQYTCVVTNELGEAVKNYHVEVLIPPSISKDDPLAEVGVKEVKTKVNSTL
TLECESWAVPPPTIRWYKDGQPVTPSSRLQVLGEGRLLQIQPTQVSDSGRYLCVATNVAGEDDQD
FNVLIQVPPMFQKVGDFSAAFEILSREEEARGGVTEYREIVENNPAYLYCDTNAIPPPDLTWYRE
DQPLSAGDEVSVLQGGRVLQIPLVRAENAGRYSCKASNEVGEDWLHYELLVLTPPVILGDTEELV
EEVTVNASSTVSLQCPALGNPVPTISWLQNGLPFSPSPRLQVLEDGQVLQVSTAEVADAASYMCV
AENQAGSAEKLFTLRVQGLDLEQVTAILNSSVSLPCDVHAHPNPEVTWYKDSQALSLGEEVFLLP
GTHTLQLGRARLSDSGMYTCEALNAAGRDQKLVQLSVLVPPAFRQAPRGPQDAVLVRVGDKAVLS
CETDALPEPTVTWYKDGQPLVLAQRTQALRGGQRLEIQEAQVSDKGLYSCKVSNVAGEAVRTFTL
TVQVPPTFENPKTETVSQVAGSPLVLTCDVSGVPAPTVTWLKDRMPVESSAVHGVVSRGGRLQLS
RLQPAQAGTYTCVAENTQAEARKDFVVAVLVAPRIRSSGVAREHHVLEGQEVRLDCEADGQPPPD
VAWLKDGSPLGQDMGPHLRFYLDGGSLVLKGLRASDAGAYTCVAHNPAGEDARLHTVNVLVPPTI
KQGADGSGTLVSRPGELVTMVCPVRGSPPIHVSWLKDGLPLPLSQRTLLHGSGHTLRISKVQLAD
AGIFTCVAASPAGVADRNFTLQVQVPPVLEPVEFQNDVVVVRGSLVELPCEARGVPLPLVSWMKD
GEPLLSQSLEQGPSLQLEAVGAGDSGTYSCVAVSEAGEARRHFQLTVMEPPHIEDSGQPTELSLT
PGAPMELLCDAQGTPQPNITWHKDGQALTRLENNSRATRVLRVENVQVRDAGLYTCLAESPAGAI
EKSFRVRVQAPPNIVGPRGPRFVVGLAPGQLVLECSVEAEPAPKITWHRDGIVLQEDAHTQFPER
GRFLQLQALSTADSGDYSCTARNAAGSTSVAFRVEIHTVPTIRSGPPAVNVSVNQTALLPCQADG
VPAPLVSWRKDRVPLDPRSPRATPIHSRFEILPEGSLRIQPVLAQDAGHYLCLASNSAGSDRQGR
DLRVLEPPAIAPSPSNLTLTAHTPALLPCEASGSPKPLVVWWKDGQKLDFRLQQGAYRLLPSNAL
LLTAPGPQDSAQFECVVSNEVGEAHRLYQVTVHVPPTIADDQTDFTVTMMAPVVLTCHSTGIPAP
TVSWSKAGAQLGARGSGYRVSPSGALEIGQALPIHAGRYTCSARNSAGVAHKHVFLTVQASPVVK
PLPSVVRAVAEEEVLLPCEASGIPRPTITWQKEGLNVATGVSTQVLPGGQLRIAHASPEDAGNYL
CIAKNSAGSAMGKTRLVVQVPPVIENGLPDLSTTEGSHAFLPCKARGSPEPNITWDKDGQPVSGA
EGKFTIQPSGELLVKNLEGQDAGTYTCTAENAVGRARRRVHLTILVLPVFTTLPGDRSLRLGDRL
WLRCAARGSPTPRIGWTVNDRPVTEGVSEQDGGSTLQRAAVSREDSGTYVCWAENRVGRTQAVSF
VHVKEAPVLQGEAFSYLVEPVGGSIQLDCVVRGDPVPDIHWIKDGLPLRGSHLRHQLQNGSLTIR
RTEARRGLAPWRDDAGRYQCLAENEMGVAKKVVILVLQTRMVPAEPHLKRQLPPIPSNNEAPSLF
PGVHGGHVGNPDFHSHLAEVLAVQLLAGSLLFSARAMPQASTAAISLLAPTSFAPFPDDISQGIL
SSSTAHQGSPQGWQKLLFFTAIPNKTTVMVTVEPQDMTVRSGDDVALRCQATGEPTPTIEWLQAG
QPLRASRRLRTLPDGSLWLENVETGDAGTYDCVAHNLLGSATARAFLVCASHAIVGSRHFRDPQV
FCEFVVPPPHFTGEPQGSWGSMTGVINGRKFGVATLNTSVMQEAHSGVSSIHSSIRHVPANVGPL
MRVLVVTIAPIYWALARESGEALNGHSLTGGRFRQESHVEFATGELLTMTQVARGLDPDGLLLLD
VVVNGVVPESLADADLQVQDFEEHYVQTGPGQLFVGSTQRFFQGGLPSFLRCNHSIQYNAARGPQ
PQLVQHLRASAISSAFDPEAEALRFQLATALQAEENEVGCPEGFELDSQGAFCVDRDECSGGPSP
CSHACLNAPGRFSCTCPTGFALAWDDRNCRDVDECAWDAHLCREGQRCVNLLGSYRCLPDCGPGF
RVADGAGCEDVDECLEGLDDCHYNQLCENTPGGHRCSCPRGYRMQGPSLPCLDVNECLQLPKACA
YQCHNLQGSYRCLCPPGQTLLRDGKACTSLERNGQNVTTVSHRGPLLPWLRPWASIPGTSYHAWV
SLRPGPMALSSVGRAWCPPGFIRQNGVCTDLDECRVRNLCQHACRNTEGSYQCLCPAGYRLLPSG
KNCQDINECEEESIECGPGQMCFNTRGSYQCVDTPCPATYRQGPSPGTCFRRCSQDCGTGGPSTL
QYRLLPLPLGVRAHHDVARLTAFSEVGVPANRTELSMLEPDPRSPFALRPLRAGLGAVYTRRALT
RAGLYRLTVRAAAPRHQSVFVLLIAVSPYPY

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

Table 9C. Patp results for NOV9			
Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAY53667 Sequence gi/3328186	+1	1529	6.5e-244
>patp:AAY87206 Human secreted protein sequence ID NO:245	+1	2235	6.4e-230
>patp:AAE06183 Human gene 57 encoded secreted protein	+1	2235	6.4e-230
>patp:AAY87120 Human secreted protein sequence SEQ ID:159	+1	2235	6.4e-230
>patp:AAE06097 Human gene 57 secreted protein HRACD80	+1	2235	6.4e-230

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 625 of 1067 bases (58%) identical to a gb:GENBANK-ID:HSLTGFBP4|acc:Y13622.1 mRNA from *Homo sapiens* (mRNA for latent transforming growth factor-beta binding protein-4). The full amino acid sequence of the protein of the invention was found to have 502 of 1665 amino acid residues (30%) identical to, and 767 of 1665 amino acid residues (46%) similar to, the 5198 amino acid residue ptnr:SPTREMBL-ACC:O76518 protein from *Caenorhabditis elegans* (HEMICENTIN PRECURSOR).

NOV9 also has homology to the proteins shown in the BLASTP data in Table 9D.

Table 9D. BLAST results for NOV9					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|14575679|gb|AAK68690.1|AF156100_1 (AF156100)	hemicentin [*Homo sapiens*]	5636	1230/3017 (40%)	1785/3017 (58%)	0.0
gi|18547943|ref|XP_053531.3|(XM_053531)	hemicentrin [*Homo sapiens*]	3645	979/2379 (41%)	1413/2379 (59%)	0.0
gi|17568539|ref|NP_509636.1|(NM_077235)	Ig superfamily repeats (I-type) [*Caenorhabditis elegans*]	5175	857/3077 (27%)	1348/3077 (42%)	0.0
gi|17568541|ref|NP_509635.1|(NM_077234)	IG (immunoglobulin) superfamily (47 domains) [*Caenorhabditis elegans*]	5198	857/3077 (27%)	1348/3077 (42%)	0.0
gi|13872813|emb|CAC37630.1|(AJ306906)	fibulin-6 [*Homo sapiens*]	2673	552/1399 (39%)	796/1399 (56%)	0.0

A multiple sequence alignment is given in Table 9E, with the NOV9 protein being shown on line 1 in Table 9E in a ClustalW analysis, and comparing the NOV9 protein with the related protein sequences shown in Table 9D. This BLASTP data is displayed graphically in the ClustalW in Table 9E.

Table 9E. ClustalW Analysis of NOV9

1) > NOV9; SEQ ID NO:18
2) > gi|14575679|/ hemicentin [*Homo sapiens*]; SEQ ID NO:72
3) > gi|18547943|/ hemicentrin [*Homo sapiens*]; SEQ ID NO:73
4) > gi|17568539|/ Ig superfamily repeats (I-type) [*Caenorhabditis elegans*]; SEQ ID NO:74
5) > gi|17568541|/ IG (immunoglobulin) superfamily (47 domains) [*Caenorhabditis elegans*]; SEQ ID NO:75
6) > gi|13872813|/ fibulin-6 [*Homo sapiens*]; SEQ ID NO:76

```
                                4090        4100        4110        4120        4130        4140
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                       2602 VSWSKAGAQLGARGSGYRVS---------------------------------------- 2621
gi|14575679|               3929 IHWTKNGIRLLPRGDGYRIL---------------------------------------- 3948
gi|18547943|               1938 IHWTKNGIRLLPRGDGYRIL---------------------------------------- 1957
gi|17568539|               3248 VTWKRNGKEIEQFNPVLHIRNATRADEGKYSCIASNEAGTAVADFLIDVFTKPTFETHET 3307
gi|17568541|               3248 VTWKRNGKEIEQFNPVLHIRNATRADEGKYSCIASNEAGTAVADFLIDVFTKPTFETHET 3307
gi|13872813|               966  IHWTKNGIRLLPRGDGYRIL---------------------------------------- 985

                                4150        4160        4170        4180        4190        4200
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                       2621 ----------------------------------------PSGALEIGQALPIHAGRY 2639
gi|14575679|               3948 ----------------------------------------SSGAIEILATQLNHAGRY 3966
gi|18547943|               1957 ----------------------------------------SSGAIEILATQLNHAGRY 1975
gi|17568539|               3308 TFNIVEGESAKIECKIDGHPKPTISWLKGGRPFNMDNIILSPRGDTLMILKAQRFDGGLY 3367
gi|17568541|               3308 TFNIVEGESAKIECKIDGHPKPTISWLKGGRPFNMDNIILSPRGDTLMILKAQRFDGGLY 3367
gi|13872813|               985  ----------------------------------------SSGAIEILATQLNHAGRY 1003

                                4210        4220        4230        4240        4250        4260
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                       2640 TCSARNSAGVAHKHVFLTVQASPVVKP-LPSVVRAVAEEEVLLPCEASGIPRPTITWQKE 2698
gi|14575679|               3967 TCVARNAAGSAHRHVTLHVHEPPVIQP-QPSELHVILNNPILLPCEATGTPSPFITWQKE 4025
gi|18547943|               1976 TCVARNAAGSAHRHVTLHVHEPPVIQP-QPSELHVILNNPILLPCEATGTPSPFITWQKE 2034
gi|17568539|               3368 TCVATNSYGDSEQDFKVNVYTKPYIDETIDQTPKAVAGGEIILKCPVLGNPTPTVTWKRG 3427
gi|17568541|               3368 TCVATNSYGDSEQDFKVNVYTKPYIDETIDQTPKAVAGGEIILKCPVLGNPTPTVTWKRG 3427
gi|13872813|               1004 TCVARNAAGSAHRHVTLHVHEPPVIQP-QPSELHVILNNPILLPCEATGTPSPFITWQKE 1062

                                4270        4280        4290        4300        4310        4320
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                       2699 GLNVAT-GVSTQVLPGGQLRIAHASPEDAGNYLCIAKNSAGSAMGKTRLVVQVPPVIEN- 2756
gi|14575679|               4026 GINVNTSGRNHAVLPSGGLQISRAVREDAGTYMCVAQNPAGTALGKIKLNVQVPPVISPH 4085
gi|18547943|               2035 GINVNTSGRNHAVLPSGGLQISRAVREDAGTYMCVAQNPAGTALGKIKLNVQVPPVISPH 2094
gi|17568539|               3428 DDAVPN-DSRHTIVNNYDLKINSVTTEDAGQYSCIAVNEAGNLTTHYAAEVIGKPTFVRK 3486
gi|17568541|               3428 DDAVPN-DSRHTIVNNYDLKINSVTTEDAGQYSCIAVNEAGNLTTHYAAEVIGKPTFVRK 3486
gi|13872813|               1063 GINVNTSGRNHAVLPSGGLQISRAVREDAGTYMCVAQNPAGTALGKIKLNVQVPPVISPH 1122

                                4330        4340        4350        4360        4370        4380
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                       2756 ---------------------GLP---------------------------------- 2759
gi|14575679|               4086 L-KEYVIAVDKPITLSCEADGLPPPDITWHKDGRAIVESIRQRVLSSG-SLQIAFVQPGD 4143
gi|18547943|               2095 L-KEYVIAVDKPITLSCEADGLPPPDITWHKDGRAIVESIRQRVLSSG-SLQIAFVQPGD 2152
gi|17568539|               3487 GGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVYLYDRYSISPDGSHITINKAKLSD 3546
gi|17568541|               3487 GGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVYLYDRYSISPDGSHITINKAKLSD 3546
gi|13872813|               1123 L-KEYVIAVDKPITLSCEADGLPPPDITWHKDGRAIVESIRQRVLSSG-SLQITFVQPGD 1180

                                4390        4400        4410        4420        4430        4440
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                       2759 ----------------------------------DLSTTEGSHAFLPCKARGSPEPNI 2783
gi|14575679|               4144 AGHYTCMAANVAGSSSTSTKLTVHVPPRIRSTEGHYTVNENSQAILPCVADGIPTPAI 4201
gi|18547943|               2153 AGHYTCMAANVAGSSSTSTKLTVHVPPRIRSTEGHYTVNENSQAILPCVADGIPTPAI 2210
gi|17568539|               3547 GGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIGNPLAIVARTIYLECPISGIPQPDV 3606
gi|17568541|               3547 GGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIGNPLAIVARTIYLECPISGIPQPDV 3606
gi|13872813|               1181 AGHYTCMAANVAGSSSTSTKLTVHVPPRIRSTEGHYTVNENSQAILPCVADGIPTPAI 1238

                                4450        4460        4470        4480        4490        4500
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                       2784 TWDKDGQPVSGAEGKFTIQPSGE-LLVKNLEGQDAGTYTCTAENAVGRARRRVHLTILVL 2842
```

```
gi|14575679|   4202 NWKKDNVLLANLLGKYTAEPYGE-LILENVVLEDSGFYTCVANNAAGEDTHTVSLTVHVL 4260
gi|18547943|   2211 NWKKDNVLLANLLGKYTAEPYGE-LILENVVLEDSGFYTCVANNAAGEDTHTVSLTVHVL 2269
gi|17568539|   3607 IWTKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGGKASHDFSLDVLSP 3666
gi|17568541|   3607 IWTKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGGKASHDFSLDVLSP 3666
gi|13872813|   1239 NWKKDNVLLANLLGKYTAEPYGE-LILENVVLEDSGFYTCVANNAAGEDTHTVSLTVHVL 1297

                         4510        4520        4530        4540        4550        4560
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9               2843 PVFTTLPGDRSLRL-GDRLWLRCAARGSP-----TPRIGWTVNDRPVTEGVSE----QDG 2892
gi|14575679|       4261 PTFTELPGDVSLNK-GEQLRLSCKATGIP-----LPKLTWTFNNNIIPAHFDS----VNG 4310
gi|18547943|       2270 PTFTELPGDVSLNK-GEQLRLSCKATGIP-----LPKLTWTFNNNIIPAHFDS----VNG 2319
gi|17568539|       3667 PEFDIHGTQPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALDGDLTDNVDISDD 3726
gi|17568541|       3667 PEFDIHGTQPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALDGDLTDNVDISDD 3726
gi|13872813|       1298 PTFTELPGDVSLNK-GEQLRLSCKATGIP-----LPKLTWTFNNNIIPAHFDS----VNG 1347

                         4570        4580        4590        4600        4610        4620
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9               2893 GSTLQRAAVSREDSGTYVCWAENRVGRTQAVSFVHVKEAPVLQG-EAFSYLVEPVGGSIQ 2951
gi|14575679|       4311 HSELVIERVSKEDSGTYVCTAENSVGFVKATGFVYVKEPPVFKG-DYPSNWIEPLGGNAI 4369
gi|18547943|       2320 HSELVIERVSKEDSGTYVCTAENSVGFVKAIGFVYVKEPPVFKG-DYPSNWIEPLGGNAI 2378
gi|17568539|       3727 GRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPPVIDISRNDVQPQVAVNQPTI 3786
gi|17568541|       3727 GRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPPVIDISRNDVQPQVAVNQPTI 3786
gi|13872813|       1348 HSELVIERVSKEDSGTYVCTAENSVGFVKAIGFVYVKEPPVFKG-DYPSHWIEPLGGNAI 1406

                         4630        4640        4650        4660        4670        4680
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9               2952 LDCVVRGDPVPDIHWIKDGLPLRGSHLRHQLQNG-SLTIRRTEARRGLAPWRDDAGRYQC 3010
gi|14575679|       4370 LNCEVKGDPTPTIQWNRKGVDIEISHRIRQLGNG-SLAIYGTVN--------EDAGDYTC 4420
gi|18547943|       2379 LNCEVKGDPTPTIQWNRKGVDIEISHRIRQLGNG-SLAIYGTVN--------EDAGDYTC 2429
gi|17568539|       3787 MRCAVTGHPFPSIKWLKNGKEVTDDENIRIVEQGQVLQILRTDS--------DHAGKWSC 3838
gi|17568541|       3787 MRCAVTGHPFPSIKWLKNGKEVTDDENIRIVEQGQVLQILRTDS--------DHAGKWSC 3838
gi|13872813|       1407 LNCEVKGDPTPTIQWNRKGVDIEISHRIRQLGNG-SLAIYGTVN--------EDAGDYTC 1457

                         4690        4700        4710        4720        4730        4740
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9               3011 LAENEMGVAKKVVILVLQTR----MVPAEPHLKR----QLPPIPSNNEAPSLFPGVHGGH 3062
gi|14575679|       4421 VATNEAGVVERSMSLTLQSPPIITLEPVETVINAGGKIILNCQATGEPQPTITWSRQGHS 4480
gi|18547943|       2430 VATNEAGVVERSMSLTLQSPPIITLEPVETVINAGGKIILNCQATGEPQPTITWSRQGHS 2489
gi|17568539|       3839 VAENDAGVKELEMVLDVFTPPVVSVKSDNPIKALGETITLFCNASGNPYPQLKWAKGGSL 3898
gi|17568541|       3839 VAENDAGVKELEMVLDVFTPPVVSVKSDNPIKALGETITLFCNASGNPYPQLKWAKGGSL 3898
gi|13872813|       1458 VATNEAGVVERSMSLTLQSPPIITLEPVETVINAGGKIILNCQATGEPQPTITWSRQGHS 1517

                         4750        4760        4770        4780        4790        4800
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9               3063 VGNPD-------------------------------FH-SHLAEVLAVQLLAG------- 3083
gi|14575679|       4481 ISWDDRVNVLSNN--SLYIADAQKEDTSEFECVARNLMGSVLVRVPVIVQVHGGF------ 4533
gi|18547943|       2490 ISWDDRVNVLSNN--SLYIADAQKEDTSEFECVARNLMGSVLVRVPVIVQVHGGF------ 2542
gi|17568539|       3899 IFDSPDGARISLKGARLDIPHLKKTDVGDYTCQALNAAGTSEASVSVDVLVPPEINRDGI 3958
gi|17568541|       3899 IFDSPDGARISLKGARLDIPHLKKTDVGDYTCQALNAAGTSEASVSVDVLVPPEINRDGI 3958
gi|13872813|       1518 ISWDDRVNVLSNN--SLYIADAQKEDTSEFECVARNLMGSVLVRVPVIVQVHGGF------ 1570

                         4810        4820        4830        4840        4850        4860
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9               3083 ---SLLFSARAMPQASTAAISLLAPTSFAPFPDDISQGILSSSTAHQGSP-Q------G- 3132
gi|14575679|       4533 ---SQWSAWRACSVTCGKGIQKRSRLCNQPLPANGGKPCQGSDLEMRNCQNKPCPV-DGS 4589
gi|18547943|       2542 ---SQWSAWRACSVTCGKGIQKRSRLCNQPLPANGGKPCQGSDLEMRNCQNKPCPV-DGS 2598
gi|17568539|       3959 DMSPRLPAQQSLTLQCLAQGKPVPQMRWTLNGTALTHSTPGITVASDSTFIQINNVSLSD 4018
gi|17568541|       3959 DMSPRLPAQQSLTLQCLAQGKPVPQMRWTLNGTALTHSTPGITVASDSTFIQINNVSLSD 4018
gi|13872813|       1570 ---SQWSAWRACSVTCGKGIQKRSRLCNQPLPANGGKPCQGSDLEMRNCQNKPCPV-DGS 1626

                         4870        4880        4890        4900        4910        4920
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9               3132 ------WQKLLFFTAIPNKT----------------------TVMVTVEP---Q------ 3155
gi|14575679|       4590 WSEWSLWEECTRSCGRGNQTRTRTCNNPSVQHGGRPCEGNAVEIIMCNIRPCPVHGA--- 4646
gi|18547943|       2599 WSEWSLWEECTRSCGRGNQTRTRTCNNPSVQHGGRPCEGNAVEIIMCNIRPCPVHGA--- 2655
gi|17568539|       4019 KGVYTCYAENVAGSDNLMYNVDVVQAPVISNGGTKQVIEGELAVIECLVEGYPAPQVSWL 4078
gi|17568541|       4019 KGVYTCYAENVAGSDNLMYNVDVVQAPVISNGGTKQVIEGELAVIECLVEGYPAPQVSWL 4078
gi|13872813|       1627 WSEWSLWEECTRSCGRGNQTRTRTCNNPSVQHGGRPCEGNAVEIIMCNIRPCPVHGA--- 1683

                         4930        4940        4950        4960        4970        4980
```

```
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                        3155 ----------------D-------MTVRSGDDVALRCQATGEP-------------TPTI 3179
gi|14575679|                4646 ----------------WSAWQPWGTCSESCGKGTQTRARLCNNP-------------PPAF 4678
gi|18547943|                2655 ----------------WSAWQPWGTCSESCGKGTQTRARLCNNP-------------PPAF 2687
gi|17568539|                4079 RNGNRVETGVQGVRYVTDGRMLTIIEARSLDSGIYLCSATNEAGSAQQAYTLEVLVSPKI 4138
gi|17568541|                4079 RNGNRVETGVQGVRYVTDGRMLTIIEARSLDSGIYLCSATNEAGSAQQAYTLEVLVSPKI 4138
gi|13872813|                1683 ----------------WSAWQPWGTCSESCGKGTQTRARLCNNP-------------PPAF 1715

                                        4990      5000      5010      5020      5030      5040
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                        3179 ------------------------------E-WLQAGQPLRASRRLRTLPDGSLWLEN--- 3206
gi|14575679|                4679 GGSYCDGAETQMQVCNERNCPIHGKWATWASWSACSVSCGGGARQRTRGCSDPVPQYGGR 4738
gi|18547943|                2688 GGSYCDGAETQMQVCNERNCPIHGKWATWASWSACSVSCGGGARQRTRGCSDPVPQYGGR 2747
gi|17568539|                4139 ITSTPGVLTPSSGSKFSLPCAVRGYPDPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4198
gi|17568541|                4139 ITSTPGVLTPSSGSKFSLPCAVRGYPDPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4198
gi|13872813|                1716 GGSYCDGAETQMQVCNERNCPVHGKWATWASWSACSVSCGGGARQRTRGCSDPVPQYGGR 1775

                                             5050      5060      5070      5080      5090      5100
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                        3206 ----------VETGDAGTYDCVAH-N------------LLGSATARAFLVCAS------- 3236
gi|14575679|                4739 K----CEGSDVQSDFCNSDPCPTHGNWSPWSGWGTCSRTCNGGQMRRYRTCDNP------ 4788
gi|18547943|                2748 K----CEGSDVQSDFCNSDPCPTHGNWSPWSGWGTCSRTCNGGQMRRYRTCDNP------ 2797
gi|17568539|                4199 ERHLIYECTAKNDAGADTLEFPVQTIVAPKISTSGNRYINGSEGTETVIKCEIESESSEF 4258
gi|17568541|                4199 ERHLIYECTAKNDAGADTLEFPVQTIVAPKISTSGNRYINGSEGTETVIKCEIESESSEF 4258
gi|13872813|                1776 K----CEGSDVQSDFCNSDPCPTHGNWSPWSGWGTCSRTCNGGQMRRYRTCDNP------ 1825

                                             5110      5120      5130      5140      5150      5160
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                        3236 -HAIVGSRHFRDP-------------------------------------QVFCEFVVP 3257
gi|14575679|                4788 -PPSNGGRACGGPDSQIQRCNTDMCPVDGSWGSWHSWSQCSASCGGGEKTRKRLCDHPVP 4847
gi|18547943|                2797 -PPSNGGRACGGPDSQIQRCNTDMCPVDGSWGSWHSWSQCSASCGGGEKTRKRLCDHPVP 2856
gi|17568539|                4259 SWSKNGVPLLPSNNLIFSEDYKLIKILSTRLSDQGEYSCTAANKAGNATQKTNLNVGVAP 4318
gi|17568541|                4259 SWSKNGVPLLPSNNLIFSEDYKLIKILSTRLSDQGEYSCTAANKAGNATQKTNLNVGVAP 4318
gi|13872813|                1825 -PPSNGGRACGGPDSQIQRCNTDMCPVDGSWGSWHSWSQCSASCGGGEKTRKRLCDHPVP 1884

                                                  5170      5180      5190      5200      5210      5220
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                        3257 -----P---------------PHFTGEPQGSWGSMTGVINGRKFGVATLNTSVMQ------ 3292
gi|14575679|                4848 VKGGRPCPGDTTQ-VTRCNVQACPGGPQRARGSVIGNINDVEFGIAFLNATITD------ 4900
gi|18547943|                2857 VKGGRPCPGDTTQ-VTRCNVQACPGGPQRARGSVIGNINDVEFGIAFLNATITD------ 2909
gi|17568539|                4319 KIMERPRTQVVHKGDQVTLWCEASGVPQPAITWYKDNELLTNTGVDETATTKKKSVIFSS 4378
gi|17568541|                4319 KIMERPRTQVVHKGDQVTLWCEASGVPQPAITWYKDNELLTNTGVDETATTKKKSVIFSS 4378
gi|13872813|                1885 VKGGRPCPGDTTQ-VTRCNVQACPGGPQRARGSVIGNINDVEFGIAFLNATITD------ 1937

                                             5230      5240      5250      5260      5270      5280
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                        3293 EAHSGVSSIHSSIRHVPANVGPLMRVLVVTIAPIYWALARESGEALNGHSLTG------- 3345
gi|14575679|                4901 SPNSDTRIIRAKITNVPRSLGSAMRKIVSILNPIYWTTAKEIGEAVNGFTLTN------- 4953
gi|18547943|                2910 SPNSDTRIIRAKITNVPRSLGSAMRKIVSILNPIYWTTAKEIGEAVNGFTLTN------- 2962
gi|17568539|                4379 ISPSQAGVYTCKAENWVASTEEDIDLIVMIPPEVVPERMNVSTNPRQTVFLSCNATGIPE 4438
gi|17568541|                4379 ISPSQAGVYTCKAENWVASTEEDIDLIVMIPPEVVPERMNVSTNPRQTVFLSCNATGIPE 4438
gi|13872813|                1938 SPNSDTRIIRAKITNVPRSLGSAMRKIVSILNPIYWTTAKEIGEAVNGFTLTN------- 1990

                                             5290      5300      5310      5320      5330      5340
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                        3345 --GRFRQESHVEFATG---ELLTMTQVARGLDPDGLLLLDVVVNGVVPESLADADLQVQD 3400
gi|14575679|                4953 --AVFKRETQVEFATG---EILQMSHIARGLDSDGSLLLDIVVSGYVLQLQSPAEVTVKD 5008
gi|18547943|                2962 --AVFKRETQVEFATG---EILQMSHIARGLDSDGSLLLDIVVSGYVLQLQSPAEVTVKD 3017
gi|17568539|                4439 PVISWMRDSNIAIQNNEKYQILGTTLAIRNVLPDDDGFYHCIAKSDAGQKIATRKLIVNK 4498
gi|17568541|                4439 PVISWMRDSNIAIQNNEKYQILGTTLAIRNVLPDDDGFYHCIAKSDAGQKIATRKLIVNK 4498
gi|13872813|                1990 --AVFKRETQVEFATG---EILQMSHIARGLDSDGSLLLDIVVSGYVLQLQSPAEVTVKD 2045

                                                  5350      5360      5370      5380      5390      5400
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                        3400 --------FEEHYVQTGPGQLFVGSTQRFFQGGLPSFLRCNHSIQYNAARGPQPQLVQHL 3452
gi|14575679|                5008 --------YTEDYIQTGPGQLYAYSTRLFTIDGISIPYTWNHTVFYDQAQGRMPFLVETL 5060
gi|18547943|                3017 --------YTEDYIQTGPGQLYAYSTRLFTIDGISIPYTWNHTVFYDQAQGRMPFLVETL 3069
gi|17568539|                4499 PSDRPAPIWVECDEKGKPKKTEYMIDRGDTPDDNPQLLPWKDVEDSSLNGSIAYRCMPGP 4558
gi|17568541|                4499 PSDRPAPIWVECDEKGKPKKTEYMIDRGDTPDDNPQLLPWKDVEDSSLNGSIAYRCMPGP 4558
gi|13872813|                2045 --------YTEDYIQTGPGQLYAYSTRLFTIDGISIPYTWNHTVFYDQAQGRMPFLVETL 2097
```

```
                                    5410        5420        5430        5440        5450        5460
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                 3453 RASAISSAFDPEAEALRFQLATALQAEENEVGCPEG------FELDSQGAFCVD-RDECS 3505
gi|14575679|         5061 HASSVESDYNQIEETLGFKIHASISKGDRSNQCPSG------FTLDSVGPFCAD-EDECA 5113
gi|18547943|         3070 HASSVESDYNQIEETLGFKIHASISKGDRSNQCPSG------FTLDSVGPFCAD-EDECA 3122
gi|17568539|         4559 RSSRTVLLHAAPQFIVKPKNTTAAIGAIVELRCSAAGPPHPTIIWAKDGKLIEDSKFEIA 4618
gi|17568541|         4559 RSSRTVLLHAAPQFIVKPKNTTAAIGAIVELRCSAAGPPHPTIIWAKDGKLIEDSKFEIA 4618
gi|13872813|         2098 HASSVESDYNQIEETLGFKIHASISKGDRSNQCPSG------FTLDSVGPFCAD-EDECA 2150

                                    5470        5480        5490        5500        5510        5520
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                 3506 GGPSPCSHACLNAPGRFSCTCPTGFALAWDDRNCRDVDECAWDAHLCREGQRCVNLLGSY 3565
gi|14575679|         5114 AG-NPCSHSCHNAMGTYYCSCPKGLTIAADGRTCQDIDECALGRHTCHAGQDCDNTIGSY 5172
gi|18547943|         3123 AG-NPCSHSCHNAMGTYYCSCPKGLTIAADGRTCQDIDECALGRHTCHAGQDCDNTIGSY 3181
gi|17568539|         4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVS-AFINVDNNILPTPKPSSNQKNVAVITCY 4677
gi|17568541|         4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVS-AFINVDNNILPTPKPSSNQKNVAVITCY 4677
gi|13872813|         2151 AG-NPCSHSCHNAMGTYYCSCPKGLTIAADGRTCQDIDECALGRHTCHAGQDCDNTIGSY 2209

                                    5530        5540        5550        5560        5570        5580
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                 3566 RCLPDCGPGFR------------------------------------------------- 3576
gi|14575679|         5173 RCVVRCGSGFRRTSDGL----------------------SCQDINECQESSPCHQRCFN 5209
gi|18547943|         3182 RCVVRCGSGFRRTSDGL----------------------SCQDINECQESSPCHQRCFN 3218
gi|17568539|         4678 ERNQAYSRGLTWEYNGVPMPKNLAGIHFMNNGSLVILDTSSLKEGDLELYTCKVRNRRRH 4737
gi|17568541|         4678 ERNQAYSRGLTWEYNGVPMPKNLAGIHFMNNGSLVILDTSSLKEGDLELYTCKVRNRRRH 4737
gi|13872813|         2210 RCVVRCGSGFRRTSDGL----------------------SCQDINECQESSPCHQRCFN 2246

                                    5590        5600        5610        5620        5630        5640
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                 3576 ------------------------------------------------------------ 3576
gi|14575679|         5210 AIGSFHCGCEPGYQLKGRKCMDVN-------ECRQNVCRPDQHCKNTRGGYKCIDLCPNG 5262
gi|18547943|         3219 AIGSFHCGCEPGYQLKGRKCMDVN-------ECRQNVCRPDQHCKNTRGGYKCIDLCPNG 3271
gi|17568539|         4738 SIPHLTSAFEGVPEVKTIDKVEVNNGDSVVLDCEVTSDPLTTHVVWTKNDQKMLDDDAIY 4797
gi|17568541|         4738 SIPHLTSAFEGVPEVKTIDKVEVNNGDSVVLDCEVTSDPLTTHVVWTKNDQKMLDDDAIY 4797
gi|13872813|         2247 AIGSFHCGCEPGYQLKGRKCMDVN-------ECRQNVCRPDQHCKNTRGGYKCIDLCPNG 2299

                                    5650        5660        5670        5680        5690        5700
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                 3577 VADGAG--CEDVDECLEGLDDCHYNQLCENTPGGHRCSCPRG------------------- 3616
gi|14575679|         5263 MTKAENGTCIDIDECKDGTHQCRYNQICENTRSSYRCVCPRG------------------- 5304
gi|18547943|         3272 MTKAENGTCIDIDECKDGTHQCRYNQICENTRGSYRCVCPRG------------------- 3313
gi|17568539|         4798 VLPNNSLVLLNVEKYDEGVYKCVASNSIGKAFDDTQLNVYG------------------- 4838
gi|17568541|         4798 VLPNNSLVLLNVEKYDEGVYKCVASNSIGKAFDDTQLNVYEGDFLPLTGFEGSGINIDDS 4857
gi|13872813|         2300 MTKAENGTCIDIDECKDGTHQCRYNQICENTRGSYRCVCPRG------------------- 2341

                                    5710        5720        5730        5740        5750        5760
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                 3616 ----------YRMQGPSLPCLDVNECLQLPKACAYQCHNLQGSYRCLCPPGQTLLRDGK 3665
gi|14575679|         5304 ----------YRSQGVGRPCMDIDECEQVPKPCAHQCSNTPGSFKCICPPGQHLLGDGK 5353
gi|18547943|         3313 ----------YRSQGVGRPCMDINECEQVPKPCAHQCSNTPGSFKCICPPGQHLLGDGK 3362
gi|17568539|         4838 ----GSSRREAYKKENEDASTTTITTTSPTTTTTETPLTTTIIPALITLPAKQYPTDDYH 4894
gi|17568541|         4858 SNAGGSSRREAYKKENEDASTTTITTTSPTTTTTETPLTTTIIPALITLPAKQYPTDDYH 4917
gi|13872813|         2341 ----------YRSQGVGRPCMDINECEQVPKPCAHQCSNTPGSFKCICPPGQHLLGDGK 2390

                                    5770        5780        5790        5800        5810        5820
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                 3666 ACTSLERNGQNVTTVSHR-----GPLLPWLRPWASIPGTS--YHAWVSLRPGPMALSSVG 3718
gi|14575679|         5354 SCAGLERLPNYGTQYSSYNLARFSPVRNNYQPQQHYRQYSHLYSSYSEYRNSRTSLSRTR 5413
gi|18547943|         3363 SCAGLERLPNYGTQYSSYNLARFSPVRNNYQPQQHYRQYSHLYSSYSEYRNSRTSLSRTR 3422
gi|17568539|         4895 EGSANDDGFGPTTQDSLFEFN--PPLHPEISVVNTDCAGTINENGDCVDKDGKTHNLKIL 4952
gi|17568541|         4918 EGSANDDGFGPTTQDSLFEFN--PPLHPEISVVNTDCAGTINENGDCVDKDGKTHNLKIL 4975
gi|13872813|         2391 SCAGLERLPNYGTQYSSYNLARFSPVRNNYQPQQHYRQYSHLYSSYSEYRNSRTSLSRTR 2450

                                    5830        5840        5850        5860        5870        5880
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9                 3718 ---RAWCPPGFIR--QNGVCTDLDECRVRNLCQHACRNTEGSYQCLCPAGYRLLPSGKNC 3773
gi|14575679|         5414 RTIRKTCPEGSEA--SHDTCVDIDECENTDACQHECKNTFGSYQCICPPGYQLTHNGKTC 5471
gi|18547943|         3423 RTIRKTCPEGSEA--SHDTCVDIDECENTDACQHECKNTFGSYQCICPPGYQLTHNGKTC 3480
gi|17568539|         4953 TG-ENHCPEGFAMNPHTRICEDLDECAFYQPCDFECINYDGGFQCNCPLGYELAEEG--C 5009
```

```
gi|17568541|   4976 TG-ENHCPEGFAMNPHTRICEDLDECAFYQPCDFECINYDGGFQCNCPLGYELAEEG--C 5032
gi|13872813|   2451 RTIRKTCPEGSEA--SHDTCVDIDECENTDACQHECKNTFGSYQCICPPGYQLTHNGKTC 2508

                                5890      5900      5910      5920      5930      5940
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9               3774 QDINECEEESIECGPGQMCFNTRGSYQCVDTPCPATYRQGPSPGTCFRRCS---QDCGTG 3830
gi|14575679|       5472 QDIDECLEQNVHCGPNRMCFNMRGSYQCIDTPCPPNYQRDPVSGFCLKNCPPNDLECALS 5531
gi|18547943|       3481 QDIDECLEQNVHCGPNRMCFNMRGSYQCIDTPCPPNYQRDPVSGFCLKNCPPNDLECALS 3540
gi|17568539|       5010 RDVNECES--VRCEDGKACFNQLGGYECIDDPCPANYSL--VDD--R-YCEP---ECENC 5059
gi|17568541|       5033 RDVNECES--VRCEDGKACFNQLGGYECIDDPCPANYSL--VDD--R-YCEP---ECENC 5082
gi|13872813|       2509 QDIDECLEQNVHCGPNRMCFNMRGSYQCIDTPCPPNYQRDPVSGFCLKNCPPNDLECALS 2568

                                5950      5960      5970      5980      5990      6000
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV9               3831 GPSTLQYRLLPLPLGVRAHHDVARLTAFSEVGVPANRTELSMLEP--------DPRSPFA 3882
gi|14575679|       5531 -PYALEYKLVSLPFGIATNQDLIRLVAYTQDGVMHPRTTFLMVDE--------EQTVPFA 5582
gi|18547943|       3540 -PYALEYKLVSLPFGIATNQDLIRLVAYTQDGVMHPRTTFLMVDE--------EQTVPFA 3591
gi|17568539|       5060 TSTPIQVHMLAIPSGLPISH-IATLTAYDKSGRVLNDTTYAISDTGAPLARGRMTSGPFT 5118
gi|17568541|       5083 TSTPIQVHMLAIPSGLPISH-IATLTAYDKSGRVLNDTTYAISDTGAPLARGRMTSGPFT 5141
gi|13872813|       2568 -PYALEYKLVSLPFGIATNQDLIRLVAYTQDGVMHPRTTFLMVDE--------EQTVPFA 2619

                                6010      6020      6030      6040      6050
                     ....|....|....|....|....|....|....|....|....|....|....|..
NOV9               3883 LRPLRAGLGAVYTRRALTRAGLYRLTVRAAAP--------RHQSVFVLLIAVSPYPY 3931
gi|14575679|       5583 LRDE-NLKGVVYTTRPLREAETYRMRVRASSYSAN--GTIEYQTTFIVYIAVSAYPY 5636
gi|18547943|       3592 LRDE-NLKGVVYTTRPLREAETYRMRVRASSYSAN--GTIEYQTTFIVYIAVSAYPY 3645
gi|17568539|       5119 IKAVKRGHAQVWTNRVLAAGDHHKVRVRAHSDHATNELHAPKETNFLVLINVGQYPF 5175
gi|17568541|       5142 IKAVKRGHAQVWTNRVLAAGDHHKVRVRAHSDHATNELHAPKETNFLVLINVGQYPF 5198
gi|13872813|       2620 LRDE-NLKGVVYTTRPLREAETYRMRVRASSYSAN--
GTIEYQTTFIVYIAVSAYPY 2673
```

The NOV9 Clustal W alignment shown in Table 9E was modified to begin at amino residue 4080. The data in Table 9E includes all of the regions overlapping with the NOV9 protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 9F lists the domain description from DOMAIN analysis results against NOV9.

Table 9F Domain Analysis of NOV9

Model	Region of Homology	Score (bits)	E value
ig	99-157	46.50	6.1e-10
ig	192-251	38.00	2.1e-07
ig	286-344	41.90	1.4e-08
ig	380-438	32.10	1.3e-05
ig	473-531	45.90	8.9e-10
ig	565-615	35.50	1.2e-06
ig	648-706	43.80	4e-09
ig	740-800	39.10	1e-07
ig	833-891	27.10	0.00041
ig	981-1039	42.90	7.1e-09
ig	1075-1133	28.90	0.00012
ig	1168-1226	30.80	3.2e-05
ig	1264-1322	39.60	7.2e-08
ig	1362-1420	43.40	5.2e-09
ig	1473-1531	29.40	8.2e-05

ig	1568-1626	35.70	1e-06
ig	1654-1712	45.20	1.5e-09
ig	1749-1807	43.80	3.8e-09
ig	1841-1899	43.00	6.8e-09
ig	1934-1994	41.40	2e-08
ig	2030-2088	49.90	5.8e-11
ig	2123-2177	49.60	6.8e-11
ig	2212-2268	37.50	3e-07
ig	2303-2361	28.60	0.00014
ig	2394-2459	28.20	0.00019
ig	2492-2552	32.90	7.2e-06
ig	2585-2643	22.80	0.0081
ig	2676-2733	37.50	3.1e-07
ig	2766-2824	43.10	6.3e-09
ig	2857-2913	44.10	3e-09
ig	2947-3012	18.00	0.22
ig	3162-3219	34.70	2.1e-06
EGF	3504-3539	38.30	1.8e-07
EGF	3589-3626	13.80	1.3
EGF	3632-3667	33.30	5.4e-06
EGF	3739-3773	38.40	1.6e-07

Consistent with other known members of the Hemicentin Precursor-like family of proteins, NOV9 has, for example, thirty-three immunoglobulin (ig) signature sequences and four epidermal growth factor (EGF) signature sequences, as well as homology to other members of the Hemicentin Precursor-like Protein Family. NOV9 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV9 nucleic acids and polypeptides can be used to identify proteins that are members of the Hemicentin Precursor-like Protein Family. The NOV9 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV9 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, *e.g.,* cellular activation, cellular differentiation, and signal transduction. These molecules can be used to treat, *e.g.*, Cardiovascular diseases, Hyperparathyroidism, Hypoparathyroidism, Lymphedema, Allergies as well as other diseases, disorders and conditions..

In addition, various NOV9 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV9 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the Hemicentin Precursor-like Protein Family.

Hemicentrin is an extracellular matrix protein with a modular sturcture. Like NOV9, the hemicentrin structure includes many immunoglobulin domains flanked by EGF domains. The protein is likely involved in cellular differentiation of epithelial tissue.

The NOV9 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of cardiac, immune and endocrine physiology. As such, the NOV9 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat cardiovascular, immune, and endocrine disorders, *e.g.,* Cardiovascular diseases, Hyperparathyroidism, Hypoparathyroidism, Lymphedema, Allergies as well as other diseases, disorders and conditions.

The NOV9 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV9 nucleic acid is expressed in Adipose, Thyroid, Colon, Lymph node, Bone, Myometrium, Prostate, Testis, Aorta, Vein.

Additional utilities for NOV9 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV10

A NOV10 polypeptide has been identified as a Selectin-like protein. The novel NOV10 nucleic acid sequences maps to the chromosome 9. Two alternative novel NOV10, NOV10a and NOV10b, nucleic acids and encoded polypeptides are provided.

NOV10a

A NOV10 variant is NOV10a (alternatively referred to herein as CG94661-01), which includes the 1268 nucleotide sequence (SEQ ID NO:19) shown in Table 10A. A NOV10a ORF begins with a ATG initiation codon at nucleotides 145-147 and ends with a TGA codon at nucleotides 871-873. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 10A, and the start and stop codons are in bold letters.

Table 10A. NOV10a Nucleotide Sequence (SEQ ID NO:19)

GCGGCCGCCACCCTCCGTGGCAAGGCGAGGCCCCGGGGGCGGGCCGGGGTCACCACGCCTGTCCCAG
GGAACCGCACAGACGGTACTCACCCTTCTTGCGATGATGTGAGATGATAAAATGCCTACATGATGAG
ATGAAGTGAGATGAAAAACATAGGCCTTGTG**ATG**GAATGGGAAATTCCAGAGATAATTTGCACGTGC
GCTAAGCTGCGGCTACCCCCGCAAGCAACCTTCCAAGTCCTTCGTGGCAATGGTGCTTCCGTGGGGA
CCGTGCTCATGTTCCGCTGCCCCTCCAACCACCAGATGGTGGGGTCTGGGCTCCTCACCTGCACCTG
GAAGGGGAGCATCGCTGAGTGGTCTTCAGGGTCCCCAGTGTGCAAACTGGTGCCACCACACGAGACC
TTTGGCTTCAAGGTGGCCGTGATCGCCTCCATTGTGAGCTGTGCCATCATCCTGCTCATGTCCATGG
CCTTCCTCACCTGCTGCCTCCTCAAGTGCGTGAAGAAGAGCAAGCGGCGGCGCTCCAACAGGTCAGC
CCAGCTGTGGTCCCAGCTGAAAGATGAGGACTTGGAGACGGTGCAGGCCGCATACCTTGGCCTCAAG

CACTTCAACAAACCCGTGAGCGGGCCCAGCCAGGCGCACGACAACCACAGCTTCACCACAGACCATG
GTGAGAGCACCAGCAAGCTGGCCAGTGTGACCCGCAGCGTGGACAAGGACCCTGGGATCCCCAGAGC
TCTAAGCCTCAGTGGCTCCTCCAGCTCACCCCAAGCCCAGGTGATGGTGCACATGGCAAACCCCAGA
CAGCCCCTGCCTGCCTCTGGGCTGGCCACAGGAATGCCACAACAGCCCGCAGCATATGCCCTAGGG**T**
GACCACGCAGTGAGGCTGGTGCCCATGCTCCACACTGGGAGGCCAGGCTGACCCCACCAGCCAGTCA
GCTACAACTCCACATCAACTCCACATGCGCCCAGCTCGAGACTGATGAGTGGAATCAGCTTCCAGGT
GTAGGGACCCCTTGAGGGGCCGAGCTGACATCCAAGGCTGAGGACCCCAGTGGGGAGTGTTCTGTTC
CGGCATATCCTGGCCGTAACGATTTTTATAGTTATGGACTACTTGAAACCACTACTGAGGGTAATTT
ACTAGCTGTGGCCTCCCACTAACTAGCATTCCTTTAAAGAGACTGGGAAATGTTTTAAGCAAATCTA
GTTTTGTATAATAAAATAAGAAAATAGCAATAAACTTCTTTTCAGCAACTACAAAAAAAAAA

The NOV10a polypeptide (SEQ ID NO:20) encoded by SEQ ID NO:19 is 242 amino acid residues in length and is presented using the one-letter amino acid code in Table 10B. The Psort profile for the NOV10a predicts that this peptide is likely to be localized at the plasma membrane with a certainty of 0.7000.

Table 10B. NOV10a protein sequence (SEQ ID NO:20)

MKNIGLVMEWEIPEIICTCAKLRLPPQATFQVLRGNGASVGTVLMFRCPSNHQMVGSGLLTCTWKGS
IAEWSSGSPVCKLVPPHETFGFKVAVIASIVSCAIILLMSMAFLTCCLLKCVKKSKRRRSNRSAQLW
SQLKDEDLETVQAAYLGLKHFNKPVSGPSQAHDNHSFTTDHGESTSKLASVTRSVDKDPGIPRALSL
SGSSSSPQAQVMVHMANPRQPLPASGLATGMPQQPAAYALG

NOV10b

Alternatively, a NOV10 variant is the novel NOV10b (alternatively referred to herein as CG94661-02), which includes the 887 nucleotide sequence (SEQ ID NO:21) shown in Table 10C. NOV10b was created by polymerase chain reaction (PCR) using the primers detailed in Example 1, Table 17. Primers were designed based on *in silico* predictions of the full length or some portion (one or more exons) of the cDNA/protein sequence of the invention. The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clone 143260::COR100348691_extn.698976.C20.

The NOV10b ORF begins with a Kozak consensus ATG initiation codon at nucleotides 72-74 and ends with a TGA codon at nucleotides 1958-1960. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 10C, and the start and stop codons are in bold letters.

Table 10C. NOV10b Nucleotide Sequence (SEQ ID NO:21)

GCACAGACGGTACTCACCCTTCTTGCGATGATGTGAGATGATAAAATGCCTACATGATGAGATGAAG
TGAG**ATG**AAAAACATAGGCCTTGTGATGGAATGGGAAATTCCAGAGATAATTTGCATGTGCGCTAAG
CTGCGGCTACCCCCGCAAGCAACCTTCCAAGTCCTTCGTGGCAATGGTGCTTCCGTGGGGACCGTGC
TCATGTTCCGCTGCCCCCCCAACCACCAGATGGTGGGGTCTGGGCTCCTCACCTGCACCTGGAAGGG
GAGCATCGCTGAGTGGTCTTCAGGGTCCCCAGTGTGCAAACTGGTGCCACCACACGAGACCTTTGGC
TTCAAGGTGGCCGTGATCGCCTCCATTGTGAGCTGTGCCATCATCCTGCTCATGTCCATGGCCTTCC
TCACCTGCTGCCTCCTCAAGTGCGTGAAGAAGAGCAAGCGGCGGCGCTCCAACAGGTCAGCCCAGCT
GTGGTCCCAGCTGAAAGATGAGGACTTGGAGACGGTGCAGGCCGCATACCTTGGCCTCAAGCACTTC

AACAAACCCGTGAGCGGGCCCAGCCAGGCGCACGACAACCACAGCTTCACCACAGACCATGGTGAGA
GCACCAGCAAGCTGGCCAGTGTGACCCGCAGCGTGGACAAGGACCCTGGGATCCCCAGAGCTCTAAG
CCTCAGTGGCTCCTCCAGCTCACCCCAAGCCCAGGTGATGGTGCACATGGCAAACCCCAGACAGCCC
CTGCCTGCCTCTGGGCTGGCCACAGGAATGCCACAACAGCCCGCAGCATATGCCCTAGGG**TGA**CCAC
GCAGTGAGGCTGGTGCCCATGCTCCACACTGGGAGGCCAGGCTGACCCCACCAGCCAGTCAGCTACA
ACTCCACATCAACTCC

Variant sequences of NOV10b are included in Example 3, Table 22. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA.

The NOV10b protein (SEQ ID NO:22) encoded by SEQ ID NO:21 is 242 amino acid residues in length and is presented using the one-letter code in Table 10D. The Psort profile for NOV10b predicts that this sequence is likely to be localized at the plasma membrane with a certainty of 0.7000.

Table 10D. NOV10b protein sequence (SEQ ID NO:22)

MKNIGLVMEWEIPEIICMCAKLRLPPQATFQVLRGNGASVGTVLMFRCPPNHQMVGSGLLTCTWKGS
IAEWSSGSPVCKLVPPHETFGFKVAVIASIVSCAIILLMSMAFLTCCLLKCVKKSKRRRSNRSAQLW
SQLKDEDLETVQAAYLGLKHFNKPVSGPSQAHDNHSFTTDHGESTSKLASVTRSVDKDPGIPRALSL
SGSSSSPQAQVMVHMANPRQPLPASGLATGMPQQPAAYALG

NOV10 Clones

Unless specifically addressed as NOV10a or NOV10b, any reference to NOV10 is assumed to encompass all variants. NOV10a differs from NOV10b at amino acid position 18 (T>M) and amino acid position 50 (S>P) as shown in Tables 10B and 10D.

A search against the Patp database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 10E.

Table 10E. Patp results for NOV10

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAM93054 Human digestive system antigen	+1	210	7.2e-17
>patp:AAR05494 Endothelial leukocyte adhesion molecule-1	+1	113	0.0016
>patp:AAR08116 Endothelial leucocyte adhesion molecule-1	+1	113	0.0016
>patp:AAW18839 E-selectin	+1	113	0.0016
>patp:AAW46733 Endothelial leukocyte adhesion molecule-1	+1	113	0.0016

In a BLAST search of public sequence databases, it was found, for example, that the NOV10a nucleic acid sequence of this invention has 438 of 447 bases (97%) identical to a gb:GENBANK-ID:HSM802384|acc:AL137623.1 mRNA from *Homo sapiens* (cDNA DKFZp434J1812 (from clone DKFZp434J1812)). The full amino acid sequence of the protein of the invention was found to have 110 of 139 amino acid residues (79%) identical to, and 123

of 139 amino acid residues (88%) similar to, the 269 amino acid residue ptnr:SPTREMBL-ACC:Q9D176 protein from *Mus musculus* (1700017I11RIK PROTEIN).

Similarly, it was found, for example, the NOV 10b nucleic acid sequence of this invention has 438 of 447 bases (97%) identical to a gb:GENBANK-ID:HSM802384|acc:AL137623.1 mRNA from *Homo sapiens* (cDNA DKFZp434J1812 (from clone DKFZp434J1812)). The full amino acid sequence of the protein of the invention was found to have 108 of 138 amino acid residues (78%) identical to, and 121 of 138 amino acid residues (87%) similar to, the 269 amino acid residue ptnr:SPTREMBL-ACC:Q9D176 protein from *Mus musculus* (1700017I11RIK PROTEIN).

Additional BLAST results are shown in Table 10F.

Table 10F. BLAST results for NOV10

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
>gi|15779059|gb|AAH14601.1|AAH14601(BC014601)	Similar to RIKEN cDNA 1700017I11 gene [*Homo sapiens*]	255	192/225 (85%)	192/225 (85%)	e-101
>gi|12834785|dbj|BAB23043.1|(AK003860)	Sushi domain (SCR repeat) containing protein~data source:Pfam, source key:PF00084, evidence:ISS~putative [*Mus musculus*]	269	130/246 (52%)	142/246 (56%)	8e-56
>gi|12850544|dbj|BAB28764.1|(AK013276)	Sushi domain (SCR repeat) containing protein~data source:Pfam, source key:PF00084, evidence:ISS~putative [*Mus musculus*]	170	71/102 (69%)	77/102 (74%)	6e-35
>gi|12838976|dbj|BAB24394.1|(AK006068)	Sushi domain (SCR repeat) containing protein~data source:Pfam, source key:PF00084, evidence:ISS~putative [*Mus musculus*]	149	55/73 (75%)	61/73 (83%)	2e-26

>gi|7494498|pir|| T18524	scavenger receptor cysteine-rich protein homolog srcrm2 - *Geodia cydonium*	2043	30/87 (34%)	42/87 (47%)	2e-04

A multiple sequence alignment is given in Table 10G, with the NOV10 protein of the invention being shown on line 1, in a ClustalW analysis comparing NOV10 with related protein sequences disclosed in Table 10F.

Table 10G. Information for the ClustalW proteins:

1) > NOV10a; SEQ ID NO:20
2) > NOV10b; SEQ ID NO:22
3) > gi|1577905/ similar to RIKEN cDNA 1700017I11 gene [*Homo sapiens*]; SEQ ID NO:77
4) > gi|1283478/ Sushi domain (SCR repeat) containing protein-data source: Pfam, source key: PF00084, evidence:ISS-putative [*Mus musculus*]; SEQ ID NO:78
5) > gi|1285054/ Sushi domain (SCR repeat) containing protein-data source: Pfam, source key: PF00084, evidence:ISS-putative [*Mus musculus*]; SEQ ID NO:79
6) > gi|1283897/ Sushi domain (SCR repeat) containing protein-data source: Pfam, source key: PF00084: ISS-putative [*Mus musculus*]; SEQ ID NO:80
7) > gi|7494498/ scavenger receptor cysteine-rich protein homolog srcrm2-*Geodia cydonium*; SEQ ID NO:81

```
                      1610        1620        1630        1640        1650
                      ....|....|....|....|....|....|....|....|....|....|
NOV10a          ---MKNIGLVMEWEIP----------------------------EIICTCAKLRL
NOV10b          ---MKNIGLVMEWEIP----------------------------EIICMCAKLRL
gi|1577905      ---MRWAAATLRGKARPRGRA----------GVTTPAPGNRTGTCAKLRL
gi|1283478      ---MRRTSATLRGRARPRWRA----------GNTTPVPVNQTGTCAQLHP
gi|1285054      ---MRRTSATLRGRARPRWRA----------GNTTPVPVNQTGTCAQLHP
gi|1283897      ---MRRTSATLRGRARPRWRA----------GNTTPVPVNQTGTCAQLHP
gi|7494498      GRYTQDTGVVIACEDGYQPTEGAADVLCTEDGTWSPRTPACSVSCDPLQV

                      1660        1670        1680        1690        1700
                      ....|....|....|....|....|....|....|....|....|....|
NOV10a          PPQATFQ-------------------------------------------
NOV10b          PPQATFQ-------------------------------------------
gi|1577905      PPQATFQ-------------------------------------------
gi|1283478      PPQGTLQ-------------------------------------------
gi|1285054      PPQGTLQ-------------------------------------------
gi|1283897      PPQGTLQ-------------------------------------------
gi|7494498      PAHASLSSPDTIPGLTVSLVCDKGYTYDASSDGDFSVVCGLDGEWNSTLG

                      1710        1720        1730        1740        1750
                      ....|....|....|....|....|....|....|....|....|....|
NOV10a          ----------------------VLRGNGASVGTVLMFRCPSNHQMVGSGLL
NOV10b          ----------------------VLRGNGASVGTVLMFRCPPNHQMVGSGLL
gi|1577905      ----------------------VLRGNGASVGTVLMFRCPSNHQMVGSGLL
gi|1283478      ----------------------VVRGDGTSLGTVLIFHCPSGHQMVGSGLL
gi|1285054      ----------------------VVRGDGTSLGTVLIFHCPSGHQMVGSGLL
gi|1283897      ----------------------VVRGDGTSLGTVLIFHCPSGHQMVGSGLL
gi|7494498      TCKLVLCPAYSFNVTTNLRVSLTQGSGTTVGTTLSFSCPSDFHLAGDSVI

                      1760        1770        1780        1790        1800
                      ....|....|....|....|....|....|....|....|....|....|
NOV10a          TCTWKGSIAEWSSGSPVCKLVPPHETFGFKVAVIASIVSCAIILLMSMAF
NOV10b          TCTWKGSIAEWSSGSPVCKLVPPHETFGFKVAVIASIVSCAIILLMSMAF
gi|1577905      TCTWKGSIAEWSSGSPVCKLVPPHETFGFKVAVIASIVSCAIILLMSMAF
gi|1283478      TCAWNGSTVDWSSGSPVCKAVPPHETFGFKVAVIASIVSCAIILLMSMAF
```

```
gi|1285054   TCAWNGSTVDWSSGSPVCKAVPPHETFGFKVAVIASIVSCAIILLMSMAF
gi|1283897   TCAWNGSTVDWSSGSPVCKAVPPHETFGFKVAVIASIVS--VPSSCSCPW
gi|7494498   SCMDSG---TWNGTVPHCSREPPTKCPTLTISDHVTASSSETTINTVVSF

                      1810      1820      1830      1840      1850
                     ....|....|....|....|....|....|....|....|....|
NOV10a       LTCC---------LLKCVKKSKRRRSNRSAQLWSQLKDEDL-ETVQAAYLG
NOV10b       LTCC---------LLKCVKKSKRRRSNRSAQLWSQLKDEDL-ETVQAAYLG
gi|1577905   LTCC---------LLKCVKKSKRRRSNRSAQLWSQLKDEDL-ETVQAAYLG
gi|1283478   LTCC---------LLKCVQKNERRRADRTAQLWYQLRGEDL-ETVQAAYLG
gi|1285054   LTCC---------LLKCVQKNERRRADR--------YSG--L-MSFQKLCLA
gi|1283897   PSS------------PAASSSVCRKMSG----------G----------VLT
gi|7494498   TCDNGYFLKGDKILECLSTGVWNGTAPTCSLPNSCPSLKLSDHVTASSTD

                      1860      1870      1880      1890      1900
                     ....|....|....|....|....|....|....|....|....|
NOV10a       LKHFN--------------------KPVSGPSQAHDNHSFTTDHGESTSKL
NOV10b       LKHFN--------------------KPVSGPSQAHDNHSFTTDHGESTSKL
gi|1577905   LKHFN--------------------KPVSGPSQAHDNHSFTTDHGESTSKL
gi|1283478   LKGHNHNNSSSVGGGNGGPSGGGGKPGIQHSQAHDNHSFTTDPGD-IREQ
gi|1285054   VR-------------------------LLITSCQVPWG--------------
gi|1283897   GR-----------------------------HNSGTS----------------
gi|7494498   TRINAVVTFTCD--DDRYTLNGNKIIACQSTGVWNGTAPTCKEIPTCPEL

                      1910      1920      1930      1940      1950
                     ....|....|....|....|....|....|....|....|....|
NOV10a       ASVTRSVDKDPGIPR-----------------------------ALSLSGS
NOV10b       ASVTRSVDKDPGIPR-----------------------------ALSLSGS
gi|1577905   ASVTRSVDKDPGIPR-----------------------------ALSLSGS
gi|1283478   AGVTHSVDKDPWTFR-----------------------------MGTPGPGGC
gi|1285054   -----SV-------------------------------------------
gi|1283897   -------------------------------------------------
gi|7494498   TPSSHVIPSTPDNSVGAEVSFQCEDGYTLQGEKKITCLPTQKWSANPPSC

                      1960      1970      1980      1990      2000
                     ....|....|....|....|....|....|....|....|....|
NOV10a       SSSPQAQVMVHMANPRQPLP---ASGLATGMPQQPAAYALG---------
NOV10b       SSSPQAQVMVHMANPRQPLP---ASGLATGMPQQPAAYALG---------
gi|1577905   SSSPQAQVMVHMANPRQPLP---ASGLATGMPQQPAAYALG---------
gi|1283478   SSSPGTYVMVHALN---------SAGLAPGNPGRPKVYLPG---------
gi|1285054   -------------------------------------------------
gi|1283897   -------------------------------------------------
gi|7494498   GSNPTSQPLSNNDNGSGTKVGPIVGGLIGGILVIVLIIVATAILFWKLSS
```

The NOV10 Clustal W alignment shown in Table 10F was modified to begin at amino residue 1600 and end at amino acid residue 2000. The data in Table 10F includes all of the regions overlapping with the NOV10 protein sequences.

The NOV10 Clustal W alignment shown in Table 10G was modified to begin at amino residue 1601. The data in Table 10G includes all of the regions overlapping with the NOV10 protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 10H lists the domain description from DOMAIN analysis results against NOV10.

Table 10H Domain Analysis of NOV10			
Model	Region of Homology	Score (bits)	E value
Sushi domain (SCR repeat)	19-78	15.8	0.0075

Consistent with other known members of the Selectin-like family of proteins, NOV10 has, for example, a Sushi domain (SCR repeat) signature sequences and homology to other members of the Selectin-like Protein Family. NOV10 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV10 nucleic acids and polypeptides can be used to identify proteins that are members of the Selectin-like Protein Family. The NOV10 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV10 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, *e.g.*, cellular adhesion and signal transduction. These molecules can be used to treat, *e.g.*, Cardiovascular diseases, 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, Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation, Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalceimia, Lesch-Nyhan syndrome, Systemic lupus erythematosus, Autoimmune disease, Asthma, Emphysema, Scleroderma, allergy as well as other diseases, disorders and conditions.

In addition, various NOV10 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV10 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the Selectin-like Protein Family.

The NOV10 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of cardiac and immune physiology. As such, the NOV10 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat cardiovascular and immune disorders, *e.g.*, Cardiovascular diseases, 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, Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation, Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalceimia, Lesch-Nyhan syndrome, Systemic lupus erythematosus , Autoimmune disease, Asthma, Emphysema, Scleroderma, allergy as well as other diseases, disorders and conditions.

The NOV10 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV10 nucleic acid is expressed in Heart, Thyroid, Parotid Salivary glands, Liver, Colon, Ascending Colon, Bone Marrow, Peripheral Blood, Lymphoid tissue, Spleen, Lymph node, Tonsils, Thymus, Cerebellum, Spinal Chord, Cervix, Mammary gland/Breast, Ovary, Placenta, Uterus, Oviduct/Uterine Tube/Fallopian tube, Vulva, Prostate, Testis, Lung, Kidney, Kidney Cortex, Retina, Skin.

Additional utilities for NOV10 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV11

A NOV11 polypeptide has been identified as a Nuclear Protein-like protein (also referred to as CG94325-01). The disclosed novel NOV11 nucleic acid (SEQ ID NO:23) of 8670 nucleotides is shown in Table 11A. The novel NOV11 nucleic acid sequences maps to the chromosome 15.

An ORF begins with an ATG initiation codon at nucleotides 204-206 and ends with a TAA codon at nucleotides 7152-7154. A putative untranslated region and/or downstream from the termination codon is underlined in Table 11A, and the start and stop codons are in bold letters.

Table 11A. NOV11 Nucleotide Sequence (SEQ ID NO:23)

ACGCGTAGAGCCGCTTTGCGCGTGCGCATCACCTAGGCGGTTAGATTTGAATACTTCACTGAGGCGA
GCCGGGCGTTGTGAGCGGACTGCTAGAGGCGGCTGTCTGTTTCCGCTCTAAGGAAACTCAGAGCGTG
TGGACCCCAAACAAGTCTGCGCAAAATTTGTCGAGGAGGTTTGCCGCGGCAGAAAAGTTTTCTTCAA
AAATGG**ATG**GGGTGTCTTCAGAGGCTAATGAAGAAAATGACAATATAGAGAGACCTGTTAGAAGACG
GCATTCTTCAATATTGAAACCCCCAAGGAGTCCTCTTCAGGACCTCAGAGGTGGGAATGAAAGAGTT
CAGGAATCCAATGCTTTGAGAAATAAGAAAAACTCTCGTCGAGTCAGCTTTGCAGATACTATAAAGG
TATTCCAGACGGAGTCTCATATGAAAATAGTGAGAAAGTCAGAAATGGAAGAAACAGAAACAGGAGA
AAATCTTCTTTTGATACAGAATAAGAAATTAGAAGATAATTACTGTGAAATTACTGGGATGAACACA
TTGCTTTCTGCTCCCATTCATACCCAGATGCAACAGAAGGAGTTTTCAATTATAGAACATACCCGTG
AAAGGAAACATGCAAATGACCAGACAGTCATTTTTTCAGATGAAAACCAGATGGACCTGACATCAAG

TCACACTGTAATGATTACCAAAGGCCTTTTAGATAATCCCATAAGTGAAAAGTCCACCAAGATAGAT
ACCACATCATTTCTAGCTAATTTAAAGCTTCACACCGAGGACTCAAGAATGAAAAAAGAAGTAAATT
TTTCCGTGGATCAAAACACTTCTTCAGAAAATAAAATAGATTTCAATGACTTCATAAAAAGATTGAA
AACAGGAAAATGTAGTGCTTTTCCTGATGTGCCTGATAAAGAAAATTTTGAGATACCTATTTATTCC
AAGGAACCGAACAGTGCCTCTTCTACACATCAAATGCATGTATCTCTTAAGGAAGATGAAAATAACA
GTAATATTACTAGGCTCTTTAGAGAAAAAGATGATGGGATGAATTTCACCCAGTGTCATACAGCCAA
TATTCAGACATTGATTCCCACATCCAGTGAGACCAACTCACGGGAATCTAAAGGTAATGATATTACA
ATTTATGGCAATGACTTTATGGACTTGACATTTAACCACACTTTGCAGATCTTACCTGCAACAGGTA
ATTTTTCTGAAATAGAAAATCAAACTCAGAATGCCATGGATGTAACAACAGGTTATGGAACTAAAGC
TTCAGGAAATAAAACAGTTTTTAAGAGTAAACAAAATACTGCTTTTCAAGACCTTTCCATAAACTCT
GCAGACAAAATACATATTACCAGAAGTCATATTATGGGGGCAGAAACTCACATAGTCTCACAGACTT
GTAATCAGGATGCCAGAATATTAGCCATGACCCCAGAATCTATATATTCTAATCCATCTATTCAAGG
TTGTAAGACTGTTTTCTATTCTAGTTGTAATGATGCCATGGAAATGACCAAATGTCTCTCAAATATG
AGAGAGGAGAAAAATTTGCTAAAGCATGACAGTAATTATTCTAAAATGTATTGCAATCCAGATGCTA
TGTCTTCTCTCACAGAGAAAACTATTTATTCCGGAGAGGAGAACATGGACATTACCAAGAGTCATAC
AGTTGCAATAGATAATCAAATTTTTAAACAAGATCAATCAAATGTGCAAATAGCAGCTGCACCAACA
CCCGAAAAAGAAATGATGCTCCAAAATCTTATGACCACATCAGAAGATGGGAAAATGAATGTAAATT
GTAACTCAGTTCCTCATGTATCTAAGGAAAGAATACAGCAGAGCCTGTCAAATCCTTTGTCTATTTC
ATTGACTGATAGAAAGACTGAACTCTTATCAGGTGAAAATACGGATTTGACTGAAAGTCACACAAGT
AACTTAGGAAGTCAGGTTCCTCTTGCAGCTTATAATCTAGCACCGGAGAGTACCAGTGAATCTCACT
CTCAGAGCAAAAGCTCTTCAGATGAATGTGAAGAAATTACCAAAAGTCGTAATGAACCATTTCAGCG
ATCAGACATAATAGCCAAAAACAGCTTAACCGACACCTGGAACAAAGACAAAGATTGGGTTTTGAAG
ATTTTGCCCTACCTTGATAAAGATTCTCCTCAGTCAGCTGATTGTAATCAGGAGATAGCAACAAGCC
ATAATATAGTCTACTGTGGTGGAGTTCTTGATAAACAAATAACTAATAGAAATACAGTATCATGGGA
ACAATCTTTGTTTTCTACCACAAAGCCATTATTTTCATCAGGACAGTTCTCTATGAAAAATCATGAT
ACTGCTATAAGTAGTCATACAGTGAAATCTGTACTAGGCCAGAATTCTAAACTGGCTGAGCCACTGA
GGAAAAGTTTAAGCAATCCCACACCTGACTATTGCCATGACAAGATGATTATATGTTCAGAGGAAGA
GCAAAATATGGATCTAACAAAGAGCCACACTGTCGTCATTGGATTTGGTCCTTCTGAACTACAAGAA
CTTGGTAAAACTAATTTAGAACACACTACTGGCCAGCTAACAACAATGAACAGACAGATAGCTGTAA
AAGTTGAAAAATGTGGTAAAAGTCCCATAGAAAAAAGTGGAGTGCTTAAATCTAACTGTATTATGGA
TGTGTTAGAGGACGAAAGTGTACAGAAACCTAAATTTCCAAAGGAAAAGCAAAATGTCAAAATTTGG
GGAAGGAAAAGTGTTGGTGGACCAAAAATTGATAAGACTATTGTATTTTCAGAAGACGATAAGAATG
ATATGGATATCACTAAGAGTTATACAATAGAAATAAACCATAGACCTTTATTAGAGAAACGTGATTG
TCATTTGGTGCCATTGGCAGGAACTTCTGAAACTATTTTATATACATGTGGGCAGGATGACATGGAG
ATCACTAGAAGTCACACAACTGCCTTAGAATGTAAAACTGTCTCACCAGATGAAATAACTACTAGGC
CTATGGACAAAACTGTAGTGTTTGTAGATAATCATGTTGAACTAGAAATGACAGAGTCCCATACTGT
TTTCATTGACTACCAAGAAAAGGAAAGAACAGACAGACCTAACTTTGAACTATCCCAAAGGAAAAGC
CTAGGAACACCAACAGTGATATGTACTCCTACTGAGGAGAGTGTTTTCTTTCCAGGAAATGGTGAAA
GTGACCGTCTAGTAGCAAATGACAGCCAGCTAACCCCTCTGGAGGAATGGTCTAATAATAGGGGCCC
TGTAGAGGTAGCTGATAACATGGAATTGTCTAAATCAGCCACTTGCAAAAACATCAAAGATGTACAA
AGTCCTGGATTTCTGAATGAACCTCTATCAAGCAAAAGTCAGAGAAGAAAAAGCCTTAAGCTAAAAA
ATGACAAGACCATTGTATTTTCAGAGAATCATAAAAATGATATGGATATTACCCAGAGTTGTATGGT
GGAAATAGATAACGAAAGTGCCCTGGAGGATAAAGAGGACTTCCATTTGGCAGGGGCTTCTAAAACT
ATTTTGTATTCATGTGGGCAGGATGACATGGAGATCACTAGGAGTCACACAACTGCCTTAGAATGTA
AAACTCTCCTGCCAAACGAAATAGCTATTAGGCCCATGGACAAAACCGTATTGTTCACAGATAATTA
CAGTGATCTGGAAGTCACCGATTCCCATACTGTTTTCATTGACTGTCAAGCCACAGAGAAAATACTT
GAAGAAAACCCTAAATTTGGAATAGGAAAAGGAAAAAACTTGGGTGTTTCCTTTCCTAAGGATAATA
GCTGTGTTCAAGAAATCGCTGAAAAACAAGCACTGGCTGTAGGAAACAAAATAGTTCTTCACACCGA
GCAAAAGCAACAACTCTTTGCTGCTACTAATAGAACTACTAATGAAATCATCAAATTTCATAGTGCT
GCTATGGATGAAAAGGTCATAGGGAAAGTTGTAGACCAGGCCTGTACATTGGAAAAAGCGCAAGTTG
AAAGCTGTCAGTTAAATAATAGAGATAGAAGAAATGTGGACTTTACAAGTAGTCATGCAACTGCTGT
TTGTGGATCCAGTGATAATTATTCCTGTTTACCAAATGTTATTTCCTGTACTGATAATTTGGAGGGT
AGTGCCATGCTCTTATGTGATAAAGATGAGGAAAAAGCCAATTATTGCCCAGTGCAAAATGATCTTG
CTTATGCAAATGATTTTGCCAGTGAATATTACTTGGAATCTGAGGGACAGCCTCTCTCTGCTCCTTG
TCCTTTGTTAGAGAAGGAAGAAGTTATTCAAACCAGTACCAAAGGACAGTTAGACTGTGTTATAACA
CTGCACAAAGATCAAGATCTGATTAAGGATCCACGAAATCTATTGGCTAATCAAACTTTAGTATATA
GTCAAGATCTGGGGGGAGATGACTAAACTTAATTCAAAGCGAGTATCTTTTAAGCTTCCAAAGGATCA
AATGAAAGTCTATGTTGATGACATTTATGTTATTCCTCAGCCTCATTTCTCAACCGACCAACCTCCA
TTACCTAAAAAAGGACAGAGTAGTATCAATAAAGAAGAAGTAATACTGTCTAAAGCTGGAAATAAGA
GTTTAAATATTATAGAAAATTCCTCTGCACCCATATGTGAAAACAAGCCCAAAATACTCAATAGTGA
GGAATGGTTTGCTGCAGCCTGTAAAAAAGAACTGAAGGAAAATATTCAAACAACTAACTATAATACA
GCTCTAGATTTCCACAGTAACTCAGACGTAACTAAGCAAGTCATTCAAACTCATGTCAATGCTGGAG

AAGCACCAGATCCTGTAATTACATCTAATGTTCCATGTTTTCATAGTATCAAACCAAATCTGAATAA
TTTGAATGGAAAAACTGGAGAGTTTTTAGCCTTTCAAACTGTTCATCTACCACCCCTTCCAGAGCAA
TTACTTGAATTAGGAAATAAGGCACACAATGATATGCATATAGTGCAAGCTACAGAAATACATAATA
TTAACATAATCTCCAGCAATGCTAAAGATAGTAGAGATGAGGAAAATAAAAAGTCTCATAATGGAGC
TGAAACCACCTCTCTACCGCCAAAGACAGTTTTTAAAGATAAAGTAAGGAGATGTTCTTTGGGAATC
TTTTTGCCTAGATTGCCCAACAAGAGAAATTGTAGTGTCACTGGTATTGATGACCTGGAACAGATTC
CAGCAGACACAACTGATATAAATCACTTAGAAACTCAGCCGGTCTCTAGCAAAGATTCAGGCATTGG
ATCTGTTGCAGGTAAACTGAACCTAAGTCCTTCTCAATATATAAATGAGGAAAATCTTCCTGTATAT
CCTGATGAGATCAATTCTTCAGACTCTATTAACATAGAAACTGAGGAAAAGGCCTTGATTGAGACAT
ACCAAAAAGAGATTTCACCATATGAAAATAAAATGGGAAAAACTTGCAATAGCCAAAAAAGAACGTG
GGTACAAGAAGAAGAAGATATTCATAAGGAGAAAAAAATCAGAAAAAATGAGATTAAGTTTAGTGAT
ACGACACAAGATCGGGAGATTTTTGATCACCATACTGAAGAGGATATAGATAAAAGTGCTAACAGTG
TATTGATAAAAAACCTGAGCAGGACCCCATCTAGTTGCAGCAGCTCTCTGGATTCAATCAAGGCTGA
TGGGACCTCTCTGGACTTCAGCACTTACCGCAGTAGTCAAATGGAATCACAGTTTCTCAGAGATACT
ATTTGTGAAGAGAGCTTGAGGGAGAAACTCCAAGATGGGAGAATAACAATAAGGGAGTTCTTTATAC
TTCTCCAGGTCCACATCTTGATACAGAAACCCCGACAGAGCAATCTCCCAGGCAATTTTACTGTAAA
CACACCACCTACTCCAGAAGACCTGATGTTAAGTCAATATGTTTACCGACCCAAGATACAGATTTAT
AGAGAAGATTGTGAGGCTCGTCGCCAAAAGATTGAAGAATTAAAGCTTTCTGCATCGAACCAAGATA
AGCTGTTGGTTGATATAAATAAGAACCTGTGGGAAAAAATGAGACACTGCTCTGACAAAGAGCTGAA
GGCCTTTGGAATTTATCTTAACAAAATAAAGTCATGTTTTACCAAGATGACTAAAGTCTTCACTCAC
CAAGGAAAAGTGGCTCTGTATGGCAAGCTGGTGCAGTCAGCTCAGAATGAGAGGGAGAAACTTCAAA
TAAAGATAGATGAGATGGATAAAATACTTAAGAAGATCGATAACTGCCTCACTGAGATGGAAACAGA
AACTAAGAATTTGGAGGATGAAGAGAAAAACAATCCTGTGGAAGAATGGGATTCTGAAATGAGAGCT
GCAGAAAAAGAATTGGAACAGCTGAAAACTGAAGAAGAGGAGCTTCAAAGAAATCTCTTAGAACTGG
AGGTACAAAAAGAGCAGACCCTTGCTCAAATAGACTTTATGCAAAAACAAAGAAATAGAACTGAAGA
GCTACTGGATCAGTTGAGCTTGTCTGAGTGGGATGTCGTTGAGTGGAGTGATGATCAAGCTGTATTC
ACCTTTGTTTATGACACGATACAACTCACCATCACCTTTGAAGAGTCAGTTGTTGGTTTCCCTTTCC
TGGACAAGCGTTATAGGAAGATTGTTGATGTCAATTTTCAATCTCTGTTAGATGAGGATCAAGCTCC
TCCTTCCTCCCTTTTAGTTCATAAGCTTATTTTCCAGTACGTTGAAGAAAAGGAATCCTGGAAGAAG
ACATGTACAACCCAGCATCAGTTACCCAAGATGCTTGAAGAATTCTCACTGGTAGTGCACCATTGCA
GACTCCTTGGAGAGGAGATTGAGTATTTAAAGAGATGGGGACCAAATTATAACCTAATGAACATAGA
TATTAATAATAATGAATTGAGACTTTTATTCTCTAGCTCCGCAGCATTTGCAAAGTTTGAAATAACT
TTGTTTCTCTCAGCCTATTATCCATCTGTACCATTACCTTCCACCATTCAGAATCACGTTGGGAACA
CTAGCCAAGATGATATTGCTACCATTCTATCTAAAGTGCCACTGGAGAACAACTACCTGAAGAATGT
AGTCAAGCAAATTTACCAAGATCTGTTTCAGGACTGCCATTTCTACCAC**TAG**ACCCTTGGACCACCA
TTGGAACAACCAAGCAGAATGTACTTGATATTATTTCAGGGTCCCATTGCTGTTCAGCCTTTGTTTT
TACGTCATTACAAGCTGAGTAAAATTCCTTCTGATGATGTTATAGTTAATCTGTATGTTTTTTATAT
CTCTGCAGAATGATGGTGATGAAGTCTGGATGGTAGGCCTCATAGCCTACTATCAACTTACTCATCT
TTGTACCAAAGGTTTAAGTAATAGGACACTTAGGAAAAATGTCTCCTAACTAAACTAGTGCTTTCTG
CTTTAGTACAAGCCCTAAGGATTAACTTAAGTATAAGAAGTGTTATCACTGACAAGAACATTAGCCA
TTTTCCCATAACTAGATAGAGCTATGATTTTTTAGGTTGCCTGGCTTCTGCCTAGCAGATATTTCTG
GAGTAGAAATGTATCTGTCTACAAACTATTATCCTTTTTCTCCGTTACTAAAATGCTATTAAGAGAA
AGTAGGGCTGGGTGTGAGCCACCACACCCAGCAATGTTTTCTTAATAAGTATAGTTTTTCTAGGGAA
AGTTAATTCATTTTTGTCTAGTACATATATGTAAATATATTAATGTTGTTTTTGTGTTTGTGATGTA
GTAAGGAGATGTACATAGAAATTCATTGAGGTATATAGATACTCATCTGTCTAGGCAGTTCCCAATT
TTCTGAAGAATGTTTTACAGCAAAATTTTCTATTTTCTTTTATTAAATAGTGACACGTCAAACAATG
TCACATCCAAAACACTAGTTTCATCAATTTCTAGCAGTAATAATAGACTTGCTGTAAGTATTGTTTT
CTGATGCCATACCCTTGTCATACATATTATTAAATGACCAATATTATGTATGAAGTAGACAAAAAAA
TTTACTCAAACTTCATTCAAATCCTAATTGTGATAATTTTTGTTTTATATTTAATTATAAACCAAAA
TACATTTGCATTTTTAAGCTAATTTGTCTCAAAATTTTGCTTTATATTTTTGGATCAGGTTAAAGTC
CTGTGGATCCCCTGAATGTTATTGTCCCTCTTGATTGGTTTTTACTTCTGAGCTATACGTCAAAAGA
CACATAAGCTTCAAAAGTCAAGACAAACCTCATTTGCCATAAAAATCAAGATATAGATGTTCTGTTC
CGTAAACTCCTTGAAAAACATTTTAAAGTCATCAATATGATCTGTTTCCCATGAAACTTAAGTTAGC
TTTCTTATTGGAGTTATTTCTTTTCTGTAAGTCTGAAAAGTAGAGATTTTGTTTTACGCATTTTAGT
AACCTGCAACAACCAACTCTAAAAAAGATTTGGCTTGTAATGACGGTCTCTGCTTTTTTGGGTTTGG
AGTACACAATTGTAATATTTACTTAGTTATTTGTGTTTTTCTTTGTTCAAGGTATTGACTAGTTTCA
TAAATTTTTTGCAAGTTTTTCTTTCATTGGTTGGAAAGCAGATTACATTTTGCACTATTAAAATAAG
TTTATTACTTTAAAAAAAAAGTCGACG

Variant sequences of NOV11 are included in Example 3, Table 23. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA.

The NOV11 protein (SEQ ID NO:24) encoded by SEQ ID NO:23 is 2316 amino acid residues in length and is presented using the one-letter amino acid code in Table 11B. Psort analysis predicts the NOV11 protein of the invention to be localized at the nucleus with a certainty of 0.8800.

Table 11B. Encoded NOV11 protein sequence (SEQ ID NO:24)

```
MDGVSSEANEENDNIERPVRRRHSSILKPPRSPLQDLRGGNERVQESNALRNKKNSRRVSFADTI
KVFQTESHMKIVRKSEMEETETGENLLLIQNKKLEDNYCEITGMNTLLSAPIHTQMQQKEFSIIE
HTRERKHANDQTVIFSDENQMDLTSSHTVMITKGLLDNPISEKSTKIDTTSFLANLKLHTEDSRM
KKEVNFSVDQNTSSENKIDFNDFIKRLKTGKCSAFPDVPDKENFEIPIYSKEPNSASSTHQMHVS
LKEDENNSNITRLFREKDDGMNFTQCHTANIQTLIPTSSETNSRESKGNDITIYGNDFMDLTFNH
TLQILPATGNFSEIENQTQNAMDVTTGYGTKASGNKTVFKSKQNTAFQDLSINSADKIHITRSHI
MGAETHIVSQTCNQDARILAMTPESIYSNPSIQGCKTVFYSSCNDAMEMTKCLSNMREEKNLLKH
DSNYSKMYCNPDAMSSLTEKTIYSGEENMDITKSHTVAIDNQIFKQDQSNVQIAAAPTPEKEMML
QNLMTTSEDGKMNVNCNSVPHVSKERIQQSLSNPLSISLTDRKTELLSGENTDLTESHTSNLGSQ
VPLAAYNLAPESTSESHSQSKSSSDECEEITKSRNEPFQRSDIIAKNSLTDTWNKDKDWVLKILP
YLDKDSPQSADCNQEIATSHNIVYCGGVLDKQITNRNTVSWEQSLFSTTKPLFSSGQFSMKNHDT
AISSHTVKSVLGQNSKLAEPLRKSLSNPTPDYCHDKMIICSEEEQNMDLTKSHTVVIGFGPSELQ
ELGKTNLEHTTGQLTTMNRQIAVKVEKCGKSPIEKSGVLKSNCIMDVLEDESVQKPKFPKEKQNV
KIWGRKSVGGPKIDKTIVFSEDDKNDMDITKSYTIEINHRPLLEKRDCHLVPLAGTSETILYTCG
QDDMEITRSHTTALECKTVSPDEITTRPMDKTVVFVDNHVELEMTESHTVFIDYQEKERTDRPNF
ELSQRKSLGTPTVICTPTEESVFFPGNGESDRLVANDSQLTPLEEWSNNRGPVEVADNMELSKSA
TCKNIKDVQSPGFLNEPLSSKSQRRKSLKLKNDKTIVFSENHKNDMDITQSCMVEIDNESALEDK
EDFHLAGASKTILYSCGQDDMEITRSHTTALECKTLLPNEIAIRPMDKTVLFTDNYSDLEVTDSH
TVFIDCQATEKILEENPKFGIGKGKNLGVSFPKDNSCVQEIAEKQALAVGNKIVLHTEQKQQLFA
ATNRTTNEIIKFHSAAMDEKVIGKVVDQACTLEKAQVESCQLNNRDRRNVDFTSSHATAVCGSSD
NYSCLPNVISCTDNLEGSAMLLCDKDEEKANYCPVQNDLAYANDFASEYYLESEGQPLSAPCPLL
EKEEVIQTSTKGQLDCVITLHKDQDLIKDPRNLLANQTLVYSQDLGEMTKLNSKRVSFKLPKDQM
KVYVDDIYVIPQPHFSTDQPPLPKKGQSSINKEEVILSKAGNKSLNIIENSSAPICENKPKILNS
EEWFAAACKKELKENIQTTNYNTALDFHSNSDVTKQVIQTHVNAGEAPDPVITSNVPCFHSIKPN
LNNLNGKTGEFLAFQTVHLPPLPEQLLELGNKAHNDMHIVQATEIHNINIISSNAKDSRDEENKK
SHNGAETTSLPPKTVFKDKVRRCSLGIFLPRLPNKRNCSVTGIDDLEQIPADTTDINHLETQPVS
SKDSGIGSVAGKLNLSPSQYINEENLPVYPDEINSSDSINIETEEKALIETYQKEISPYENKMGK
TCNSQKRTWVQEEEDIHKEKKIRKNEIKFSDTTQDREIFDHHTEEDIDKSANSVLIKNLSRTPSS
CSSSLDSIKADGTSLDFSTYRSSQMESQFLRDTICEESLREKLQDGRITIREFFILLQVHILIQK
PRQSNLPGNFTVNTPPTPEDLMLSQYVYRPKIQIYREDCEARRQKIEELKLSASNQDKLLVDINK
NLWEKMRHCSDKELKAFGIYLNKIKSCFTKMTKVFTHQGKVALYGKLVQSAQNEREKLQIKIDEM
DKILKKIDNCLTEMETETKNLEDEEKNNPVEEWDSEMRAAEKELEQLKTEEEELQRNLLELEVQK
EQTLAQIDFMQKQRNRTEELLDQLSLSEWDVVEWSDDQAVFTFVYDTIQLTITFEESVVGFPFLD
KRYRKIVDVNFQSLLDEDQAPPSSLLVHKLIFQYVEEKESWKKTCTTQHQLPKMLEEFSLVVHHC
RLLGEEIEYLKRWGPNYNLMNIDINNNELRLLFSSSAAFAKFEITLFLSAYYPSVPLPSTIQNHV
GNTSQDDIATILSKVPLENNYLKNVVKQIYQDLFQDCHFYH
```

A search against the Patp database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 11C.

Table 11C. Patp results for NOV11

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAW88398 Human testis secreted protein do15_4	+1	2444	1.7e-253
>patp:AAU71933 Human bone marrow tissue polypeptide #11	+1	2444	1.7e-253
>patp:AAU71961 Human bone marrow tissue polypeptide #39	+1	2444	1.7e-253
>patp:AAU71933 Human bone marrow tissue polypeptide #11	+1	2444	1.7e-253
>patp:AAU71961 Human bone marrow tissue polypeptide #39	+1	2444	1.7e-253

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 5584 of 5584 bases (100%) identical to a gb:GENBANK-ID:AB046790|acc:AB046790.1 mRNA from *Homo sapiens* (mRNA for KIAA1570 protein, partial cds). The full amino acid sequence of the protein of the invention was found to have 1790 of 1793 amino acid residues (99%) identical to, and 1792 of 1793 amino acid residues (99%) similar to, the 1833 amino acid residue ptnr:SPTREMBL-ACC:Q9NR92 protein from *Homo sapiens* (AF15Q14 PROTEIN).

NOV11 also has homology to the proteins shown in the BLASTP data in Table 11D.

Table 11D. BLAST results for NOV11

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|18308012|gb|AAL67803.1|AF461041_1(AF461041)	AF15q14 isoform 2 [*Homo sapiens*]	2316	2316/2316 (100%)	2316/2316 (100%)	0.0
gi|9966807|ref|NP_065113.1|(NM_020380)	AF15q14 protein [*Homo sapiens*]	1833	1790/1793 (99%)	1792/1793 (99%)	0.0
gi|14749154|ref|XP_031524.1|(XM_031524)	AF15q14 protein [*Homo sapiens*]	1833	1789/1793 (99%)	1791/1793 (99%)	0.0
gi|10047205|dbj|BAB13396.1|(AB046790)	KIAA1570 protein [*Homo sapiens*]	1360	1360/1360 (100%)	1360/1360 (100%)	0.0
gi|14749150|ref|XP_012461.3|(XM_012461)	similar to KIAA1570 protein [*Homo sapiens*]	915	900/900 (100%)	900/900 (100%)	0.0

A multiple sequence alignment is given in Table 11E, with the NOV11 protein being shown on line 1 in Table 11E in a ClustalW analysis, and comparing the NOV11 protein with the related protein sequences shown in Table 11D. This BLASTP data is displayed graphically in the ClustalW in Table 11E.

Table 11E. ClustalW Analysis of NOV11

1) > NOV11; SEQ ID NO:24
2) > gi|18308012|/ AF15q14 isoform 2 [*Homo sapiens*]; SEQ ID NO:82
3) > gi|9966807|/ AF15q14 protein [*Homo sapiens*]; SEQ ID NO:83
4) > gi|14749154|/ AF15q14 protein [*Homo sapiens*]; SEQ ID NO:84
5) > gi|10047205|/ KIAA1570 protein [*Homo sapiens*]; SEQ ID NO:85

6) > gi|14749150|/ similar to KIAA1570 protein [*Homo sapiens*]; SEQ ID NO:86

```
                                10        20        30        40        50        60
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                1   MDGVSSEANEENDNIERPVRRRHSSILKPPRSPLQDLRGGNERVQESNALRNKKNSRRVS  60
gi|18308012|         1   MDGVSSEANEENDNIERPVRRRHSSILKPPRSPLQDLRGGNERVQESNALRNKKNSRRVS  60
gi|9966807|          1   MDGVSSEANEENDNIERPVRRRHSSILKPPRSPLQDLRGGNERVQESNALRNKKNSRRVS  60
gi|14749154|         1   MDGVSSEANEENDNIERPVRRRHSSILKPPRSPLQDLRGGNERVQESNALRNKKNSRRVS  60
gi|10047205|         1   ------------------------------------------------------------  1
gi|14749150|         1   ------------------------------------------------------------  1

                                70        80        90        100       110       120
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                61  FADTIKVFQTESHMKIVRKSEMEETETGENLLLIQNKKLEDNYCEITGMNTLLSAPIHTQ  120
gi|18308012|         61  FADTIKVFQTESHMKIVRKSEMEETETGENLLLIQNKKLEDNYCEITGMNTLLSAPIHTQ  120
gi|9966807|          61  FADTIKVFQTESHMKIVRKSEMEETETGENLLLIQNKKLEDNYCEITGMNTLLSAPIHTQ  120
gi|14749154|         61  FADTIKVFQTESHMKIVRKSEMEETETGENLLLIQNKKLEDNYCEITGMNTLLSAPIHTQ  120
gi|10047205|         1   ------------------------------------------------------------  1
gi|14749150|         1   ------------------------------------------------------------  1

                                130       140       150       160       170       180
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                121 MQQKEFSIIEHTRERKHANDQTVIFSDENQMDLTSSHTVMITKGLLDNPISEKSTKIDTT  180
gi|18308012|         121 MQQKEFSIIEHTRERKHANDQTVIFSDENQMDLTSSHTVMITKGLLDNPISEKSTKIDTT  180
gi|9966807|          121 MQQKEFSIIEHTRERKHANDQTVIFSDENQMDLTSSHTVMITKGLLDNPISEKSTKIDTT  180
gi|14749154|         121 MQQKEFSIIEHTRERKHANDQTVIFSDENQMDLTSSHTVMITKGLLDNPISEKSTKIDTT  180
gi|10047205|         1   ------------------------------------------------------------  1
gi|14749150|         1   ------------------------------------------------------------  1

                                190       200       210       220       230       240
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                181 SFLANLKLHTEDSRMKKEVNFSVDQNTSSENKIDFNDFIKRLKTGKCSAFPDVPDKENFE  240
gi|18308012|         181 SFLANLKLHTEDSRMKKEVNFSVDQNTSSENKIDFNDFIKRLKTGKCSAFPDVPDKENFE  240
gi|9966807|          181 SFLANLKLHTEDSRMKKEVNFSVDQNTSSENKIDFNDFIKRLKTGKCSAFPDVPDKENFE  240
gi|14749154|         181 SFLANLKLHTEDSRMKKEVNFSVDQNTSSENKIDFNDFIKRLKTGKCSAFPDVPDKENFE  240
gi|10047205|         1   ------------------------------------------------------------  1
gi|14749150|         1   ------------------------------------------------------------  1

                                250       260       270       280       290       300
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                241 IPIYSKEPNSASSTHQMHVSLKEDENNSNITRLFREKDDGMNFTQCHTANIQTLIPTSSE  300
gi|18308012|         241 IPIYSKEPNSASSTHQMHVSLKEDENNSNITRLFREKDDGMNFTQCHTANIQTLIPTSSE  300
gi|9966807|          241 IPIYSKEPNSASSTHQMHVSLKEDENNSNITRLFREKDDGMNFTQCHTANIQTLIPTSSE  300
gi|14749154|         241 IPIYSKEPNSASSTHQMHVSLKEDENNSNITRLFREKDDGMNFTQCHTANIQTLIPTSSE  300
gi|10047205|         1   ------------------------------------------------------------  1
gi|14749150|         1   ------------------------------------------------------------  1

                                310       320       330       340       350       360
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                301 TNSRESKGNDITIYGNDFMDLTFNHTLQILPATGNFSEIENQTQNAMDVTTGYGTKASGN  360
gi|18308012|         301 TNSRESKGNDITIYGNDFMDLTFNHTLQILPATGNFSEIENQTQNAMDVTTGYGTKASGN  360
gi|9966807|          301 TNSRESKGNDITIYGNDFMDLTFNHTLQILPATGNFSEIENQTQNAMDVTTGYGTKASGN  360
gi|14749154|         301 TNSRESKGNDITIYGNDFMDLTFNHTLQILPATGNFSEIENQTQNAMDVTTGYGTKASGN  360
gi|10047205|         1   ------------------------------------------------------------  1
gi|14749150|         1   ------------------------------------------------------------  1

                                370       380       390       400       410       420
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                361 KTVFKSKQNTAFQDLSINSADKIHITRSHIMGAETHIVSQTCNQDARILAMTPESIYSNP  420
gi|18308012|         361 KTVFKSKQNTAFQDLSINSADKIHITRSHIMGAETHIVSQTCNQDARILAMTPESIYSNP  420
gi|9966807|          361 KTVFKSKQNTAFQDLSINSADKIHITRSHIMGAETHIVSQTCNQDARILAMTPESIYSNP  420
gi|14749154|         361 KTVFKSKQNTAFQDLSINSADKIHITRSHIMGAETHIVSQTCNQDARILAMTPESIYSNP  420
gi|10047205|         1   ------------------------------------------------------------  1
gi|14749150|         1   ------------------------------------------------------------  1

                                430       440       450       460       470       480
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                421 SIQGCKTVFYSSCNDAMEMTKCLSNMREEKNLLKHDSNYSKMYCNPDAMSSLTEKTIYSG  480
```

```
gi|18308012|   421  SIQGCKTVFYSSCNDAMEMTKCLSNMREEKNLLKHDSNYSKMYCNPDAMSSLTEKTIYSG 480
gi|9966807|    421  SIQGCKTVFYSSCNDAMEMTKCLSNMREEKNLLKHDSNYSKMYCNPDAMSSLTEKTIYSG 480
gi|14749154|   421  SIQGCKTVFYSSCNDAMEMTKCLSNMREEKNLLKHDSNYAKMYCNPDAMSSLTEKTIYSG 480
gi|10047205|   1    ------------------------------------------------------------ 1
gi|14749150|   1    ------------------------------------------------------------ 1

                          490       500       510       520       530       540
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11          481  EENMDITKSHTVAIDNQIFKQDQSNVQIAAAPTPEKEMMLQNLMTTSEDGKMNVNCNSVP 540
gi|18308012|   481  EENMDITKSHTVAIDNQIFKQDQSNVQIAAAPTPEKEMMLQNLMTTSEDGKMNVNCNSVP 540
gi|9966807|    481  EENMDITKSHTVAIDNQIFKQDQSNVQIAAAPTPEKEMMLQNLMTTSEDGKMNVNCNSVP 540
gi|14749154|   481  EENMDITKSHTVAIDNQIFKQDQSNVQIAAAPTPEKEMMLQNLMTTSEDGKMNVNCNSVP 540
gi|10047205|   1    ------------------------------------------------------------ 1
gi|14749150|   1    ------------------------------------------------------------ 1

                          550       560       570       580       590       600
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11          541  HVSKERIQQSLSNPLSISLTDRKTELLSGENTDLTESHTSNLGSQVPLAAYNLAPESTSE 600
gi|18308012|   541  HVSKERIQQSLSNPLSISLTDRKTELLSGENTDLTESHTSNLGSQVPLAAYNLAPESTSE 600
gi|9966807|    541  HVSKERIQQSLSNPLSISLTDRKTELLSGENTDLTESHTSNLGSQVPLAAYNLAPESTSE 600
gi|14749154|   541  HVSKERIQQSLSNPLSISLTDRKTELLSGENMDLTESHTSNLGSQVPLAAYNLAPESTSE 600
gi|10047205|   1    ------------------------------------------------------------ 1
gi|14749150|   1    ------------------------------------------------------------ 1

                          610       620       630       640       650       660
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11          601  SHSQSKSSSDECEEITKSRNEPFQRSDIIAKNSLTDTWNKDKDWVLKILPYLDKDSPQSA 660
gi|18308012|   601  SHSQSKSSSDECEEITKSRNEPFQRSDIIAKNSLTDTWNKDKDWVLKILPYLDKDSPQSA 660
gi|9966807|    601  SHSQSKSSSDECEEITKSRNEPFQRSDIIAKNSLTDTWNKDKDWVLKILPYLDKDSPQSA 660
gi|14749154|   601  SHSQSKSSSDECEEITKSRNEPFQRSDIIAKNSLTDTWNKDKDWVLKILPYLDKDSPQSA 660
gi|10047205|   1    ------------------------------------------------------------ 1
gi|14749150|   1    ------------------------------------------------------------ 1

                          670       680       690       700       710       720
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11          661  DCNQEIATSHNIVYCGGVLDKQITNRNTVSWEQSLFSTTKPLFSSGQFSMKNHDTAISSH 720
gi|18308012|   661  DCNQEIATSHNIVYCGGVLDKQITNRNTVSWEQSLFSTTKPLFSSGQFSMKNHDTAISSH 720
gi|9966807|    661  DCNQEIATSHNIVYCGGVLDKQITNRNTVSWEQSLFSTTKPLFSSGQFSMKNHDTAISSH 720
gi|14749154|   661  DCNQEIATSHNIVYCGGVLDKQITNRNTVSWEQSLFSTTKPLFSSGQFSMKNHDTAISSH 720
gi|10047205|   1    ------------------------------------------------------------ 1
gi|14749150|   1    ------------------------------------------------------------ 1

                          730       740       750       760       770       780
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11          721  TVKSVLGQNSKLAEPLRKSLSNPTPDYCHDKMIICSEEEQNMDLTKSHTVVIGFGPSELQ 780
gi|18308012|   721  TVKSVLGQNSKLAEPLRKSLSNPTPDYCHDKMIICSEEEQNMDLTKSHTVVIGFGPSELQ 780
gi|9966807|    721  TVKSVLGQNSKLAEPLRKSLSNPTPDYCHDKMIICSEEEQNMDLTKSHTVVIGFGPSELQ 780
gi|14749154|   721  TVKSVLGQNSKLAEPLRKSLSNPTPDYCHDKMIICSEEEQNMDLTKSHTVVIGFGPSELQ 780
gi|10047205|   1    ------------------------------------------------------------ 1
gi|14749150|   1    ------------------------------------------------------------ 1

                          790       800       810       820       830       840
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11          781  ELGKTNLEHTTGQLTTMNRQIAVKVEKCGKSPIEKSGVLKSNCIMDVLEDESVQKPKFPK 840
gi|18308012|   781  ELGKTNLEHTTGQLTTMNRQIAVKVEKCGKSPIEKSGVLKSNCIMDVLEDESVQKPKFPK 840
gi|9966807|    781  ELGKTNLEHTTGQLTTMNRQIAVKVEKCGKSPIEKSGVLKSNCIMDVLEDESVQKPKFPK 840
gi|14749154|   781  ELGKTNLEHTTGQLTTMNRQIAVKVEKCGKSPIEKSGVLKSNCIMDVLEDESVQKPKFPK 840
gi|10047205|   1    ------------------------------------------------------------ 1
gi|14749150|   1    ------------------------------------------------------------ 1

                          850       860       870       880       890       900
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11          841  EKQNVKIWGRKSVGGPKIDKTIVFSEDDKNDMDITKSYTIEINHRPLLEKRDCHLVPLAG 900
gi|18308012|   841  EKQNVKIWGRKSVGGPKIDKTIVFSEDDKNDMDITKSYTIEINHRPLLEKRDCHLVPLAG 900
gi|9966807|    841  EKQNVKIWGRKSVGGPKIDKTIVFSEDDKNDMDITKSYTIEINHRPLLEKRDCHLVPLAG 900
gi|14749154|   841  EKQNVKIWGRKSVGGPKIDKTIVFSEDDKNDMDITKSYTIEINHRPLLEKRDCHLVPLAG 900
gi|10047205|   1    ------------------------------------------------------------ 1
gi|14749150|   1    ------------------------------------------------------------ 1

                          910       920       930       940       950       960
```

```
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                901 TSETILYTCGQDDMEITRSHTTALECKTVSPDEITTRPMDKTVVFVDNHVELEMTESHTV 960
gi|18308012|         901 TSETILYTCGQDDMEITRSHTTALECKTVSPDEITTRPMDKTVVFVDNHVELEMTESHTV 960
gi|9966807|          901 TSETILYTCGQDDMEITRSHTTALECKTVSPDEITTRPMDKTVVFVDNHVELEMTESHTV 960
gi|14749154|         901 TSETILYTCRQDDMEITRSHTTALECKTVSPDEITTRPMDKTVVFVDNHVELEMTESHTV 960
gi|10047205|         1 ----------------------------------------------------------SHTV 4
gi|14749150|         1 ------------------------------------------------------------ 1

                                  970       980       990       1000      1010      1020
                                    ....|....|....|....|....|....|....|....|....|....|....|
NOV11                961 FIDYQEKERTDRPNFELSQRKSLGTPTVICTPTEESVFFPGNGESDRLVANDSQLTPLEE 1020
gi|18308012|         961 FIDYQEKERTDRPNFELSQRKSLGTPTVICTPTEESVFFPGNGESDRLVANDSQLTPLEE 1020
gi|9966807|          961 FIDYQEKERTDRPNFELSQRKSLGTPTVICTPTEESVFFPGNGESDRLVANDSQLTPLEE 1020
gi|14749154|         961 FIDYQEKERTDRPNFELSQRKSLGTPTVICTPTEESVFFPGNGESDRLVANDSQLTPLEE 1020
gi|10047205|         5 FIDYQEKERTDRPNFELSQRKSLGTPTVICTPTEESVFFPGNGESDRLVANDSQLTPLEE 64
gi|14749150|         1 ------------------------------------------------------------ 1

                                  1030      1040      1050      1060      1070      1080
                                    ....|....|....|....|....|....|....|....|....|....|....|
NOV11               1021 WSNNRGPVEVADNMELSKSATCKNIKDVQSPGFLNEPLSSKSQRRKSLKLKNDKTIVFSE 1080
gi|18308012|        1021 WSNNRGPVEVADNMELSKSATCKNIKDVQSPGFLNEPLSSKSQRRKSLKLKNDKTIVFSE 1080
gi|9966807|         1021 WSNNRGPVEVADNMELSKSATCKNIKDVQSPGFLNEPLSSKSQRRKSLKLKNDKTIVFSE 1080
gi|14749154|        1021 WSNNRGPVEVADNMELSKSATCKNIKDVQSPGFLNEPLSSKSQRRKSLKLKNDKTIVFSE 1080
gi|10047205|        65 WSNNRGPVEVADNMELSKSATCKNIKDVQSPGFLNEPLSSKSQRRKSLKLKNDKTIVFSE 124
gi|14749150|        1 ------------MELSKSATCKNIKDVQSPGFLNEPLSSKSQRRKSLKLKNDKTIVFSE 47

                                  1090      1100      1110      1120      1130      1140
                                    ....|....|....|....|....|....|....|....|....|....|....|
NOV11               1081 NHKNDMDITQSCMVEIDNESALEDKEDFHLAGASKTILYSCGQDDMEITRSHTTALECKT 1140
gi|18308012|        1081 NHKNDMDITQSCMVEIDNESALEDKEDFHLAGASKTILYSCGQDDMEITRSHTTALECKT 1140
gi|9966807|         1081 NHKNDMDITQSCMVEIDNESALEDKEDFHLAGASKTILYSCGQDDMEITRSHTTALECKT 1140
gi|14749154|        1081 NHKNDMDITQSCMVEIDNESALEDKEDFHLAGASKTILYSCGQDDMEITRSHTTALECKT 1140
gi|10047205|        125 NHKNDMDITQSCMVEIDNESALEDKEDFHLAGASKTILYSCGQDDMEITRSHTTALECKT 184
gi|14749150|        48 NHKNDMDITQSCMVEIDNESALEDKEDFHLAGASKTILYSCGQDDMEITRSHTTALECKT 107

                                  1150      1160      1170      1180      1190      1200
                                    ....|....|....|....|....|....|....|....|....|....|....|
NOV11               1141 LLPNEIAIRPMDKTVLFTDNYSDLEVTDSHTVFIDCQATEKILEENPKFGIGKGKNLGVS 1200
gi|18308012|        1141 LLPNEIAIRPMDKTVLFTDNYSDLEVTDSHTVFIDCQATEKILEENPKFGIGKGKNLGVS 1200
gi|9966807|         1141 LLPNEIAIRPMDKTVLFTDNYSDLEVTDSHTVFIDCQATEKILEENPKFGIGKGKNLGVS 1200
gi|14749154|        1141 LLPNEIAIRPMDKTVLFTDNYSDLEVTDSHTVFIDCQATEKILEENPKFGIGKGKNLGVS 1200
gi|10047205|        185 LLPNEIAIRPMDKTVLFTDNYSDLEVTDSHTVFIDCQATEKILEENPKFGIGKGKNLGVS 244
gi|14749150|        108 LLPNEIAIRPMDKTVLFTDNYSDLEVTDSHTVFIDCQATEKILEENPKFGIGKGKNLGVS 167

                                  1210      1220      1230      1240      1250      1260
                                    ....|....|....|....|....|....|....|....|....|....|....|
NOV11               1201 FPKDNSCVQEIAEKQALAVGNKIVLHTEQKQQLFAATNRTTNEIIKFHSAAMDEKVIGKV 1260
gi|18308012|        1201 FPKDNSCVQEIAEKQALAVGNKIVLHTEQKQQLFAATNRTTNEIIKFHSAAMDEKVIGKV 1260
gi|9966807|         1201 FPKDNSCVQEIAEKQALAVGNKIVLHTEQKQQLFAATNRTTNEIIKFHSAAMDEKVIGEV 1260
gi|14749154|        1201 FPKDNSCVQEIAEKQALAVGNKIVLHTEQKQQLFAATNRTTNEIIKFHSAAMDEKVIGKV 1260
gi|10047205|        245 FPKDNSCVQEIAEKQALAVGNKIVLHTEQKQQLFAATNRTTNEIIKFHSAAMDEKVIGKV 304
gi|14749150|        168 FPKDNSCVQEIAEKQALAVGNKIVLHTEQKQQLFAATNRTTNEIIKFHSAAMDEKVIGKV 227

                                  1270      1280      1290      1300      1310      1320
                                    ....|....|....|....|....|....|....|....|....|....|....|
NOV11               1261 VDQACTLEKAQVESCQLNNRDRRNVDFTSSHATAVCGSSDNYSCLPNVISCTDNLEGSAM 1320
gi|18308012|        1261 VDQACTLEKAQVESCQLNNRDRRNVDFTSSHATAVCGSSDNYSCLPNVISCTDNLEGSAM 1320
gi|9966807|         1261 VDQACTLEKAQVESCQLNNRDRRNVDFTSSHATAVCGSSDNYSCLPNVISCTDNLEGSAM 1320
gi|14749154|        1261 VDQACTLEKAQVESCQLNNRDRRNVDFTSSHATAVCGSSDNYSCLPNVISCTDNLEGSAM 1320
gi|10047205|        305 VDQACTLEKAQVESCQLNNRDRRNVDFTSSHATAVCGSSDNYSCLPNVISCTDNLEGSAM 364
gi|14749150|        228 VDQACTLEKAQVESCQLNNRDRRNVDFTSSHATAVCGSSDNYSCLPNVISCTDNLEGSAM 287

                                  1330      1340      1350      1360      1370      1380
                                    ....|....|....|....|....|....|....|....|....|....|....|
NOV11               1321 LLCDKDEEKANYCPVQNDLAYANDFASEYYLESEGQPLSAPCPLLEKEEVIQTSTKGQLD 1380
gi|18308012|        1321 LLCDKDEEKANYCPVQNDLAYANDFASEYYLESEGQPLSAPCPLLEKEEVIQTSTKGQLD 1380
gi|9966807|         1321 LLCDKDEEKANYCPVQNDLAYANDFASEYYLESEGQPLSAPCPLLEKEEVIQTSTKGQLD 1380
gi|14749154|        1321 LLCDKDEEKANYCPVQNDLAYANDFASEYYLESEGQPLSAPCPLLEKEEVIQTSTKGQLD 1380
gi|10047205|        365 LLCDKDEEKANYCPVQNDLAYANDFASEYYLESEGQPLSAPCPLLEKEEVIQTSTKGQLD 424
gi|14749150|        288 LLCDKDEEKANYCPVQNDLAYANDFASEYYLESEGQPLSAPCPLLEKEEVIQTSTKGQLD 347
```

```
                                1390        1400        1410        1420        1430        1440
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                   1381 CVITLHKDQDLIKDPRNLLANQTLVYSQDLGEMTKLNSKRVSFKLPKDQMKVYVDDIYVI 1440
gi|18308012|            1381 CVITLHKDQDLIKDPRNLLANQTLVYSQDLGEMTKLNSKRVSFKLPKDQMKVYVDDIYVI 1440
gi|9966807|             1381 CVITLHKDQDLIKDPRNLLANQTLVYSQDLGEMTKLNSKRVSFKLPKDQMKVYVDDIYVI 1440
gi|14749154|            1381 CVITLHKDQDLIKDPRNLLANQTLVYSQDLGEMTKLNSKRVSFKLPKDQMKVYVDDIYVI 1440
gi|10047205|            425  CVITLHKDQDLIKDPRNLLANQTLVYSQDLGEMTKLNSKRVSFKLPKDQMKVYVDDIYVI 484
gi|14749150|            348  CVITLHKDQDLIKDPRNLLANQTLVYSQDLGEMTKLNSKRVSFKLPKDQMKVYVDDIYVI 407

                                1450        1460        1470        1480        1490        1500
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                   1441 PQPHFSTDQPPLPKKGQSSINKEEVILSKAGNKSLNIIENSSAPICENKPKILNSEEWFA 1500
gi|18308012|            1441 PQPHFSTDQPPLPKKGQSSINKEEVILSKAGNKSLNIIENSSAPICENKPKILNSEEWFA 1500
gi|9966807|             1441 PQPHFSADQPPLPKKGQSSINKEEVILSKAGNKSLNIIENSSAPICENKPKILNSEEWFA 1500
gi|14749154|            1441 PQPHFSTDQPPLPKKGQSSINKEEVILSKAGNKSLNIIENSSAPICENKPKILNSEEWFA 1500
gi|10047205|            485  PQPHFSTDQPPLPKKGQSSINKEEVILSKAGNKSLNIIENSSAPICENKPKILNSEEWFA 544
gi|14749150|            408  PQPHFSTDQPPLPKKGQSSINKEEVILSKAGNKSLNIIENSSAPICENKPKILNSEEWFA 467

                                1510        1520        1530        1540        1550        1560
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                   1501 AACKKELKENIQTTNYNTALDFHSNSDVTKQVIQTHVNAGEAPDPVITSNVPCFHSIKPN 1560
gi|18308012|            1501 AACKKELKENIQTTNYNTALDFHSNSDVTKQVIQTHVNAGEAPDPVITSNVPCFHSIKPN 1560
gi|9966807|             1501 AACKKELKENIQTTNYNTALDFHSNSDVTKQVIQTHVNAGEAPDPVITSNVPCFHSIKPN 1560
gi|14749154|            1501 AACKKELKENIQTTNYNTALDFHSNSDVTKQVIQTHVNAGEAPDPVITSNVPCFHSIKPN 1560
gi|10047205|            545  AACKKELKENIQTTNYNTALDFHSNSDVTKQVIQTHVNAGEAPDPVITSNVPCFHSIKPN 604
gi|14749150|            468  AACKKELKENIQTTNYNTALDFHSNSDVTKQVIQTHVNAGEAPDPVITSNVPCFHSIKPN 527

                                1570        1580        1590        1600        1610        1620
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                   1561 LNNLNGKTGEFLAFQTVHLPPLPEQLLELGNKAHNDMHIVQATEIHNINIISSNAKDSRD 1620
gi|18308012|            1561 LNNLNGKTGEFLAFQTVHLPPLPEQLLELGNKAHNDMHIVQATEIHNINIISSNAKDSRD 1620
gi|9966807|             1561 LNNLNGKTGEFLAFQTVHLPPLPEQLLELGNKAHNDMHIVQATEIHNINIISSNAKDSRD 1620
gi|14749154|            1561 LNNLNGKTGEFLAFQTVHLPPLPEQLLELGNKAHNDMHIVQATEIHNINIISSNAKDSRD 1620
gi|10047205|            605  LNNLNGKTGEFLAFQTVHLPPLPEQLLELGNKAHNDMHIVQATEIHNINIISSNAKDSRD 664
gi|14749150|            528  LNNLNGKTGEFLAFQTVHLPPLPEQLLELGNKAHNDMHIVQATEIHNINIISSNAKDSRD 587

                                1630        1640        1650        1660        1670        1680
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                   1621 EENKKSHNGAETTSLPPKTVFKDKVRRCSLGIFLPRLPNKRNCSVTGIDDLEQIPADTTD 1680
gi|18308012|            1621 EENKKSHNGAETTSLPPKTVFKDKVRRCSLGIFLPRLPNKRNCSVTGIDDLEQIPADTTD 1680
gi|9966807|             1621 EENKKSHNGAETTSLPPKTVFKDKVRRCSLGIFLPRLPNKRNCSVTGIDDLEQIPADTTD 1680
gi|14749154|            1621 EENKKSHNGAETTSLPPKTVFKDKVRRCSLGIFLPRLPNKRNCSVTGIDDLEQIPADTTD 1680
gi|10047205|            665  EENKKSHNGAETTSLPPKTVFKDKVRRCSLGIFLPRLPNKRNCSVTGIDDLEQIPADTTD 724
gi|14749150|            588  EENKKSHNGAETTSLPPKTVFKDKVRRCSLGIFLPRLPNKRNCSVTGIDDLEQIPADTTD 647

                                1690        1700        1710        1720        1730        1740
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                   1681 INHLETQPVSSKDSGIGSVAGKLNLSPSQYINEENLPVYPDEINSSDSINIETEEKALIE 1740
gi|18308012|            1681 INHLETQPVSSKDSGIGSVAGKLNLSPSQYINEENLPVYPDEINSSDSINIETEEKALIE 1740
gi|9966807|             1681 INHLETQPVSSKDSGIGSVAGKLNLSPSQYINEENLPVYPDEINSSDSINIETEEKALIE 1740
gi|14749154|            1681 INHLETQPVSSKDSGIGSVAGKLNLSPSQYINEENLPVYPDEINSSDSINIETEEKALIE 1740
gi|10047205|            725  INHLETQPVSSKDSGIGSVAGKLNLSPSQYINEENLPVYPDEINSSDSINIETEEKALIE 784
gi|14749150|            648  INHLETQPVSSKDSGIGSVAGKLNLSPSQYINEENLPVYPDEINSSDSINIETEEKALIE 707

                                1750        1760        1770        1780        1790        1800
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                   1741 TYQKEISPYENKMGKTCNSQKRTWVQEEEDIHKEKKIRKNEIKFSDTTQDREIFDHHTEE 1800
gi|18308012|            1741 TYQKEISPYENKMGKTCNSQKRTWVQEEEDIHKEKKIRKNEIKFSDTTQDREIFDHHTEE 1800
gi|9966807|             1741 TYQKEISPYENKMGKTCNSQKRTWVQEEEDIHKEKKIRKNEIKFSDTTQDREVS------ 1794
gi|14749154|            1741 TYQKEISPYENKMGKTCNSQKRTWVQEEEDIHKEKKIRKNEIKFSDTTQDREVS------ 1794
gi|10047205|            785  TYQKEISPYENKMGKTCNSQKRTWVQEEEDIHKEKKIRKNEIKFSDTTQDREIFDHHTEE 844
gi|14749150|            708  TYQKEISPYENKMGKTCNSQKRTWVQEEEDIHKEKKIRKNEIKFSDTTQDREIFDHHTEE 767

                                1810        1820        1830        1840        1850        1860
                                ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11                   1801 DIDKSANSVLIKNLSRTPSSCSSSLDSIKADGTSLDFSTYRSSQMESQFLRDTICEESLR 1860
gi|18308012|            1801 DIDKSANSVLIKNLSRTPSSCSSSLDSIKADGTSLDFSTYRSSQMESQFLRDTICEESLR 1860
gi|9966807|             1794 -------SVLNQRMFLNFGFCFVFLN--------------CGYS---------------- 1817
gi|14749154|            1794 -------SVLNQRMFLNFGFCFVFLN--------------CGYS---------------- 1817
```

```
gi|10047205|   845 DIDKSANSVLIKNLSRTPSSCSSSLDSIKADGTSLDFSTYRSSQMESQFLRDTICEESLR 904
gi|14749150|   768 DIDKSANSVLIKNLSRTPSSCSSSLDSIKADGTSLDFSTYRSSQMESQFLRDTICEESLR 827

                          1870      1880      1890      1900      1910      1920
                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11             1861 EKLQDGRITIREFFILLQVHILIQKPRQSNLPGNFTVNTPPTPEDLMLSQYVYRPKIQIY 1920
gi|18308012|      1861 EKLQDGRITIREFFILLQVHILIQKPRQSNLPGNFTVNTPPTPEDLMLSQYVYRPKIQIY 1920
gi|9966807|       1817 ----------------QILILVS----------G----------------RQKIIIS 1832
gi|14749154|      1817 ----------------QILILVS----------G----------------RQKIIIS 1832
gi|10047205|      905  EKLQDGRITIREFFILLQVHILIQKPRQSNLPGNFTVNTPPTPEDLMLSQYVYRPKIQIY 964
gi|14749150|      828  EKLQDGRITIREFFILLQVHILIQKPRQSNLPGNFTVNTPPTPEDLMLSQYVYRPKIQIY 887

                          1930      1940      1950      1960      1970      1980
                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11             1921 REDCEARRQKIEELKLSASNQDKLLVDINKNLWEKMRHCSDKELKAFGIYLNKIKSCFTK 1980
gi|18308012|      1921 REDCEARRQKIEELKLSASNQDKLLVDINKNLWEKMRHCSDKELKAFGIYLNKIKSCFTK 1980
gi|9966807|       1833 T---------------------------------------------------------- 1833
gi|14749154|      1833 T---------------------------------------------------------- 1833
gi|10047205|      965  REDCEARRQKIEELKLSASNQDKLLVDINKNLWEKMRHCSDKELKAFGIYLNKIKSCFTK 1024
gi|14749150|      888  REDCEARRQKIEEYES------------------------------------------- 903

                          1990      2000      2010      2020      2030      2040
                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11             1981 MTKVFTHQGKVALYGKLVQSAQNEREKLQIKIDEMDKILKKIDNCLTEMETETKNLEDEE 2040
gi|18308012|      1981 MTKVFTHQGKVALYGKLVQSAQNEREKLQIKIDEMDKILKKIDNCLTEMETETKNLEDEE 2040
gi|9966807|       1833 ---------------------------------------------------------- 1833
gi|14749154|      1833 ---------------------------------------------------------- 1833
gi|10047205|      1025 MTKVFTHQGKVALYGKLVQSAQNEREKLQIKIDEMDKILKKIDNCLTEMETETKNLEDEE 1084
gi|14749150|      903  ----FLHRTKISCW--LI------------------------------------------ 915

                          2050      2060      2070      2080      2090      2100
                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11             2041 KNNPVEEWDSEMRAAEKELEQLKTEEEELQRNLLELEVQKEQTLAQIDFMQKQRNRTEEL 2100
gi|18308012|      2041 KNNPVEEWDSEMRAAEKELEQLKTEEEELQRNLLELEVQKEQTLAQIDFMQKQRNRTEEL 2100
gi|9966807|       1833 ---------------------------------------------------------- 1833
gi|14749154|      1833 ---------------------------------------------------------- 1833
gi|10047205|      1085 KNNPVEEWDSEMRAAEKELEQLKTEEEELQRNLLELEVQKEQTLAQIDFMQKQRNRTEEL 1144
gi|14749150|      915  ---------------------------------------------------------- 915

                          2110      2120      2130      2140      2150      2160
                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11             2101 LDQLSLSEWDVVEWSDDQAVFTFVYDTIQLTITFEESVVGFPFLDKRYRKIVDVNFQSLL 2160
gi|18308012|      2101 LDQLSLSEWDVVEWSDDQAVFTFVYDTIQLTITFEESVVGFPFLDKRYRKIVDVNFQSLL 2160
gi|9966807|       1833 ---------------------------------------------------------- 1833
gi|14749154|      1833 ---------------------------------------------------------- 1833
gi|10047205|      1145 LDQLSLSEWDVVEWSDDQAVFTFVYDTIQLTITFEESVVGFPFLDKRYRKIVDVNFQSLL 1204
gi|14749150|      915  ---------------------------------------------------------- 915

                          2170      2180      2190      2200      2210      2220
                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11             2161 DEDQAPPSSLLVHKLIFQYVEEKESWKKTCTTQHQLPKMLEEFSLVVHHCRLLGEEIEYL 2220
gi|18308012|      2161 DEDQAPPSSLLVHKLIFQYVEEKESWKKTCTTQHQLPKMLEEFSLVVHHCRLLGEEIEYL 2220
gi|9966807|       1833 ---------------------------------------------------------- 1833
gi|14749154|      1833 ---------------------------------------------------------- 1833
gi|10047205|      1205 DEDQAPPSSLLVHKLIFQYVEEKESWKKTCTTQHQLPKMLEEFSLVVHHCRLLGEEIEYL 1264
gi|14749150|      915  ---------------------------------------------------------- 915

                          2230      2240      2250      2260      2270      2280
                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV11             2221 KRWGPNYNLMNIDINNNELRLLFSSSAAFAKFEITLFLSAYYPSVPLPSTIQNHVGNTSQ 2280
gi|18308012|      2221 KRWGPNYNLMNIDINNNELRLLFSSSAAFAKFEITLFLSAYYPSVPLPSTIQNHVGNTSQ 2280
gi|9966807|       1833 ---------------------------------------------------------- 1833
gi|14749154|      1833 ---------------------------------------------------------- 1833
gi|10047205|      1265 KRWGPNYNLMNIDINNNELRLLFSSSAAFAKFEITLFLSAYYPSVPLPSTIQNHVGNTSQ 1324
gi|14749150|      915  ---------------------------------------------------------- 915

                          2290      2300      2310
                  ....|....|....|....|....|....|....|.
NOV11             2281 DDIATILSKVPLENNYLKNVVKQIYQDLFQDCHFYH 2316
gi|18308012|      2281 DDIATILSKVPLENNYLKNVVKQIYQDLFQDCHFYH 2316
```

```
gi|9966807|    1833 ------------------------------------ 1833
gi|14749154|   1833 ------------------------------------ 1833
gi|10047205|   1325 DDIATILSKVPLENNYLKNVVKQIYQDLFQDCHFYH 1360
gi|14749150|   915  ------------------------------------ 915
```

NOV11 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV11 nucleic acids and polypeptides can be used to identify proteins that are members of the Nuclear Protein-like Protein Family. The NOV11 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV11 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, *e.g.,* cellular activation, cellular replication, and signal transduction. These molecules can be used to treat, *e.g.*, Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmume disease, allergies, immunodeficiencies, transplantation, Graft vesus host, Cardiovascular diseases, Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Systemic lupus erythematosus , Autoimmune disease, Asthma, Emphysema, Scleroderma, allergy, as well as other diseases, disorders and conditions.

In addition, various NOV11 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of sequence relatedness to previously described proteins. The NOV11 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of cardiac, immune, and nerve physiology. As such, the NOV11 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat cardiovascular, immune, and nervous system disorders, *e.g.,* Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmume disease, allergies, immunodeficiencies, transplantation, Graft vesus host, Cardiovascular diseases, Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Systemic

lupus erythematosus , Autoimmune disease, Asthma, Emphysema, Scleroderma, allergy, as well as other diseases, disorders and conditions.

The NOV11 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV11 nucleic acid is expressed in Adipose, Aorta, Artery, Coronary Artery, Umbilical Vein, Thyroid, Liver, Small Intestine, Duodenum, Colon, Ascending Colon, Bone Marrow, Lymph node, Tonsils, Thymus, Cartilage, Muscle, Brain, Cervix, Uterus, Vulva, Prostate, Testis, Lung, Bronchus, Urinary Bladder, Kidney, Skin, Epidermis, Dermis.

Additional utilities for NOV11 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV12

A NOV12 polypeptide has been identified as a Plasma Membrane Protein-like protein (also referred to as CG94282-01). The disclosed novel NOV12 nucleic acid (SEQ ID NO:25) of 8811 nucleotides is shown in Table 12A. The novel NOV12 nucleic acid sequences maps to the chromosome 12.

An ORF begins with an ATG initiation codon at nucleotides 1-3 and ends with a TAG codon at nucleotides 4378-4380. A putative untranslated region and/or downstream from the termination codon is underlined in Table 12A, and the start and stop codons are in bold letters.

Table 12A. NOV12 Nucleotide Sequence (SEQ ID NO:25)

ATGCTGTTCAAGCTCCTGCAGAGACAAACCTATACCTGCCTGTCCCACAGGTATGGGCTCTACGTGTGCTTCTT
GGGCGTCGTTGTCACCATCGTCTCCGCCTTCCAGTTCGGAGAGTGGGTAGAAGCCAGGGATCCTGCCAAACATC
CTATAGTGCACAGGACAGCCCCTACAACAAAGAATCATCCAGCCCAAAATGTCGATAGTGCTGAAGTTGAGAAA
TCCGGAATTAGAAGGGGCAAGAATGGCTGCAGGGCAGTTAGTCTACAGGACTGGCCTGGGACTAGAGGATGTGC
CAATTTCACCTTCGCCTTCTGCCATGATTGTAAGTTTTCTGAGGTCTCCCAGAAACGCTTCCTGTACATCCTGC
AGAACTGTCATTGGTTAACTGATTGGGGTTGGACTTGGTTGGCTCTGCTCCACGGGTCTCTCATCCTCCAGGGA
CCAGCCAGCGAACCTGGTTGTGTTCTTCTCAAGGCAAAGGTGGTTCTGGAATGGAGCCGAGATCAATACCATGT
TTTGTTTGATTCCTATAGAGACAATATTGCTGGAAAGTCCTTTCAGAATCGGCTTTGTCTGCCCATGCCGATTG
ACGTTGTTTACACCTGGGTGAATGGCACAGATCTTGAACTACTGAAGGAACTACAGCAGGTCAGAGAACAGATG
GAGGAGGAGCAGAAAGCAATGAGAGAAATCCTTGGGAAAAACACAACGGAACCTACTAAGAAGAGTGAGAAGCA
GTTAGAGTGTTTGCTAACACACTGCATTAAGGTGCCAATGCTTGTCCTGGACCCAGCCCTGCCAGCCAACATCA
CCCTGAAGGACCTGCCATCTCTTTATCCTTCTTTTCATTCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAA
AACCCTTCTACCAATGTCTCAGTTGTTGTTTTTGACAGTACTAAGGATGGGACATTGCTCACTCAGAAGGTGAC
TTTTGAGTGGAAATGTGAAGAAGGTGAGGTAGCCAGCAATGCGAATATCTGGGGAAAGACTGATCTGGGTTCGC
CCAGGAGGCCTTTGCCATGGCCTGTGGCCCTGGAGCCACCTAGGGCTCAGCTCAGCTCTGCCCTACAGATTCTC
ACTAGGCCACGGGTATCTCAGGACAGAGCCAACACAAGTTATGAAATTAAACTAGACACACCCCTTCTTCGAGG
TTACGCCAAGCCAGTGCCTGGGCCTGAAACTGGCCTGCAGCCCCTCAGCTTCGCCCACTGCCTTCCGACCCTGG
ACCTTCGCAAAGTGAACGAGCTTCGGGACTTCGTGAAAATGTATAAGCAGGATCCGAGCATTCTGCATACCAAG
GAAACGTGCTTTCTGAGGGAGCAGGTGGAGAGCATGGGGGAAAGCTATTATAAATCAGAAGAAAATATCAAGGA
ATTAAAAACAGGTAGTAAGAAGGTGGAGGAAAACATAAGCACAGACGAACTATCAAGTGAGGAAAGTGATCTAG
AAATTGATAACGAAGCTGTGATTGAACCAGACACTGATTCCCCTCAAGAAATGGGAGATGGAGAGGCCAGTGTA
GCGCTTCTAAAACTGAATAACCCCAAGGATTTTCAAGAATTGAATAAGCAAACTAAGAAGAACATGACCATTGA
TGGAAAAGAACTGACCATAAGTCCTGCATATTTATTATGGGATCTGAGCGCCATCAGCCAGTCTAAGCAGGATG
AAGACATCTCTGCCAGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCATTGCGATCTATCGAGAGGCATGCA

CCATGGGTTCGGAATATTTTCATTGTCACCAACGGGCAGATTCCATCCTGGCTGAACCTTGACAATCCTCGAGT
GACAATAGTAACACACCAGGATGTTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTCACCTGCTATTGAAA
GTCACATTCATCGCATCGAAGGGCTGTCCCAGAAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGAT
GTCTGGCCAGATGATTTTTACAGTCACTCCAAAGGCCAGAAGGTTTATTTGACATGGCCTGTGCCAAACTGTGC
CGAGGGCTGCCCAGGTTCCTGGATTAAGGATGGCTATTGTGACAAGGCTTGTAATAATTCAGCCTGCGATTGGG
ATGGTGGGGATTGCTCTGGAAACAGTGGAGGGAGTCGCTATATTGCAGGAGGTGGAGGTACTGGGAGTATTGGA
GTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTACTGTAATCAGGGATGTGCGAATTCCTG
GCTCGCTGATAAGTTCTGTGACCAAGCATGCAATGTCTTGTCCTGTGGGTTTGATGCTGGCGACTGTGGGCAAG
AAAACTCAGACTCAAAGAATAGGAAAACAGAGGAAAAATGCCCAGTTAAAAAAAAAAAAATCATGTTTCTGTTT
TTTCCTCTAGATCATTTTCATGAATTGTATAAAGTGATCCTTCTCCCAAACCAGACTCACTATATTATTCCAAA
AGGTGAATGCCTGCCTTATTTCAGCTTTGCAGAAGTAGCCAAAAGAGGAGTTGAAGGTGCCTATAGTGACAATC
CAATAATTCGACATGCTTCTATTGCCAACAAGTGGAAAACCATCCACCTCATAATGCACAGTGGAATGAATGCC
ACCACAATACATTTTAATCTCACGTTTCAAAATACAAACGATGAAGAGTTCAAAATGCAGATAACAGTGGAGGT
GGACACAAGGGAGGGACCAAAACTGAATTCTACAGCCCAGAAGGGTTACGAAAATTTAGTTAGTCCCATAACAC
TTCTTCCAGAGGCGGAAATCCTTTTTGAGGATATTCCCAAAGAAAAACGCTTCCCGAAGTTTAAGAGACATGAT
GTTAACTCAACAAGGAGAGCCCAGGAAGAGGTGAAAATTCCCCTGGTAAATATTTCACTCCTTCCAAAAGACGC
CCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACATGGAGACATCACTTTGAAAGGATACAATTTGTCCA
AGTCAGCCTTGCTGAGATCATTTCTGATGAACTCACAGCATGCTAAAATAAAAAATCAAGCTATAATAACAGAT
GAAACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTCATAAAAGCATCTTGCCAAACAGCTTAGGAGT
GTCTGAAAGATTGCAGAGGTTGACTTTTCCTGCAGTGAGTGTAAAAGTGAATGGTCATGACCAGGGTCAGAATC
CACCCCTGGACTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACACCCAAAAAACCATAGGCGGAAATGTG
ACAAAAGAAAAGCCCCCATCTCTGATTGTTCCACTGGAAAGCCAGATGACAAAAGAAAAGAAAATCACAGGGAA
AGAAAAAGAGAACAGTAGAATGGAGGAAAATGCTGAAAATCACATAGGCGTTACTGAAGTGTTACTTGGAAGAA
AGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTTGCCATGGGAGAAAAAAAAGTATTTCCAAGATCTTCTC
GACGAAGAAGAGTCATTGAAGACACAATTGGCATACTTCACTGATAGCAAAAATACTGGGAGGCAACTAAAAGA
TACATTTGCAGATTCCCTCAGATATGTAAATAAAATTCTAAATAGCAAGTTTGGATTCACATCGCGGAAAGTCC
CTGCTCACATGCCTCACATGATTGACCGGATTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAATTTGAC
AAGACGTCATTTCACAAAGTGCGCCATTCTGAGGATATGCAGTTTGCCTTCTCTTATTTTTATTATCTCATGAG
TGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGTCTGACA
GAGAAATCCGAACACTGGCTACCAGAATTCACGAACTGCCGTTAAGTTTGCAGGATTTGACAGGTCTGGAACAC
ATGCTAATAAATTGCTCAAAAATGCTTCCTGCTGATATCACGCAGCTAAATAATATTCCACCAACTCAGGAATC
CTACTATGATCCCAACCTGCCACCGGTCACTAAAAGTCTAGTAACAAACTGTAAACCAGTAACTGACAAAATCC
ACAAAGCATATAAGGACAAAAACAAATATAGGTTTGAAATCATGGGAGAAGAAGAAATCGCTTTTAAAATGATT
CGTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATAAGAAAAAACCCTAGGATCTCACTCTGTTGTCC
AAGCTGGAATGCAGTAATGCAAACATGGCTCACTGTAGCCTCGACCTCGTGGGCTCAAGCAATCCTCCCACCTC
AGCCTCCTGAC**TAG**TGGAACCACAGACATGAGCTGCTGCACCCAGCTAAAATGGAGTATTTTTAATTTCTGGGT
CTTTTAAATGCATTTGGAGGTCTTTAGTTTTACCTCACTGAAATTAGGATTTTAATTATAAATAATCAAAGATG
TGAACCTTACAGACATTTTAAAGCCATTATATTTTTTCTATAAACCCTGTTCTCGTTTGGAGGAGAAAGAAATT
GGAATTTTCAAAAAAAATAAAAATACCTTTAACACCTATTTAGTGTCTTTAGTAATCCAGTAAAATACTTGATT
TTTTACTAAATGTTTCCCACAAGCCAAGCAAACCATAAGCTACAATAATAATTACCTAGCGTACAGCCCTCTTT
GCATATGCTGTTCCCTCCACTTGAAGTGTACTGTTTAATTTCTTAAAATAACTTTAGCTTTTAAGAACCAATTT
TGATGGGAGTACAGACTTCCCCCATTTTCTTGATGAGTTCTCTCCGTCATGTGTAGTAATAATGTGAGAATTTG
CAGTTTTTAGTTGTAGCCTATACTTTTAGGTCTTTGTGCCAATTTGAAAGTTATTGGGTTAGAGTATTCATAGA
CATTTTCATGGTACTTAAAGGGACAGGGGTTTAGTAAAAAGACACATGGCAAGCCAGGCTTTTTCCACAGTTTG
CCAGGCCCAGCTGCCTCTTGTGTACCTGAACAGATTTTATCATTAACCCTTGTTTATGTTGTTTTGTTTTATTT
CGACGAAGGCTTATTTTAAGTCAGGCATGGAAAACTAGACTTCAGACTGACTTCAGCTTTAAGGACATGTTTAT
CCCGTTAACAGGGAGTCTGGGATAGACAATCTCCAGGCTTTGTTTTTCTCTGAATTTCTTAGCTCTGCTTGTGA
TGGCTTCATCATCAGGCCACAGACCATTAACACATTCTAGAACTTTAACATTGGTTAAATAATACCATCTAATA
GCCTGTCTTCAGCATTTCCCCAGTTGCCTCCAAATGCCCTTCATAGCTGTTCTCTGCCTCTGTTTGTTTTTAAT
CCAAGATACACTCAAGGCTCATATATTAGGTTGACATAGCTCTTTAGTATCCTTTAATTTAAAGCAGTCTCCAG
GTTTAGAGAAAGATGAATGAGCTTTCACATACCCCTCACTTGTCTTCTTCAGAAGTGTAGGCTACAACTAAAAC
TTCCTTCTTCAGAAGGAAGACAAGTGATTTATATTTATTTACTTCCATTTCTATTTGACCTTGTTTCATCTAAA
CACACCCACTCCACCACTGCTACCTGATTAATATTAAGTGAATCTCAAACATTGTATCATTTTAGCTCCACGTT
TTTTGTATGTATCTCCAAAATATAAAGATTCTTAAAAATATAACCACAATACCATTATCACCCTAAAAAAATCA
ATAATGATTCCTTAATATGACCAACTACTTTGTCAATGTACACCTTTCACTCCTCTTAACTTTCATAAAGACTT
ATGTGTTTTTTTGGTTTTTAAGTTTGTTGGTTTGAAGTTAAATCTATGGGTTTTCCCTCCATCTCTCTTTTTTT
AACCGTATAATTTTTTGCATGTGTATATGAATAAATCTGATTATAGATTCTATAGCTATCTTACACTTTGTCCC
TCTCTGATTGAATCCTAGTTAACAAGTTTCTATGTCTCTTGTATTTCCCATAAATTGGTAGTTGGATCTGAAGG
CTTTATCAGGTTTGTTTGATTTTTTTTTTTTTTAATTTTGGCGAATCTACTTCAAAAGTATTGGCCTACCTACAA
GCCACTTTAATGGGCCCTTAGTTTAGTGACCTTTGCCTTGAAAGGAACTTGAAACAAGCAAGGAAGCACCACTG
TAATCTGCTTTTTTGCCAGAACTGTAGCATCTTACAGCTTGGTTAGAGACATAGTAAGCAGAAATTATCAAATT
CATATAATCTGTAGCTATAAGGCACTGTCTCTCTCTCTCAATTATTTACATGATTTTTCTTTGTAATATAACTA
TCATTTCAGAGAACTTGGTTTTGATTTTTTTTTTTTTAATCTTTTTGAGACAGAGTCTCGCTTTATCACCCAGGC
TGGAGTGCAGTGGTGCAATCTAAAGATTGCTGACTGCAACCTCTGCCTCCCGAGTTCAGCAATTCTAGTGCCTC
AGCCTCTCGAGTAGCTGGGATTACAGGCATGCCACCACACCCGGCTAATTTTTTTGTATTTTTAGTAAAGACAG
GGTTTCACCATGTTGGCTAGGCTGGTCTCAAATTTTTGACCTCAAGTAATCAGCCTACCTTGATCTCCCAAAGT
GCTGGGATTACAGGCATGAGCCACCATGCATGGCCTTCAGAGAACTTGGTTTTAGGTACTTACGGATTGTCTTT
CTTTTTTTTCCTCACTGCAGCCTCTCCCTCCCAGGTTCAAGCGATTCTCCTACCTCAGCTTCCTGAAGAGCTGG
GACCACAGGAAGTTTGTTTGCCTGAATGACAACATTGACCACAATCATAAAGATGCTCAGACAGTGAAGGCTGT

TCTCAGGGACTTCTATGAATCCATGTTCCCCATACCTTCCCAATTTGAACTGCCAAGAGAGTATCGAAACCGTT
TCCTTCATATGCATGAGCTGCAGGAATGGAGGGCTTATCGAGACAAATTGAAGTTTTGGACCCATTGTGTACTA
GCAACATTGATTATGTTTACTATATTCTCATTTTTTGCTGAGCAGTTAATTGCACTTAAGCGGAAGATATTTCC
CAGAAGGAGGATACACAAAGAAGCTAGTCCCAATCGAATCAGAGTATAGAAGATCTTCATTTGAAAACCATCTA
CCTCAGCATTTACTGAGCATTTTAAAACTCAGCTTCACAGAGATGTCTTTGTGATGTGATGCTTAGCAGTTTGG
CCCGAAGAAGGAAAATATCCAGTACCATGCTGTTTTGTGGCATGAATATAGCCCACTGACCAGGAATTATTTAA
CCAACCCACTGAAAACTTGTGTGTTGAGCAGCTCTGAACTGATTTTACTTTTAAAGAATTTGCTCATGGACCTG
TCATCCTTTTTATAAAAAGGCTCACTGACAAGAGACAGCTGTTAATTTCCCACAGCAATCATTGCAGACTAACT
TTATTAGGAGAAGCCTATGCCAGCTGGGAGTGATTGCTAAGAGGCTCCAGTCTTTGCATTCCAAAGCCTTTTGC
TAAAGTTTTGCACTTTTTTTTTTTCATTTCCCATTTTTAAGTAGTTACTAAGTTAACTAGTTATTCTTGCTTCT
GAGTATAACGAATTGGGATGTCTAAACCTATTTTTATAGATGTTATTTAAATAATGCAGCAATATCACCTCTTA
TTGACAATACCTAAATTATGAGTTTTATTAATATTTAAGACTGTAAATGGTCTTAAACCACTAACTACTGAAGA
GCTCAATGATTGACATCTGAAATGCTTTGTAATTATTGACTTCAGCCCCTAAGAATGCTATGATTTCACGTGCA
GGTCTAATTTCAAAGGGCTAGAGTTAGTACTACTTACCAGATGTAATTATGTTTTGGAAATGTACATATTCAAA
CAGAAGTGCCTCATTTTAGAAATGAGTAGTGCTGATGGCACTGGCACATTACAGTGGTGTCTTGTTTAATACTC
ATTGGTATATTCCAGTAGCTATCTCTCTCAGTTGGTTTTTGATAGAACAGAGGCCAGCAAACTTTCTTTGTAAA
AGGCTGGTTAGTAAATTATTGCAGGCCACCTGTGTCTTTGTCATACATTCTTCTTGCTGTTGTTTAGTTTGTTT
TTTTTCAAACAACCCTCTAAAAATGTAAAAACCATGTTTAGCTTGCAGCTGTACAAAAACTGCCCACCAGCCAG
ATGTGACCCTCAGGCCATCATTTGCCAATCACTGAGAATTAGTTTTTGTTGTTGTTGTTGTTGTTGTTTTTGAG
ACAGAGTCTCTCTCTGTTGCCCAGGCTGGAGTGCAGTGGCGCAATCTCAGCTCACTGCAACCTCCGCCTCCCGG
GTTCAAGCAGTTCTGTCTCAGCCTTCTGAGTAGCTGGGACTACAGGTGCATGCCACCACACCCTGCTAATTTTT
GTATTTTTAGTAGAGACGGGGGTTCCACCATATTGGTCAGGCTTATCTTGAACTCCTGACCTCAGGTGATCCAC
CTGCCTCTGCCTCCCAAAGTGCTGAGATTACAGGCATAAGCCAGTGCACCCAGCCGAGAATTAGTATTTTTATG
TATGGTTAAACCTTGGCGTCTAGCCATATTTTATGTCATAATACAATGGATTTGTGAAGAGCAGATTCCATGAG
TAACTCTGACAGGTATTTTAGATCATGATCTCAACAATATTCTTCCAAAATGGCATACATCTTTTGTACAAAGA
ACTTGAAATGTAAATACTGTGTTTGTGCTGTAAGAGTTGTGTATTTCAAAAACTGAAATCTCATAAAAAGTTAA
ATTTT

Variant sequences of NOV12 are included in Example 3, Table 24. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA.

The NOV12 protein (SEQ ID NO:26) encoded by SEQ ID NO:25 is 1459 amino acid residues in length and is presented using the one-letter amino acid code in Table 12B. Psort analysis predicts the NOV12 protein of the invention to be localized at the plasma membrane with a certainty of 0.6500.

Table 12B. Encoded NOV12 protein sequence (SEQ ID NO:26)

MLFKLLQRQTYTCLSHRYGLYVCFLGVVVTIVSAFQFGEWVEARDPAKHPIVHRTAPTTKNHPAQ
NVDSAEVEKSGIRRGKNGCRAVSLQDWPGTRGCANFTFAFCHDCKFSEVSQKRFLYILQNCHWLT
DWGWTWLALLHGSLILQGPASEPGCVLLKAKVVLEWSRDQYHVLFDSYRDNIAGKSFQNRLCLPM
PIDVVYTWVNGTDLELLKELQQVREQMEEEQKAMREILGKNTTEPTKKSEKQLECLLTHCIKVPM
LVLDPALPANITLKDLPSLYPSFHSASDIFNVAKPKNPSTNVSVVVFDSTKDGTLLTQKVTFEWK
CEEGEVASNANIWGKTDLGSPRRPLPWPVALEPPRAQLSSALQILTRPRVSQDRANTSYEIKLDT
PLLRGYAKPVPGPETGLQPLSFAHCLPTLDLRKVNELRDFVKMYKQDPSILHTKETCFLREQVES
MGESYYKSEENIKELKTGSKKVEENISTDELSSEESDLEIDNEAVIEPDTDSPQEMGDGEASVAL
LKLNNPKDFQELNKQTKKNMTIDGKELTISPAYLLWDLSAISQSKQDEDISASRFEDNEELRYSL
RSIERHAPWVRNIFIVTNGQIPSWLNLDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIHRIEGL
SQKFIYLNDDVMFGKDVWPDDFYSHSKGQKVYLTWPVPNCAEGCPGSWIKDGYCDKACNNSACDW
DGGDCSGNSGGSRYIAGGGGTGSIGVGQPWQFGGGINSVSYCNQGCANSWLADKFCDQACNVLSC
GFDAGDCGQENSDSKNRKTEEKCPVKKKKIMFLFFPLDHFHELYKVILLPNQTHYIIPKGECLPY
FSFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHFNLTFQNTNDEEFKMQIT
VEVDTREGPKLNSTAQKGYENLVSPITLLPEAEILFEDIPKEKRFPKFKRHDVNSTRRAQEEVKI

PLVNISLLPKDAQLSLNTLDLQLEHGDITLKGYNLSKSALLRSFLMNSQHAKIKNQAIITDETND
SLVAPQEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGQNPPLDLETTARFRVETHTQKT
IGGNVTKEKPPSLIVPLESQMTKEKKITGKEKENSRMEENAENHIGVTEVLLGRKLQHYTDSYLG
FLPWEKKKYFQDLLDEEESLKTQLAYFTDSKNTGRQLKDTFADSLRYVNKILNSKFGFTSRKVPA
HMPHMIDRIVMQELQDMFPEEFDKTSFHKVRHSEDMQFAFSYFYYLMSAVQPLNISQVFDEVDTD
QSGVLSDREIRTLATRIHELPLSLQDLTGLEHMLINCSKMLPADITQLNNIPPTQESYYDPNLPP
VTKSLVTNCKPVTDKIHKAYKDKNKYRFEIMGEEEIAFKMIRTNVSHVVGQLDDIRKNPRISLCC
PSWNAVMQTWLTVASTSWAQAILPPQPPD

A search against the Patp database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 12C.

Table 12C. Patp results for NOV12

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:ABB30279 Peptide #2930 encoded by breast cell	+1	1900	7.5e-196
>patp:AAM56268 Human brain expressed single exon probe	+1	1900	7.5e-196
>patp:AAM16457 Peptide #2891 encoded by probe	+1	1900	7.5e-196
>patp:AAM28952 Peptide #2989 encoded by probe	+1	1900	7.5e-196
>patp:AAM04186 Peptide #2868 encoded by probe	+1	1900	7.5e-196

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 6444 of 6447 bases (99%) identical to a gb:GENBANK-ID:AB033034|acc:AB033034.1 mRNA from *Homo sapiens* (mRNA for KIAA1208 protein, partial cds). The full amino acid sequence of the protein of the invention was found to have 663 of 663 amino acid residues (100%) identical to, and 663 of 663 amino acid residues (100%) similar to, the 663 amino acid residue ptnr:SPTREMBL-ACC:Q9ULL2 protein from *Homo sapiens* (KIAA1208 PROTEIN).

NOV12 also has homology to the proteins shown in the BLASTP data in Table 12D.

Table 12D. BLAST results for NOV12

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|6382022|dbj|BAA86522.1|(AB033034)	KIAA1208 protein [*Homo sapiens*]	663	663/663 (100%)	663/663 (100%)	0.0

gi|16551459|dbj|BAB 71102.1|(AK056137)	unnamed protein product [*Homo sapiens*]	847	585/613 (95%)	585/613 (95%)	0.0
gi|2137411|pir||I49 528	hypothetical protein [*Mus musculus*]	384	277/400 (69%)	307/400 (76%)	e-142
gi|11360271|pir||T5 0618	hypothetical protein DKFZp762B226.1 [*Homo sapiens*]	248	134/137 (97%)	135/137 (97%)	2e-73
gi|7303923|gb|AAF58 967.1|(AE003834)	CG8027 gene product [*Drosophila melanogaster*]	652	84/155 (54%)	114/155 (73%)	9e-49

A multiple sequence alignment is given in Table 12E, with the NOV12 protein being shown on line 1 in Table 12E in a ClustalW analysis, and comparing the NOV12 protein with the related protein sequences shown in Table 12D. This BLASTP data is displayed graphically in the ClustalW in Table 12E.

Table 12E. ClustalW Analysis of NOV12

1) > NOV12; SEQ ID NO:26
2) > gi|6382022|/ KIAA1208 protein [*Homo sapiens*]; SEQ ID NO:87
3) > gi|16551459|/ unnamed protein product [*Homo sapiens*]; SEQ ID NO:88
4) > gi|2137411|/ hypothetical protein - mouse (fragment); SEQ ID NO:89
5) > gi|11360271|/ hypothetical protein DKFZp762B226.1 - human (fragment); SEQ ID NO:90
6) > gi|7303923|/ CG8027 gene product [*Drosophila melanogaster*]; SEQ ID NO:91

```
                                10        20        30        40        50        60
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                1   ----MLFKLLQRQTYTCLSHRYGLYVCFLGVVVTIVSAFQFGEWVEARDPAKHPIVHRTA 56
gi|6382022|          1   RKTEEKCPVKKKKIMFLFFPLDHFHELYKVILLPNQTHYIIPKGECLP------------ 48
gi|16551459|         1   ----MLFKLLQRQTYTCLSHRYGLYVCFLGVVVTIVSAFQF------------------- 37
gi|2137411|          1   ------------------------------------------------------------ 1
gi|11360271|         1   ------------------------------------------------------------ 1
gi|7303923|          1   ------------------------------------------------------------ 1

                                70        80        90        100       110       120
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                57  PTTKNHPAQNVDSAEVEKSGIRRGKNGCRAVSLQDWPGTRGCANFTFAFCHDCKFSEVSQ 116
gi|6382022|          48  ------------------------------------------------------------ 48
gi|16551459|         37  ----------------------G------------------------------------- 38
gi|2137411|          1   ------------------------------------------------------------ 1
gi|11360271|         1   ------------------------------------------------------------ 1
gi|7303923|          1   ------------------------------------------------------------ 1

                               130       140       150       160       170       180
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                117 KRFLYILQNCHWLTDWGWTWLALLHGSLILQGPASEPGCVLLKAKVVLEWSRDQYHVLFD 176
gi|6382022|          48  ----------------------------------------YFSFAEVA--KRGVEG 62
gi|16551459|         38  ----------------------------------------EVVLEWSRDQYHVLFD 54
gi|2137411|          1   ------------------------------------------------------------ 1
gi|11360271|         1   ------------------------------------------------------------ 1
gi|7303923|          1   ------------------------------------------------------------ 1

                               190       200       210       220       230       240
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                177 SYRDNIAGKSFQNRLCLPMPIDVVYTWVNGTDLELLKELQQVREQMEEEQKAMREILGKN 236
gi|6382022|          63  AYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHFNLTFQNT-NDEEFKMQITVEVDTRE 121
gi|16551459|         55  SYRDNIAGKSFQNRLCLPMPIDVVYTWVNGTDLELLKELQQVREQMEEEQKAMREILGKN 114
gi|2137411|          1   ------------------------------------------------------------ 1
gi|11360271|         1   ----------------------------------------PTRPVFDEVD 10
gi|7303923|          1   ------------------------------------------------------------ 1
```

```
                                  250        260        270        280        290        300
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                 237  TTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDLPSLYPSFHSASDIFNVAKPK 296
gi|6382022|           122  GPKLNSTAQKGYENLVSPITLLP--------EAEILFEDIP----------------KEK 157
gi|16551459|          115  TTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDLPSLYPSFHSASDIFNVAKPK 174
gi|2137411|             1  ----ILTTQKAYESLVSPVTPLP--------QADVPFEDVP----------------KEK  32
gi|11360271|           11  TDQSGVLSDREIRTLATRIHELP-----------LSLQDLT--------------------  40
gi|7303923|            1  ---MEQLRLQSRRKLLALTMTGG---------LCLIAIWVVAN-----------------  30

                                  310        320        330        340        350        360
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                 297  NPSTNVSVVVFDSTKDGTLLTQKVTFEWKCEEGEVASNANIWGKTDLGSPRRPLPWPVAL 356
gi|6382022|           158  R----------FPKFKR---------------HDVNS-------------------------- 169
gi|16551459|          175  NPSTNVSVVVFDSTKD-------------VEDAHS-------G------------------ 197
gi|2137411|            33  R----------FPKIRR---------------HDVNA-------------------------- 44
gi|11360271|           40  ----------------G------------------------------------------------ 41
gi|7303923|            30  ------------------------------------------------------------- 30

                                  370        380        390        400        410        420
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                 357  EPPRAQLSSALQILTRPRVSQDRANTSYEIKLDTPLLRGYAKPVPGPETGLQPLSFAHCL 416
gi|6382022|           169  ----------------------------------------------------TR---- 171
gi|16551459|          197  ---------LLKGNSRQTVWRGYLTT------D--------KEVPG-LVLMQDLAFLSGF 233
gi|2137411|            44  ----------------------------------------------------TG---- 46
gi|11360271|           41  ------------------------------------------------------------ 41
gi|7303923|            30  ------------------------------------------------------------ 30

                                  430        440        450        460        470        480
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                 417  PTLDLRKVNELRDFVKMYKQDPSILHTKETCFLREQVESMGESYYKSEENIKELKTGSKK 476
gi|6382022|           171  ----------------------------------------------------------R 172
gi|16551459|          234  P--------------PTFK-----------------ET--N----------QLKT---K 246
gi|2137411|            46  ----------------------------------------------------------R 47
gi|11360271|           41  ------------------------------------------------------------ 41
gi|7303923|            30  ------------------------------------------------------------ 30

                                  490        500        510        520        530        540
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                 477  VEENISTDELSSEESDLEIDNEAVIEPDTDSPQEMGDGEASVALLKLNNPKDFQELNKQT 536
gi|6382022|           173  AQEE-------------VKIPLVN----------------ISLLPKDAQLSLNTLDLQL 202
gi|16551459|          247  LPEN-----LSS-----KVKLLQLY-----S-------EASVALLKLNNPKDFQELNKQT 284
gi|2137411|            48  FQEE-------------VKIPRVN----------------ISLLPKEAQVRLSNLDLQL  77
gi|11360271|           41  ----------------LEHMLIN--------------------CSKMLPADITQLNNIP  64
gi|7303923|            30  ----------------YLSAS--------------------EEGQTGGFSSACTAI  50

                                  550        560        570        580        590        600
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                 537  KKN-MTIDGKELTISPAYLLWDLSAISQSKQDEDISASRFEDNEELRYSLRSIERHAPWV 595
gi|6382022|           203  EHGDITLKG--YNLSKSALLR--SFLMNSQHAKIKNQAII--TDETNDSL-VAPQ----- 250
gi|16551459|          285  KKN-MTIDGKELTISPAYLLWDLSAISQSKQDEDISASRFEDNEELRYSLRSIERHAPWV 343
gi|2137411|            78  ERGDITLKG--YNLSKSALLR--SFLGNSLDTKIKPQAR---TDETKGNL-EVPQ----- 124
gi|11360271|           65  PTQESYYDPNLPPVTKSLVTN-----CKPVTDKIHKAYK----DKNKYRFEIMGE----- 110
gi|7303923|            51  DAVYTWVNG----SDPNFIED------IRRFDDKYDPSRFDDKNELRYSLRSLEKHAAWI 100

                                  610        620        630        640        650        660
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                 596  RNIFIVTNGQIPSWLNLDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIHRIEGLSQKFI 655
gi|6382022|           250  ----------------EKQVHKSI-LPNSLGVSERLQRLTFPAVSVKVNGHDQGQNPPL 292
gi|16551459|          344  RNIFIVTNGQIPSWLNLDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHIHRIEGLSQKFI 403
gi|2137411|           124  ----------------ENPSHR---RPHGFAGEHRSERWTAPAETVTVKGRDHALNPPP 164
gi|11360271|          110  ----------------EEIAFK-------------MIRTNVSHVVGQLDDIRKNPRKFV 140
gi|7303923|           101  RHVYIVTNGQIPSWLDLSYERVTVVPHEVLAPDPDQLPTFSSSAIETFLHRIPKLSKRFL 160

                                  670        680        690        700        710        720
                                  ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12                 656  YLNDDVMFGKDVWPDDFYSHSKGQKVYLTWPVPNCAEGCPGSWIKDGYCDKACNNSACDW 715
gi|6382022|           293  DLETTARFRVETHTQKTIGGNVTKEKPPSLIVP------LESQMTK---EKKIT-----G 338
gi|16551459|          404  YLNDDVMFGKDVWPDDFYSHSKGQKVYLTWPVPNCAEGCPGSWIKDGYCDKACNNSACDW 463
gi|2137411|           165  VLETNARL-----AQPTLGVTVSKENLSPLIVP------PESHLPK---EE--------- 201
```

```
gi|11360271|   141  CLNDNIDHN----HKDAQTVKAVLRDFYESMFP--------------------------  169
gi|7303923|    161  YLNDDIFLGAPLYPEDLYTEAEGVRVYQAWMVPGCALDCPWTYTGDGACDRHCNIDACQF  220

                          730       740       750       760       770       780
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12          716  DGGDCSGNSGGSRYIAGGGGTGSIGVGQPWQFGGGINSVSYCNQGCANSWLADKFCDQAC  775
gi|6382022|    339  KEKENSRMEENAE-----------------------NHIG------VTEVLLGRKLQHYT  369
gi|16551459|   464  DGGDCSGNSGGSRYIAGGGGTGSIGVGQPWQFGGGINSVSYCNQGCANSWLADKFCDQAC  523
gi|2137411|    201  ---ESDRAEG-------------------------NAVP------VKELVPGRRCSRII  226
gi|11360271|   169  ---------------------------------IP--------SQFELPREYRNRFL  185
gi|7303923|    221  DGGDCSETGPAS-------------------DAHVIPPSKEVLEVQPAAVPQSRVHRFP  260

                          790       800       810       820       830       840
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12          776  NVLSCGFDAGDCGQENSDSKNRKTEEKCPVKKKKIMFLFFPLDHFHELYKVILLPNQTHY  835
gi|6382022|    370  DSY-LGFLPWEKKK-----------------------------YFQDLLDEEESLKT-Q-  397
gi|16551459|   524  NVLSCGFDAGDCGQ-----------------------------DHFHELYKVILLPNQTHY  555
gi|2137411|    227  Q----AFCPGKKKK-----------------------------YFQDLLDAEESLKT-Q-  251
gi|11360271|   186  H-------------------------------------------MHELQEWRAYRDK---  199
gi|7303923|    261  QMGLQKLFRRSSANFKDVMRHRN-------------------VSTLKELRRIVERFNKAKL  302

                          850       860       870       880       890       900
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12          836  IIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHFNL  895
gi|6382022|    397  -------LAYFTDSKNTGRQLKDTFADS-----------LRYVNKILNSKFGFTSRKVPA  439
gi|16551459|   556  IIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHFNL  615
gi|2137411|    251  -------LAYFTDSKHTGRQLKDTFADS-----------LRYVNKILNSKFGFTSRKVPA  293
gi|11360271|   199  -------LKFWTHCVLATLIMFTIFS-------------------------------------  218
gi|7303923|    303  MSLN--PELETSSSEPQTTQRHGLRKEDFKSSTDIYSHSLIATNMLLNRAYGFKARHVLA  360

                          910       920       930       940       950       960
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12          896  TFQNTNDEEFKMQITVEVDTREGPKLNSTAQKGYENLVSPITLLPEAEILFEDIPKEKRF  955
gi|6382022|    440  HMPHMIDR-IVMQELQDMFPEEFDKTSFHKVRHSEDMQFAFSYFYYLMSAVQPLNISQVF  498
gi|16551459|   616  TFQNTNDEEFKMQITVEVDTREGPKLNSTAQKGYENLVSPITLLPEAEILFEDIPKEKRF  675
gi|2137411|    294  HMPHMIDR-IVMQELQDMFPEEFDKTSFHKVRHSEDMQFAFSYFYYLMSAVQPLNISQVF  352
gi|11360271|   218  ----------FFAEQLIALKRKIFPRRRIHKEASPNRIRV---------------------  248
gi|7303923|    361  HVGFLIDK-DIVEAMQRRFHQQILDTAHQRFRAPTDLQYAFAYYSFLMSETKVMSVEEIF  419

                          970       980       990      1000      1010      1020
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12          956  PKFKRHDVNSTRRAQEEVKIPLVNISLLPKDAQLSLNTLDLQLEHG---DITLKGYNLSK 1012
gi|6382022|    499  DEVDTD--QSGVLSDREIRTLATRIHELPLSLQDLTGLEHMLINCS---KMLPADITQLN  553
gi|16551459|   676  PKFKRHDVNSTRRAQEEVKIPLVNISLLPKDAQLSLNTLDLQLEHG---DITLKGYNLSK  732
gi|2137411|    353  HEVDTD--QSGVLSDREIRTLATRIHDLPLTCRI-------------------------  384
gi|11360271|   248  ---------------------------------------------------------  248
gi|7303923|    420  DEFDTDG--SATWSDREVRTFLTRIYQPPLDWSAMRYFEEVVQNCTRNLGMHLKVDTVEH  477

                         1030      1040      1050      1060      1070      1080
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12         1013  SALLRSFLMNSQHAKIKNQAITTD-ETNDSLVAPQEKQVHKSILPNSLGVSERLQRLTFP 1071
gi|6382022|    554  NIPPTQESYYDPNLPPVTKSLVTN-CKPVTDKIHKAYKDKNKYRFEIMGEEEIAFKMIRT  612
gi|16551459|   733  SALLRSFLMNSQHAKIKNQAITTD-ETNDSLVAPQEKQVHKSILPNSLGVSERLQRLTFP  791
gi|2137411|    384  ---------------------------------------------------------  384
gi|11360271|   248  ---------------------------------------------------------  248
gi|7303923|    478  STLVYERYEDSNLPTITRDLVVRCPLLAEALAANFAVRPKYNFHVSPKRTSHSNFMMLTS  537

                         1090      1100      1110      1120      1130      1140
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12         1072  AVSVKVNGHDQGQNPPLDLETTARFRVETHTQKTIGGNVTKEKPPSLIVPLESQMTKEKK 1131
gi|6382022|    613  NVSHVVGQLDDIRKNPRISLCCPSWNAVMQTWLTVASTSWAQAILPPQPPD---------  663
gi|16551459|   792  AVSVKVNGHDQGQNPPLDLETTARFRVETHTQKTIGGNVTKEKPPSLIVPLESQMT----  847
gi|2137411|    384  ---------------------------------------------------------  384
gi|11360271|   248  ---------------------------------------------------------  248
gi|7303923|    538  NLTEVVESLDRLRRNPRKFNCINDNLDANRGEDNEMVRHLLEDFYLSFFPRRSKFELPPQ  597

                         1150      1160      1170      1180      1190      1200
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV12         1132  ITGKEKENSRMEENAENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYFQDLLDEEESLK 1191
gi|6382022|    663  ---------------------------------------------------------  663
```

```
gi|16551459|   847  ---------------------------------------------------------- 847
gi|2137411|    384  ---------------------------------------------------------- 384
gi|11360271|   248  ---------------------------------------------------------- 248
gi|7303923|    598  YRNRFESWRDFQRWKRRKRAVLVIGYGVSLLLVVCLLRFMCHHKAKLVRRCVQRL----- 652
```

The NOV12 Clustal W alignment shown in Table 12E was modified to end at amino residue 1200. The data in Table 1E includes all of the regions overlapping with the NOV12 protein sequences.

NOV12 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV12 nucleic acids and polypeptides can be used to identify proteins that are members of the Plasma Membrane Protein-like Protein Family. The NOV12 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV12 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, *e.g.,* cellular activation and signal transduction. These molecules can be used to treat, *e.g.*, Diabetes,Von Hippel-Lindau (VHL) syndrome , Pancreatitis, Obesity, 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, Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation, Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection as well as other diseases, disorders and conditions.

In addition, various NOV12 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the sequence relatedness to previously described proteins. The NOV12 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of cardiac, nerve, and immune physiology. As such, the NOV12 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat cardiovascular, immune, and nervous system disorders, *e.g.,* Diabetes,Von Hippel-Lindau (VHL) syndrome, Pancreatitis, Obesity, 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, Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation,

Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection as well as other diseases, disorders and conditions.

The NOV12 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV12 nucleic acid is expressed in Pancreas, Uterus, Epidermis, Heart, Coronary Artery, Adrenal Gland/Suprarenal gland, Pancreas, Parathyroid Gland, Salivary Glands, Liver, Small Intestine, Bone Marrow, Peripheral Blood, Lymphoid tissue, Lymph node, Cartilage, Brain, Hypothalamus, Spinal Chord, Mammary gland/Breast, Uterus, Prostate, Testis, Lung, Kidney, Epidermis, Hair Follicle.

Additional utilities for NOV12 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV13

A NOV13 polypeptide has been identified as a BHLH Factor MATH6-like protein (also referred to as CG94399-01). The disclosed novel NOV13 nucleic acid (SEQ ID NO:27) of 2244 nucleotides is shown in Table 13A. The novel NOV13 nucleic acid sequences maps to the chromosome 2.

An ORF begins with an ATG initiation codon at nucleotides 105-107 and ends with a TGA codon at nucleotides 1062-1064. A putative untranslated region and/or downstream from the termination codon is underlined in Table 13A, and the start and stop codons are in bold letters.

Table 13A. NOV13 Nucleotide Sequence (SEQ ID NO:27)

ACGCGTGAAGGGCGGGCGAAGCGGGAGAGCCAGAGACTCCTCGGCGCTGAGCGCGGCGGCGGCCCGG
GCAGCCCCACGCCCCTGCCTCGCGCGCCGCCCGCGCC**ATG**AAGCACATCCCGGTCCTCGAGGACGGG
CCGTGGAAGACCGTGTGCGTGAAGGAGCTGAACGGCCTTAAGAAGCTCAAGCGGAAAGGCAAGGAGC
CGGCGCGGCGCGCGAACGGCTATAAAACTTTCCGACTGGACTTGGAAGCGCCCGAGCCCCGCGCCGT
AGCCACCAACGGGCTGCGGGACAGGACCCATCGGCTGCAGCCGGTCCCGGTACCGGTCCGGTGCCAG
TCCCAGTGGCGCCGGCCGTTCCCCCAAGAGGGGGCACGGACACAGCCGGGGAGCGCGGGGGCTCTCG
GGCGCCCGAGGTCTCCGACGCGCGGAAACGTGCTTCGCCCTAGGCGCAGTGGGGCCAGGACTCCCCA
CGCCGCCGCCGCCGCCGCCTCCTGCGCCCCAGAGCCAGGCACCTGGGGGCCCAGAGGCACAGCCTTT
CGGGAGCCGGGTCTGCGTCCTCGCATCTTGCTGTGCGCACCGCCCGCGCCCCGCGCCGTCAGCACCC
CCAGCACCGCCAGCGCCCCCGGAGTCCACTGTGCGCCCTGGCCCCCGACGCGCCCCGGGGAAAGTTC
CTACTCGTCAATTTCACACGTAATTTACAATAACCACCAGGATTCCTCCGCGTCGCCTAGGAAACGA
CCGGGCGAAGCGACTGCCGCCTCCTCCGAGATCAAAGCCCTGCAGCAGACCCGGAGGCTCCTGGCGA

ACGCCAGGGAGCGGACGCGGGTGCACACCATCAGCGCAGCCTTCGAGGCGCTCAGGAAGCAGGTGCC
GTGCTACTCATATGGGCAGAAGCTGTCCAAACTGGCCATCCTGAGGATCGCCTGTAACTACATCCTG
TCCCTGGCGCGGCTGGCTGACCTTGACTACAGTGCCGACCACAGCAACCTCAGCTTCTCCGAGTGTG
TGCAGCGCTGCACCCGCACCCTGCAGGCCGAGGGACGTGCCAAGAAGCGCAAGGAG**TGA**CTGGCTGC
AGGCAAGACCAAGGCCACCACTGTGGGCCCTCCTTCCAGTCAGGCCTGAGGACAAGGTGAGCTCGCT
GAGTCCAGCCTCGTGGTCTTCTCCAAGATGGCGCCCCACTTGGAGCCTACAGCCTCTCAGGGTCGGA
TCGGAGCACGCCTGCCTCCCTCTCCCCTCCGCCCTCACCCAGCCAATCCGAGGCTGCTTCGCACGTT
GCCCTCTGCCTGGTGGGGAGGGGAGAGCTCAGCCCCCGACTCACTCAGACCCCAAGGCCCACTGTCC
AGCTGCAGAAATTCGTTGCCAAAGATTGGACAGAGACACCGAAGGAAATGGGGTGGTGAAACCCCAC
AGCGAAAAGCCACACCGTTGCTCTGTGACTTTTGCTCCTCCTGTTGCCTGAGCCCCATCTCAAGCCA
AAGATGAGTCAGTGGTTCTGCTAGGAACTCATGGAATGGATGGGCATTTGATGACCCCTGGGGGTCA
TCTTGGCCCTCTGACCTGGTGCTCTCTCTCCACTGGGCCTTGTGCTGGTTGAGTGCAAGACAAGCCT
TAGGGGCTGTGAGAGGGAGGCTGGGGTGCCTGGGCGGGGCTGGGAGTGGGACCTGAGATCCCTGCCC
ACTCTCTCCCCTTCATTGGCTTGCCCAGGCCACTGGCCCCAGTTCTCAGTGTCCCTTGGGGTCCAGG
CTCCTTGGGCCCTAAGCATCACCAGAAGGGAGTAAGCAGGGAGAGAAGCAATATTACTCCCTCCCCT
ACACCAGGGACTTGCCCCAGGGCGGCTACCTATGGGTCTTTGCTTCCCCAGCCAGCCTCTCCTCACT
GTGACCCACCCCCATGGGCCCCCGTCCCAGGCAGCCAGCACCATGGGCAGGCCCTGCCATGGACAGA
AAAAGAGTTTTTCTCTTGTTCAGCCTGCACGTGGCCTGAGGAAGGAGTAGAGGCTGGGTTGGCTGGA
GCCGTCCTACTGGGCAAGATGGCGCCCCACTTGGAGGGCGGTGGTCTGTTACAGGGTGTGCAGGGGC
AGAGAAGGAAGGGACCAGGGGACTGGGCCAGTATGTGGAGGATGGGGCCTGCGTGTTCAAAGCCAAG
GCCCGCCCCTTCCTTGTGCTCAAATGGCCAAAGCTGTTCACGTCTGTGCTCAACCATCTGCTTCAAA
TTGAAGTAAAAGCCCCAAAATGTCAAGAAAAAA

Variant sequences of NOV13 are included in Example 3, Table 25. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA.

The NOV13 protein (SEQ ID NO:28) encoded by SEQ ID NO:27 is 319 amino acid residues in length and is presented using the one-letter amino acid code in Table 13B. Psort analysis predicts the NOV13 protein of the invention to be localized at the nucleus with a certainty of *0.7000*.

Table 13B. Encoded NOV13 protein sequence (SEQ ID NO:28)

MKHIPVLEDGPWKTVCVKELNGLKKLKRKGKEPARRANGYKTFRLDLEAPEPRAVATNGLRDRTH
RLQPVPVPVRCQSQWRRPFPQEGARTQPGSAGALGRPRSPTRGNVLRPRRSGARTPHAAAAAASC
APEPGTWGPRGTAFREPGLRPRILLCAPPAPRAVSTPSTASAPGVHCAPWPPTRPGESSYSSISH
VIYNNHQDSSASPRKRPGEATAASSEIKALQQTRRLLANARERTRVHTISAAFEALRKQVPCYSY
GQKLSKLAILRIACNYILSLARLADLDYSADHSNLSFSECVQRCTRTLQAEGRAKKRKE

A search against the Patp database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 13C.

Table 13C. Patp results for NOV13			
Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAM93743 Human polypeptide, SEQ ID NO: 3717	+1	1197	2.3e-121
>patp:AAB95274 Human protein sequence SEQ ID NO:17476	+1	1197	2.3e-121
>patp:AAU16607 Human novel secreted protein, Seq ID 1560	+1	534	4.2e-51
>patp:ABG00300 Novel human diagnostic protein #291	+1	334	6.8e-33
>patp:ABG00300 Novel human diagnostic protein #291	+1	334	6.8e-33

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 372 of 657 bases (56%) identical to a gb:GENBANK-ID:HSBBICP4A|acc:L14320.1 mRNA from Bovine herpesvirus 1 (Bovine herpesvirus type 1 early-intermediate transcription control protein (BICP4) gene, complete cds). The full amino acid sequence of the protein of the invention was found to have 238 of 322 amino acid residues (73%) identical to, and 244 of 322 amino acid residues (75%) similar to, the 322 amino acid residue ptnr:TREMBLNEW-ACC:BAB39468 protein from *Mus musculus* (BHLH FACTOR MATH6)..

NOV13 also has homology to the proteins shown in the BLASTP data in Table 13D.

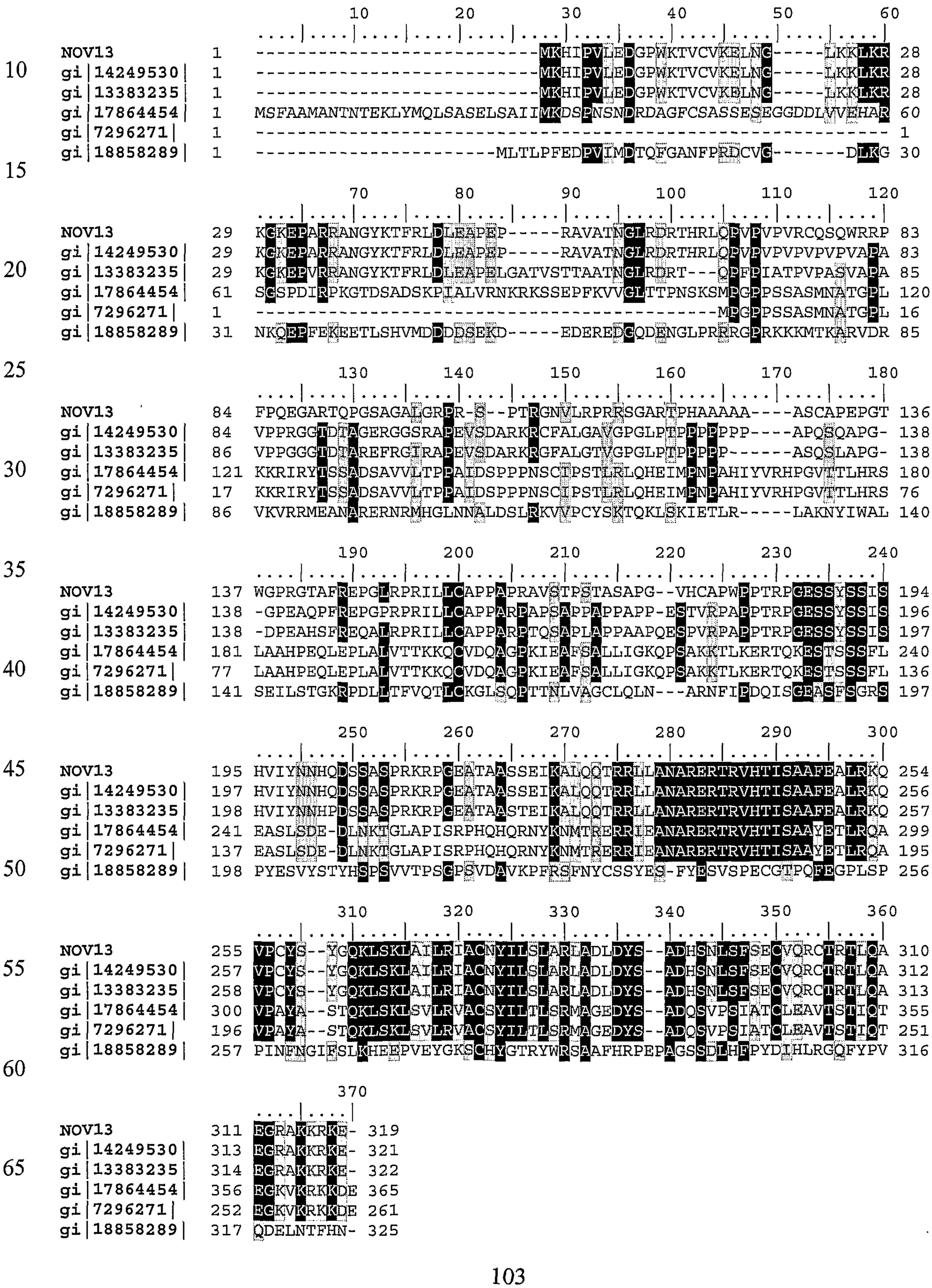
Table 13D. BLAST results for NOV13					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|14249530|ref|NP_116216.1|(NM_032827)	hypothetical protein FLJ14708 [*Homo sapiens*]	321	246/321 (76%)	246/321 (76%)	2e-93
gi|13383235|dbj|BAB39468.1|(AB049066)	bHLH factor Math6 [*Mus musculus*]	322	233/329 (70%)	240/329 (72%)	3e-86
gi|17864454|ref|NP_524820.1|(NM_080081)	net [*Drosophila melanogaster*]	365	55/97 (56%)	76/97 (77%)	5e-21
gi|7296271|gb|AAF51562.1| (AE003590)	CG11450 gene product [*Drosophila melanogaster*]	261	55/97 (56%)	76/97 (77%)	2e-20
gi|18858289|ref|NP_571891.1|(NM_131816)	atonal homolog 2a [*Danio rerio*]	325	36/82 (43%)	51/82 (61%)	2e-10

A multiple sequence alignment is given in Table 13E, with the NOV13 protein being shown on line 1 in Table 13E in a ClustalW analysis, and comparing the NOV13 protein with the related protein sequences shown in Table 13D. This BLASTP data is displayed graphically in the ClustalW in Table 13E.

Table 13E. ClustalW Analysis of NOV13

1) > NOV13; SEQ ID NO:28
2) > gi|14249530|/ hypothetical protein FLJ14708 [*Homo sapiens*]; SEQ ID NO:92

3) > gi|13383235|/ bHLH factor Math6 [*Mus musculus*]; SEQ ID NO:93
4) > gi|17864454|/ net [*Drosophila melanogaster*]; SEQ ID NO:94
5) > gi|7296271|/ CG11450 gene product [*Drosophila melanogaster*]; SEQ ID NO:95
6) > gi|18858289|/ atonal homolog 2a [*Danio rerio*]; SEQ ID NO:96



The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 13F lists the domain description from DOMAIN analysis results against NOV13.

Table 13F Domain Analysis of NOV13

Model	Region of Homology	Score (bits)	E value
Helix-loop-helix domain	234-280	55.8	4.0e-09
Helix-loop-helix DNA-binding domain (HLH)	229-281	61.4	1.9e-14

Consistent with other known members of the BHLH Factor MATH6-like family of proteins, NOV13 has, for example, a Helix-loop-helix domain and a Helix-loop-helix DNA binding domain (HLH) signature sequence as well as homology to other members of the BHLH Factor MATH6-like Protein Family. NOV13 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts.

For example, NOV13 nucleic acids and polypeptides can be used to identify proteins that are members of the BHLH Factor MATH6-like Protein Family. The NOV13 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV13 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, *e.g.*, cellular activation, cellular replication, and signal transduction. These molecules can be used to treat, *e.g.*, Diabetes,Von Hippel-Lindau (VHL) syndrome , Pancreatitis, Obesity, Inflammatory bowel disease, Diverticular disease, Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Systemic lupus erythematosus , Autoimmune disease, Asthma, Emphysema, Scleroderma, allergy as well as other diseases, disorders and conditions.

In addition, various NOV13 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV13

nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the BHLH Factor MATH6-like Protein Family.

A number of eukaryotic proteins, probably sequence specific DNA- binding proteins that act as transcription factors belong to this family. They share a conserved domain that is formed of two amphipathic helices joined by a variable length linker region that could form a loop (Littlewood and Evan, *Protein Prof.* 2: 621-702 (1995).) This 'helix-loop-helix' (HLH) domain mediates protein dimerization and has been found in a large variety of proteins (Garrell and Campuzano, *Bioessays* 13: 493-498 (1991); Kato and Dang, *FASEB J.* 6: 3065-72 (1992).) Most of these proteins have an short basic region adjacent to the HLH domain that specifically binds to DNA. They are referred as basic helix-loop-helix proteins (bHLH), and are classified in two groups: class A (ubiquitous) and class B (tissue-specific). The HLH proteins lacking the basic domain (Emc, Id) function as negative regulators since they form heterodimers, but fail to bind DNA. The hairy-related proteins (hairy, E(spl), deadpan) also repress transcription although they can bind DNA. The proteins of this subfamily act together with co-repressor proteins, like groucho, through their C-terminal motif WRPW.

MATH6 (Inoue, *et al.*, *Genes to Cells* 6: 977-86 (2001)) is a distant homolog of *Drosophila* proneuronal gene A*tonal*. Murine expression is higest in developing nervous system (ventricular zone and mantle layer, spinal cord, dorsal root ganglia). MATH6 is expressed by neuronal precursor cells and designated neurons, *e.g.*, cerebellar Purkinje cells.

The closest mammalian homolog to MATH6 is NeuroD. NeuroD point mutations and NeuroD gene knockout animals have severe diabetes and die perinatally. The NeuroD knockout animals lack beta-Islet cells and could not be rescued with insulin administration. Also, the NeuroD knockout animals are deaf due to a loss of inner ear sensory neurons.

The NOV13 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of metabolism as well as nerve and immune physiology. As such, the NOV13 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat metabolic, hearing, nervous system, immune disorders, *e.g.*, Diabetes,Von Hippel-Lindau (VHL) syndrome , Pancreatitis, Obesity, Inflammatory bowel disease, Diverticular disease, Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Systemic

lupus erythematosus , Autoimmune disease, Asthma, Emphysema, Scleroderma, allergy as well as other diseases, disorders and conditions.

The NOV13 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV13 nucleic acid is expressed in Pancreas, Umbilical Vein, Small Intestine, Cartilage, Synovium/Synovial membrane, Brain, Placenta, Oviduct/Uterine Tube/Fallopian tube, Lung, Brain, Uterus.

Additional utilities for NOV13 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV14

A NOV14 polypeptide has been identified as a Putative Protein-Tyrosine Phosphatase-like protein (also referred to as CG94366-01). The disclosed novel NOV14 nucleic acid (SEQ ID NO:29) of nucleotides is shown in Table 14A. The novel NOV14 nucleic acid sequences maps to the chromosome 22.

An ORF begins with an ATG initiation codon at nucleotides 248-250 and ends with a TAA codon at nucleotides 1679-1681. A putative untranslated region and/or downstream from the termination codon is underlined in Table 14A, and the start and stop codons are in bold letters.

Table 14A. NOV14 Nucleotide Sequence (SEQ ID NO:29)

ATTGAGTTTGAAATAACTGCCACCACAAAGTCTGTCACACATTGAGACTGAGGTCATAATAAAGAGG
TTTACTTAAATAGGGAAGCATTACTATTTTCCCCCGCCTAAGATTTTGGTTGTCGCCATATAAATCC
TCATTTCTAATAAAGAGAAAAAGACATTCCAGGTTCCAATAGTGCTATACAC**ATG**AATAGTCAGAAA
TTAATTGGTTTCTGTCTAGAATAATGAAAAGTAATTTTTCCAAAATATGAATTCAGAATTAAGTCTC
CTCTCTGACTGTTTTCTCTTATCATCCGCTAGTCCACAGACAAACGAATTTAAAGGAGCAACCGAGG
AGGCACCTGCGAAAGAAAGCCCACACACAGGTGAATTTAAAGGAGCAGCCCTGGTGTCACCTATCAG
TAAAAGAATGTTAGAACGACTTTCCAAGTTTGAAGTTGGAGATGCTGAAAATGTTGCTTCATATGAC
AGCAAGATTAAGAAAATTGTTCATTCAATTGTGTCATCCTTTGCAGTTGGGATATTTGGAGTTTTCC
TGATCTTGTTGGATGTGGCTCTGATCTTTGCTGACCTAATTTTCACTGATAGCAAAGTTTATATTCC
TTTGGAGTATCGTTCTATTTCTCTAGCTATTGCTTTATTTTTTCTCATGGATGTTCTTCTTCGAGTA
TTTGTAGAAAGGAGACAGCATTATTTTTCTGATTTACTTAACATTTTAGATACTGCCGTTACTGTGA
TTATTCTGCTGGTTGATGTCGTTTACATTTTTTTTGACGTTAAGTTTCTTAAGGATATTCCCAGATG
GACACGTTTATTTCGACTTCTACGACTTATCATTCTGATAAGAGTTTTTCGTCTGGCTCATCTAAAA
AGACAACTTGGAAAGCTGATAAGAAGGCTGGTAAGTAGGNGATACGAAAGGGATGGATTTGACCTAG
ACCTCACTTATATTACAGAACGTATTGTCGCTATGTCATTTCCATCTTCGGGAGGCCAGTCTTTCTA
TCGGAATCCAATTAAGGAAGTCGTACAGTTTCTAGACAAGAAACATCCAAACCACTATCGAGTCTAC
AATCTATGCAGTGAAAGAGCTTATGATCCTAAGCACTTCCATAATAGGGTCAGTAGAATCATGATCG
ATGATCATAATGTCCCCACTCTAAGGGAGATGGTAGCATTCTCCAAGGAAGTGTTGGAGTGGATGGC
TCAAGATTCTGAAAACATCGTAGTGATTCACTGTAAAGGAGGCAAAGGTAGAACCGGAACTATGGTT
TGTGCCTGCCTGATTGCCAGTGAAATATTTTTAACTGCAGAGGAAAGATTGTACTATTTTGGAGAAC
GGCGAACAGATAAAACCAATGGCACTAAATATCAGGGAGTAGAAACTCCTTCTCAGAATAGATATGT
TGGATATTTTGCACAAGTGAAACATAGCTACAACTGGAATCTCCCTCCAAGAAAAACACTGTTTATA
AAAAGATTAGTTATTTATTCGATTCATGGTAAGTGTTTAGATCTAAAAGTCCAAATAGTAATGAAGA

AAAAGATTGTCTTTTCCTGCACTTCCTTAAACAGTTGTCGGGTAAGAGAAAACATGGAAACAGACAG
GGTAATAATTGATGTGTTCAACTGTCCACCTCTGTATGATGATGTGAAAGTGCAATTTTTTTTTTCT
TTT**TAG**GATTTTCCTAAATACTATCACAACTACCCTTTTTTCTTCTGGTTTAACACATCTTTAATAC
AAAATAACAGGCTTTATCTACAAAGAAATGAATTGGATAATCTTCATAAACAAAAAACATGGAAAAT
TTATCAACCAGAATATGCAGTAGAGATATATTTTGATGAGAAATGACTTAAGTTATGTTGTAACTGG
TAGCTGATTAAGTATAGTTCCCTGCACCCCTTCTGGGAAAGAATTATGTTCTTTCTAACCCTGCCAC
ATAGTTATATGTTCTAAATCTTCCTTGCTGGTACATCTATATTGATATATGTATACACATGTTCTTT
ATAAATCTATTAAATATATACAGATAAA

The NOV14 protein (SEQ ID NO:30) encoded by SEQ ID NO:29 is 477 amino acid residues in length and is presented using the one-letter amino acid code in Table 14B. Psort analysis predicts the NOV14 protein of the invention to be localized at the plasma membrane with a certainty of 0.6000.

Table 14B. Encoded NOV14 protein sequence (SEQ ID NO:30)

MNSELSLLSDCFLLSSASPQTNEFKGATEEAPAKESPHTGEFKGAALVSPISKRMLERLSKFEVG
DAENVASYDSKIKKIVHSIVSSFAVGIFGVFLILLDVALIFADLIFTDSKVYIPLEYRSISLAIA
LFFLMDVLLRVFVERRQHYFSDLLNILDTAVTVIILLVDVVYIFFDVKFLKDIPRWTRLFRLLRL
IILIRVFRLAHLKRQLGKLIRRLVSRXYERDGFDLDLTYITERIVAMSFPSSGGQSFYRNPIKEV
VQFLDKKHPNHYRVYNLCSERAYDPKHFHNRVSRIMIDDHNVPTLREMVAFSKEVLEWMAQDSEN
IVVIHCKGGKGRTGTMVCACLIASEIFLTAEERLYYFGERRTDKTNGTKYQGVETPSQNRYVGYF
AQVKHSYNWNLPPRKTLFIKRLVIYSIHGKCLDLKVQIVMKKKIVFSCTSLNSCRVRENMETDRV
IIDVFNCPPLYDDVKVQFFFSF

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

Table 14C. Patp results for NOV14

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAG67459 Amino acid sequence of a human polypeptide	+1	1895	2.5e-195
>patp:AAG67638 Amino acid sequence of a human protein	+1	1895	2.5e-195
>patp:AAB73230 Human phosphatase AA493915_h	+1	574	1.8e-111
>patp:AAW34402 Protein encoded by gene IMAGE clone 264611	+1	473	1.2e-44
>patp:AAY07450 Human TS10q23.3 gene bases 453-2243	+1	473	1.2e-44

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1105 of 1427 bases (77%) identical to a gb:GENBANK-ID:AF007118|acc:AF007118.1 mRNA from *Homo sapiens* (putative tyrosine phosphatase mRNA, complete cds). The full amino acid sequence of the protein of the invention was found to have 369 of 462 amino acid residues (79%) identical to, and 402 of 462 amino acid residues (87%) similar to, the 551 amino acid residue ptnr:SWISSNEW-

ACC:P56180 protein from *Homo sapiens* (PUTATIVE PROTEIN-TYROSINE PHOSPHATASE TPTE (EC 3.1.3.48)).

NOV14 also has homology to the proteins shown in the BLASTP data in Table 14D.

Table 14D. BLAST results for NOV14

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|7019559|ref|NP_037447.1|(NM_013315)	transmembrane phosphatase with tensin homology; tensin, putative protein-tyrosine phosphatase [*Homo sapiens*]	551	369/462 (79%)	402/462 (86%)	0.0
gi|16166555|ref|XP_055073.1|(XM_055073)	similar to transmembrane phosphatase with tensin homology [*Homo sapiens*]	551	367/462 (79%)	401/462 (86%)	0.0
gi|18640756|ref|NP_570141.1|(NM_130785)	similar to PUTATIVE PROTEIN-TYROSINE PHOSPHATASE TPTE [*Homo sapien*]	445	316/462 (68%)	343/462 (73%)	e-156
gi|14787415|emb|CAC44243.1|(AJ311311)	tyrosine phosphatase isoform A [*Mus musculus*]	664	275/432 (63%)	336/432 (77%)	e-141
gi|14787417|emb|CAC44244.1|(AJ311312)	tyrosine phosphatase isoform B [*Mus musculus*]	645	275/432 (63%)	336/432 (77%)	e-141

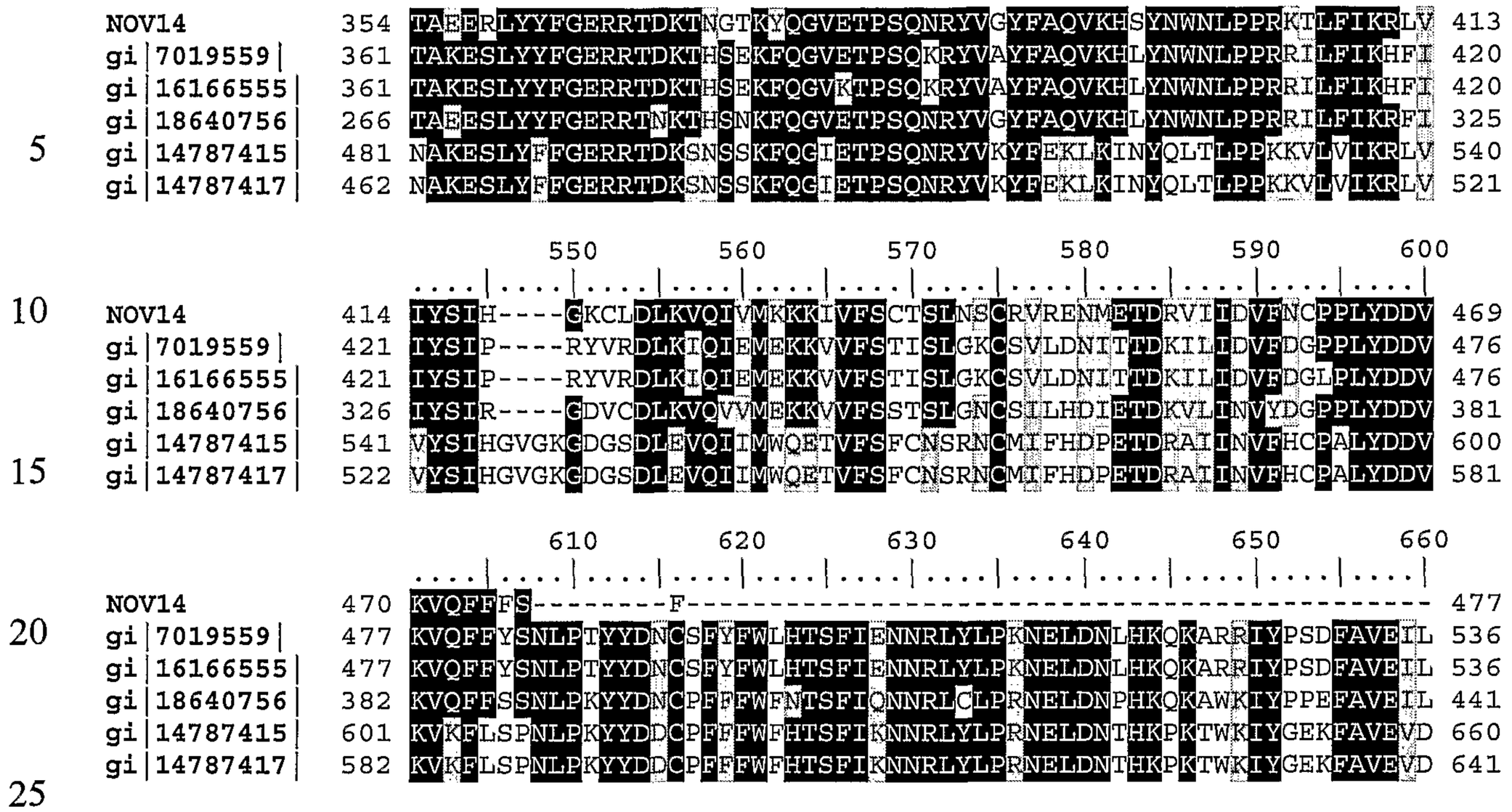
A multiple sequence alignment is given in Table 14E, with the NOV14 protein being shown on line 1 in Table 14E in a ClustalW analysis, and comparing the NOV14 protein with the related protein sequences shown in Table 14D. This BLASTP data is displayed graphically in the ClustalW in Table 14E.

Table 14E. ClustalW Analysis of NOV14

1) > NOV14; SEQ ID NO:30
2) > gi|7019559|/ transmembrane phosphatase with tensin homology; tensin, putative protein-tyrosine phosphatase *[Homo sapiens]*; SEQ ID NO:97
3) > gi|16166555|/ similar to transmembrane phosphatase with tensin homology [*Homo sapiens*]; SEQ ID NO:98
4) > gi|18640756|/ similar to PUTATIVE PROTEIN-TYROSINE PHOSPHATASE TPTE; similar to transmembrane phosphatase with tensin homology; tensin, putative protein-tyrosine phosphatase [*Homo sapiens*]; SEQ ID NO:99
5) > gi|14787415|/ tyrosine phosphatase isoform A [*Mus musculus*]; SEQ ID NO:100
6) > gi|14787417|/ tyrosine phosphatase isoform B [*Mus musculus*]; SEQ ID NO:101

```
                                    10        20        30        40        50        60
                           ....|....|....|....|....|....|....|....|....|....|....|....|
NOV14                  1   ------------------------------------------------MNSELSLLS 9
gi|7019559|            1   ------------------------------------------------MNESPDPTD 9
```

```
gi|16166555|   1   ----------------------------------------------------MNESPDPTD  9
gi|18640756|   1   ----------------------------------------------------MNESP----  5
gi|14787415|   1   MYGEKKSHLYLWMEHYGYDMPANIYKMYSQPSRKTDDANKKVSVSASRTIKLNGSTGYDT 60
gi|14787417|   1   ------------MHFGNINSTNWF------FRDKKHQNKKVSVSASRTIKLNGSTGYDT 41

                                  70        80        90        100       110       120
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV14              9   -DCFLLS---SASPQTNEFKGATEEAPAKESPHTG----------------EFKGAALV 48
gi|7019559|       10   LAGVIIELGPNDSPQTSEFKGATEEAPAKESPHTS----------------EFKGAARV 52
gi|16166555|      10   LAGVIIELGPNDSPQTSEFKGATEEAPAKESPHTS----------------EFKGAARV 52
gi|18640756|       5   --------------QTNEFKGTTEEAPAKESPHTS----------------EFKGAALV 34
gi|14787415|      61   NEQITLITNGSSLSYPDEIKSASYADPISTKAYTNDSSVYDPGGASSSTTLYELNSLSEV 120
gi|14787417|      42   NEQITLITNGSSLSYPDEIKSASYADPISTKAYTNDSSVYDPGGASSSTTLYELNSLSEV 101

                                  130       140       150       160       170       180
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV14              49   S----------------------------------------------PISKRMLER 58
gi|7019559|        53   S----------------------------------------------PISESVLAR 62
gi|16166555|       53   S----------------------------------------------PISESVLAR 62
gi|18640756|       35   S----------------------------------------------PISK----- 39
gi|14787415|      121   SKEIITQGESALLRDKEATSELKIPSTLQTQTSMSTNTLSLSDLSSDYQEEQMNKCKLNQ 180
gi|14787417|      102   SKEIITQGESALLRDKEATSELKIPSTLQTQTSMSTNTLSLSDLSSDYQEEQMNKCKLNQ 161

                                  190       200       210       220       230       240
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV14              59   LSKFEVGD--AENVASYDSKIKKIVHSIVSSFAVGIFGVFLILLDVALIFADLIFTDSKV 116
gi|7019559|        63   LSKFEVED--AENVASYDSKIKKIVHSIVSSFAFGLFGVFLVLLDVTLIIADLIFTDSKL 120
gi|16166555|       63   LSKFEVED--AENVASYDSKIKKIVHSIVSSFAFGLFGVFLVLLDVTLIIADLIFTDSKL 120
gi|18640756|       39   -------------------------R----------IFGVFLVLLDVTLLLADLIFTDSKL 65
gi|14787415|      181   MSKLYDDDERTDIQKSYWNVVKKFVGILVSSVAFRIFGIFLVILDVFLVVVDLNVSEKKI 240
gi|14787417|      162   MSKLYDDDERTDIQKSYWNVVKKFVGILVSSVAFRIFGIFLVILDVFLVVVDLNVSEKKI 221

                                  250       260       270       280       290       300
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV14             117   YIPLEYRSISLAIALFFLMDVLLRVFVERRQHYFSDLLNILDTAVTVIILLVDVVYIFFD 176
gi|7019559|       121   YIPLEYRSISLAIALFFLMDVLLRVFVERRQQYFSDLFNILDTAIIVILLLVDVVYIFFD 180
gi|16166555|      121   YIPLEYRSISLAIALFFLMDVLLRVFVERRQQYFSDLFNILDTAIIVILLLVDVVYIFFD 180
gi|18640756|       66   YIPLEYRSISLAIGLFFLMDVLLRVFVEG---------------------------- 94
gi|14787415|      241   YIPLDYRSISLAIALFFLVDILLRVSVEGRRRYFSDVLNTLDAVVIGVTVVVAVIYALYD 300
gi|14787417|      222   YIPLDYRSISLAIALFFLVDILLRVSVEGRRRYFSDVLNTLDAVVIGVTVVVAVIYALYD 281

                                  310       320       330       340       350       360
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV14             177   VKFLKDIPRWTRLFRLLRLIILIRVFRLAHLKRQLGKLIRRLVS---RXYERDGFDLDLT 233
gi|7019559|       181   IKLLRNIPRWTHLLRLLRLIILLRIFHLFHQKRQLEKLIRRRVSENKRRYTRDGFDLDLT 240
gi|16166555|      181   IKLLRNIPRWTHLLRLLRLIILLRIFHLFHQKRQLEKLIRRRVSENKRRYTRDGFDLDLT 240
gi|18640756|       94   ----------WTHLVRLLRLIILIRIFHLLHQKRQLEKLMRRLVSENKRRYTRDGFDLDLT 145
gi|14787415|      301   KHFLRDIPRLAVLLRPLRLLILIRILQLAHQKRQLERLTRKLVSGNKRRYKKDGFDLDLT 360
gi|14787417|      282   KHFLRDIPRLAVLLRPLRLLILIRILQLAHQKRQLERLTRKLVSGNKRRYKKDGFDLDLT 341

                                  370       380       390       400       410       420
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV14             234   YITERIVAMSFPSSGGQSFYRNPIKEVVQFLDKKHPNHYRVYNLCSERAYDPKHFHNRVS 293
gi|7019559|       241   YVTERIIAMSFPSSGRQSFYRNPIKEVVRFLDKKHRNHYRVYNLCSERAYDPKHFHNRVV 300
gi|16166555|      241   YVTERIIAMSFPSSGRQSFYRNPIKEVVRFLDKKHRNHYRVYNLCSERAYDPKHFHNRVV 300
gi|18640756|      146   YVTERIIAMSFPSSGRQSFYRNPIEEVVRFLDKKHRNHYRVYNLCSERAYDPKHFHNRVS 205
gi|14787415|      361   YVTERIIAMSFPSSGRESFYRNPIKEVVRFLDTKHPNHYQVYNLCSERAYDPKHFHYRVR 420
gi|14787417|      342   YVTERIIAMSFPSSGRESFYRNPIKEVVRFLDTKHPNHYQVYNLCSERAYDPKHFHYRVR 401

                                  430       440       450       460       470       480
                          ....|....|....|....|....|....|....|....|....|....|....|....|
NOV14             294   RIMIDDHNVPTLREMVAFSKEVLEWMAQDSENIVVIHCKGGKGRTGTMVCACLIASEIFL 353
gi|7019559|       301   RIMIDDHNVPTLHQMVVFTKEVNEWMAQDLENIVAIHCKGGTDRTGTMVCAFLIASEICS 360
gi|16166555|      301   RIMIDDHNVPTLHQMVVFTKEVNEWMAQDLENIVAIHCKGGTDRTGTMVCAFLIASEICS 360
gi|18640756|      206   RIMIDDHNVPTLHEMVVFTKEVNEWMAQDLENIVAIHCKGGKGRTGTMVCALLIASEIFL 265
gi|14787415|      421   RIMIDDHNVPTLEEMLLFSKEVNNWMAQDPENVVAIHCKGGKGRTGTMVCACLIASEIVL 480
gi|14787417|      402   RIMIDDHNVPTLEEMLLFSKEVNNWMAQDPENVVAIHCKGGKGRTGTMVCACLIASEIVL 461

                                  490       500       510       520       530       540
                          ....|....|....|....|....|....|....|....|....|....|....|....|
```



The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 14F lists the domain description from DOMAIN analysis results against NOV14.

Table 14F Domain Analysis of NOV14

Model	Region of Homology	Score (bits)	E value
Dual specificity phosphatase, catalytic doma	231-378	-48.8	3.2

Consistent with other known members of the Putative Protein-Tyrosine Phosphatase-like family of proteins, NOV14 has, for example, a dual specificity protein phosphatase signature sequence and homology to other members of the Putative Protein-Tyrosine Phosphatase-like Protein Family. NOV14 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV14 nucleic acids and polypeptides can be used to identify proteins that are members of the Putative Protein-Tyrosine Phosphatase-like Protein Family. The NOV14 nucleic acids and

polypeptides can also be used to screen for molecules, which inhibit or enhance NOV14 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, *e.g.,* cellular activation, cellular replication, and signal transduction. These molecules can be used to treat, *e.g.*, Cardiovascular diseases, cystitis, incontinence as well as other diseases, disorders and conditions

In addition, various NOV14 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV14 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the Putative Protein-Tyrosine Phosphatase-like Protein Family.

Cellular processes involving growth, differentiation, transformation and metabolism are often regulated in part by protein phosphorylation and dephosphorylation. The protein tyrosine phosphatases (PTPs), which hydrolyze the phosphate monoesters of tyrosine residues, all share a common active site motif and are classified into 3 groups. These include the receptor-like PTPs, the intracellular PTPs, and the dual-specificity PTPs, which can dephosphorylate at serine and threonine residues as well as at tyrosines. Diamond *et al.* (1994) described a PTP from regenerating rat liver that is a member of a fourth class. The gene, which they designated Prl1, was one of many immediate-early genes. Overexpression of Prl1 in stably transfected cells resulted in a transformed phenotype, which suggested that it may play some role in tumorigenesis. By using an *in vitro* prenylation screen, Cates *et al.* (1996) isolated 2 human cDNAs encoding PRL1 homologs, designated PTP(CAAX1) and PTP(CAAX2)(PRL2), that are farnesylated *in vitro* by mammalian farnesyl:protein transferase. Overexpression of these PTPs in epithelial cells caused a transformed phenotype in cultured cells and tumor growth in nude mice. The authors concluded that PTP(CAAX1) and PTP(CAAX2) represent a novel class of isoprenylated, oncogenic PTPs. Peng *et al.* (1998) reported that the human PTP(CAAX1) gene, or PRL1, is composed of 6 exons and contains 2 promoters. The predicted mouse, rat, and human PRL1 proteins are identical. Zeng *et al.* (1998) determined that the human PRL1 and PRL2 proteins share 87% amino acid sequence identity.

The NOV14 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of cardiac and renal physiology. As such, the NOV14 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to

treat cardiovascular and urogenital system disorders, *e.g.,* Cardiovascular diseases, cystitis, incontinence as well as other diseases, disorders and conditions. The NOV14 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV14 nucleic acid is expressed in Urinary bladder.

Additional utilities for NOV14 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV15

A NOV15 polypeptide has been identified as a Leucine Rich Repeat (LRR)-like protein (also referred to as CG95387-02). The disclosed novel NOV15 nucleic acid (SEQ ID NO:31) of 3136 nucleotides is shown in Table 15A. The novel NOV15 nucleic acid sequences maps to the chromosome 19.

An ORF begins with an ATG initiation codon at nucleotides 330-332 and ends with a TAA codon at nucleotides 2331-2333. A putative untranslated region and/or downstream from the termination codon is underlined in Table 15A, and the start and stop codons are in bold letters.

Table 15A. NOV15 Nucleotide Sequence (SEQ ID NO:31)

ACTCCTGACCTAAAGTGATCCACTCGCCTTGGCCTCCCAAAGTGCTAGGATTACAGCCCTCATTCTC
TTTTGCTCCTCAGGTGACACAGGACAAGATCATCTGTCTACCCAATCATGAGCTCCAGGAGAACTTA
TCAGAGGCCCCGTGCCAGCAATTGCTGCCTCGGGGGATCCCTGAGCAGATTGGGGCCCTGCAGGAGG
TTAAAGGCCTTAAGAACAATTTGGACCTGCAGCAATACAGCTTTATTAACCAGCTGTGTTATGAGAC
GGCCCTGCACTGGTATGCCAAGTACTTCCCTTACCTCGTGGTCATTCACACACTCATCTTC**ATG**GTC
TGCACCAGTTTCTGGTTCAAGTTCCCTGGCACCAGCTCCAAGATTGAACACTTCATCTCCATCCTGG
GCAAGTGTTTCGACTCTCCATGGACCACCAGGGCCCTATCCGAGGTCTCCGGGGAGAACCAGAAGGG
CCCAGCAGCCACCGAACGGGCTGCGGCATCCATAGTGGCCATGGCAGGGACCGGGCCGGGGAAGGCA
GGGGAGGGTGAGAAGGAGAAAGTGCTGGCGGAACCGGAGAAGGTGGTGACCGAGCCTCCAGTTGTCA
CCCTGTTGGACAAGAAGGAGGGTGAGCAAGCCAAAGCCCTGTTTGAGAAGGTGAAGAAGTTCCGCAT
GCACGTGGAAGAGGGCGACATCCTGTACACCATGTACATCCGACAGACGGTGCTGAAAGTGTGTAAG
TTCCTGGCCATCCTGGTCTACAACCTGGTCTATGTGGAGAAGATCAGTTTCCTGGTGGCCTGTAGGG
TGGAGACGTCAGAGGTCACGGGCTACGCCAGCTTCTGCTGCAACCACACCAAGGCCCACCTCTTCTC
CAAGCTGGCCTTCTGTTACATCTCCTTTGTGTGCATCTACGGACTTACCTGCATCTACACGCTCTAC
TGGCTCTTCCACCGGCCCCTCAAGGAGTACTCCTTCCGTTCCGTGCGGGAGGAGACTGGCATGGGGG
ACATTCCTGACGTCAAGAATGACTTCGCCTTCATGCTGCACCTCATCGATCAGTACGACTCCCTCTA
CTCCAAGCGCTTCGCCGTCTTCCTGTCCGAGGTCAGCGAAAGCCGTCTAAAGCAGCTCAATCTCAAC
CACGAGTGGACCCCCGAGAAGCTTCGACAGAAGCTGCAGCGCAATGCCGCGGGCCGGCTGGAGCTGG
CCCTCTGCATGCTGCCGGGTCTGCCCGACACCGTCTTTGAGCTCAGTGAGGTGGAGTCACTCAGGCT
GGAGGCCATCTGCGATATCACCTTCCCCCCGGGGCTGTCACAGCTGGTGCACTTGCAGGAGCTCAGC
TTGCTCCACTCGCCCGCCAGGCTACCCTTCTCCTTGCAGGTCTTCCTGCGGGACCACCTGAAGGTGA
TGCGCGTCAAATGCGAGGAGCTCCGCGAGGTGCCGCTTTGGGTGTTTGGGCTGCGGGGCTTGGAGGA
GCTGCACCTGGAGGGGCTTTTCCCCCAGGAGCTAGCTCGGGCAGCCACCCTGGAGAGCCTCCGGGAG
CTGAAGCAGCTCAAGGTGTTGTCCCTCCGGAGCAACGCCGGGAAGGTGCCAGCCAGTGTGACCGACG
TTGCTGGCCACCTGCAGAGGCTCAGCCTGCACAACGATGGGGCCCGTCTGGTTGCCCTGAACAGCCT
CAAGAAGCTGGCGGCATTGCGGGAGCTGGAGCTGGTGGCCTGCGGGCTGGAGCGCATCCCCCATGCA
GTGTTCAGCCTGGGTGCGCTGCAGGAACTTGACCTCAAGGACAACCACCTGCGCTCCATCGAGGAAA

TCCTCAGCTTCCAGCACTGCCGGAAGCTGGTCACGCTCAGGCTGTGGCACAACCAGATCGCCTACGT
CCCTGAGCACGTGCGGAAGCTCAGGAGCCTGGAGCAGCTCTACCTCAGCTACAACAAGCTGGAGACC
CTGCCCTCCCAGCTCGGCCTGTGCTCAGGCCTCCGTCTGCTGGATGTGTCCCACAATGGGCTACACT
CCCTGCCACCCGAGGTGGGCCTCCTGCAGAACCTACAGCACCTGGCCCTCTCCTACAATGCCCTGGA
GGCCCTGCCCGAAGAGCTCTTCTTCTGCCGCAAGCTGCGGACGTTGCTTCTGGGCGACAACCAACTG
AGCCAGCTCTCGCCCCACGTGGGTGCCCTCAGAGCCCTCAGCCGCCTGGAGCTCAAAGGCAACCGCT
TAGAGGCGCTGCCAGAAGAACTTGGCAACTGTGGGGGGCTCAAGAAGGCGGGGCTCCTGGTGGAAGA
CACGCTTTACCAGGGTCTGCCGGCAGAAGTGCGGGACAAGATGGAGGAGGAA**TGA**AGCTGGGGTGGG
GCCGTTTTAGGTAGAGCCTTAAAAATGCTTCTGCCCTGGAATCTCAACCATTATCTTCCAAGATAGG
AAGCCAAGTGGGTCTAGGCCAGGAGATGGGGGGGGGCGGGGGCAGCTGTGTCATCTTTCTGGGGCCC
AGGAGGATCTGGGCTGGTTTGTCTGGGGAGACAGACAGGATGTTGTGGAGGTGGGGTGGAACCTGGT
ATGGAGGGATTAACTCAGTCATGGCATTCTCCGACCAAAACCACACCTGTGTCTCTGGCAGGCTGGC
TGGCCTTGCTCCCATCCCTAGAACTGCTGCCTCTCCCTGGATATTCCAGCTCAATTAGTGCCACATA
TGGGGGAAACGACACATCCCAGTGGGATTTCCAACACTCCCCCTCCCCATGCAACAAAGCAACTTAC
TTCTGGAGTTCTCTCCCAAGGAGAGGACACAGACACAGTTGTTTGCTGTGTTATATGTTAGCTCCGA
ACAATGGTTCTCATTTGGCTAAGCATCAAAATCACCTAGGGAGCCGGTGCAAAACAAAATATCCCAG
TCCCCTCCCCTGAAACACTGACTCAGGAGGTTTGGTTGGGGGCCAGGAGTCTGTTCCTAAATATTCC
AGGTAGTTCTGGTGCAGGTAAGTGGCCCTGAGACAGTATGTTGGGAAATGCTGACGTAAAGGTATCA
GGGCCGGGCGCTGTGGCTCATGACTATAATCCCAGCTGTTTGAGAGGCCAATGCAGGAGGATGGTTG
AGCTCAGGAGTTCGAGATCAGCCTGGGTAACATAGCGAGACCCCACCTCTGCCA

Variant sequences of NOV15 are included in Example 3, Table 26. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA.

The NOV15 protein (SEQ ID NO:32) encoded by SEQ ID NO:31 is 667 amino acid residues in length and is presented using the one-letter amino acid code in Table 15B. Psort analysis predicts the NOV15 protein of the invention to be localized at the plasma membrane with a certainty of 0.7900.

Table 15B. Encoded NOV15 protein sequence (SEQ ID NO:32)

MVCTSFWFKFPGTSSKIEHFISILGKCFDSPWTTRALSEVSGENQKGPAATERAAASIVAMAGTG
PGKAGEGEKEKVLAEPEKVVTEPPVVTLLDKKEGEQAKALFEKVKKFRMHVEEGDILYTMYIRQT
VLKVCKFLAILVYNLVYVEKISFLVACRVETSEVTGYASFCCNHTKAHLFSKLAFCYISFVCIYG
LTCIYTLYWLFHRPLKEYSFRSVREETGMGDIPDVKNDFAFMLHLIDQYDSLYSKRFAVFLSEVS
ESRLKQLNLNHEWTPEKLRQKLQRNAAGRLELALCMLPGLPDTVFELSEVESLRLEAICDITFPP
GLSQLVHLQELSLLHSPARLPFSLQVFLRDHLKVMRVKCEELREVPLWVFGLRGLEELHLEGLFP
QELARAATLESLRELKQLKVLSLRSNAGKVPASVTDVAGHLQRLSLHNDGARLVALNSLKKLAAL
RELELVACGLERIPHAVFSLGALQELDLKDNHLRSIEEILSFQHCRKLVTLRLWHNQIAYVPEHV
RKLRSLEQLYLSYNKLETLPSQLGLCSGLRLLDVSHNGLHSLPPEVGLLQNLQHLALSYNALEAL
PEELFFCRKLRTLLLGDNQLSQLSPHVGALRALSRLELKGNRLEALPEELGNCGGLKKAGLLVED
TLYQGLPAEVRDKMEEE

A search against the Patp database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 15C.

Table 15C. Patp results for NOV15

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAY70473 Human CNAP-1	+1	2147	5.0e-222
>patp:AAG75413 Human colon cancer antigen protein	+1	1882	6.0e-194
>patp:AAM41692 Human polypeptide SEQ ID NO 6623	+1	1878	1.6e-193
>patp:AAU20426 Human secreted protein, Seq ID No 418	+1	1835	5.8e-189
>patp:AAB92855 Human protein sequence SEQ ID NO:11424	+1	1795	1.0e-184

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 2227 of 2228 bases (99%) identical to a gb:GENBANK-ID:AK027073|acc:AK027073.1 mRNA from *Homo sapiens* (cDNA: FLJ23420 fis, clone HEP22352). The full amino acid sequence of the protein of the invention was found to have 444 of 444 amino acid residues (100%) identical to, and 444 of 444 amino acid residues (100%) similar to, the 444 amino acid residue ptnr:SPTREMBL-ACC:Q9H5H8 protein from *Homo sapiens* (CDNA: FLJ23420 FIS, CLONE HEP22352).

NOV15 also has homology to the proteins shown in the BLASTP data in Table 15D.

Table 15D. BLAST results for NOV15

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|13376597|ref|NP_079337.1|(NM_025061)	hypothetical protein FLJ23420 [*Homo sapiens*]	444	444/444 (100%)	444/444 (100%)	0.0
gi|14150009|ref|NP_115646.1|(NM_032270)	hypothetical protein DKFZp586J1119 [*Homo sapiens*]	708	404/667 (60%)	520/667 (77%)	0.0
gi|19343671|gb|AAH25473.1| (BC025473)	Similar to hypothetical protein DKFZp586J1119 [*Mus musculus*]	708	404/671 (60%)	526/671 (78%)	0.0
gi|7243272|dbj|BAA92675.1|(AB037858)	KIAA1437 protein [*Homo sapiens*]	811	367/666 (55%)	84/666 (72%)	0.0
gi|8922442|ref|NP_060573.1|(NM_018103)	hypothetical protein FLJ10470 [*Homo sapiens*]	682	345/673 (51%)	470/673 (69%)	0.0

A multiple sequence alignment is given in Table 15E, with the NOV15 protein being shown on line 1 in Table 15E in a ClustalW analysis, and comparing the NOV15 protein with the related protein sequences shown in Table 15D. This BLASTP data is displayed graphically in the ClustalW in Table 15E.

Table 15E. ClustalW Analysis of NOV15

1) > NOV15; SEQ ID NO:32
2) > gi|13376597|/ hypothetical protein FLJ23420 [*Homo sapiens*]; SEQ ID NO:102
3) > gi|14150009|/ hypothetical protein DKFZp586J1119 [*Homo sapiens*]; SEQ ID NO:103
4) > gi|19343671|/ Similar to hypothetical protein DKFZp586J1119 [*Mus musculus*]; SEQ ID NO:104
5) > gi|7243272|/ KIAA1437 protein [*Homo sapiens*]; SEQ ID NO:105
6) > gi|8922442|/ hypothetical protein FLJ10470 [*Homo sapiens*]; SEQ ID NO:106

```
gi|14150009|   379 LHQCSVKIHSAALSFLKENLKVLSVKFDDMRELPPWMYGLRNLEELYLVGSLSHDISRNV 438
gi|19343671|   379 LHQCSVKIHSAALSFLKENLKVLSVKFDDMRELPPWMYGLRNLEELYLVGSLSHDISKNV 438
gi|7243272|    477 LYHTAAKIEAPALAFLRENLRALHIKFTDIKEIPLWIYSLKTLEELHLTGNLSAENNRYI 536
gi|8922442|    344 LCHCPAKVEQTAFSFLRDHLRCLHVKFTDVAEIPAWVYLLKNLRELYLIGNLNSENNKMI 403

                          550       560       570       580       590       600
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|
NOV15              398 TLESLRELKQLKVLSLRSNAGKVPASVTDVAGHLQRLSLHNDGARLVALNSLKKLAALRE 457
gi|13376597|       175 TLESLRELKQLKVLSLRSNAGKVPASVTDVAGHLQRLSLHNDGARLVALNSLKKLAALRE 234
gi|14150009|       439 TLESLRDLKSLKILSIKSNVSKIPQAVVDVSSHLQKMCIHNDGTKLVMLNNLKKMTNLTE 498
gi|19343671|       439 TLESLRDLKSLKILSIKSNVSKIPQAVVDVSSHLQKMCVHNDGTKLVMLNNLKKMTNLTE 498
gi|7243272|        537 VIDGLRELKRLKVLRLKSNLSKLPQVVTDVGVHLQKLSINNEGTKLIVLNSLKKMANLTE 596
gi|8922442|        404 GLESLRELRHLKILHVKSNLTKVPSNITDVAPHLTKLVIHNDGTKLLVLNSLKKMMNVAE 463

                          610       620       630       640       650       660
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|
NOV15              458 LELVACGLERIPHAVFSLGALQELDLKDNHLRSIEEILSFQHCRKLVTLRLWHNQIAYVP 517
gi|13376597|       235 LELVACGLERIPHAVFSLGALQELDLKDNHLRSIEEILSFQHCRKLVTLRLWHNQIAYVP 294
gi|14150009|       499 LELVHCDLERIPHAVFSLLSLQELDLKENNLKSIEEIVSFQHLRKLTVLKLWHNSITYIP 558
gi|19343671|       499 LELVHCDLERIPHAVFSLLSLQELDLKENNLKSIEEIVSFQHLRKLTVLKLWYNSIAYIP 558
gi|7243272|        597 LELIRCDLERIPHSIFSLHNLQEIDLKDNNLKTIEEIISFQHLHRLTCLKLWYNHIAYIP 656
gi|8922442|        464 LELQNCELERIPHAIFSLSNLQELDLKSNNIRTIEEIISFQHLKRLTCLKLWHNKIVTIP 523

                          670       680       690       700       710       720
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|
NOV15              518 EHVRKLRSLEQLYLSYNKLETLPSQLGLCSGLRLLDVSHNGLHSLPPEVGLLQNLQHLAL 577
gi|13376597|       295 EHVRKLRSLEQLYLSYNKLETLPSQLGLCSGLRLLDVSHNGLHSLPPEVGLLQNLQHLAL 354
gi|14150009|       559 EHIKKLTSLERLSFSHNKIEVLPSHLFLCNKIRYLDLSYNDIRFIPPEIGVLQSLQYFSI 618
gi|19343671|       559 EHIKKLTSLERLFFSHNKVEVLPSHLFLCNKIRYLDLSYNDIRFIPPEIGVLQSLQYFSI 618
gi|7243272|        657 IQIGNLTNLERLYLNRNKIEKIPTQLFYCRKLRYLDLSHNNLTFLPADIGLLQNLQNLAI 716
gi|8922442|        524 PSITHVKNLESLYFSNNKLESLPVAVFSLQKLRCLDVSYNNISMIPIEIGLLQNLQHLHI 583

                          730       740       750       760       770       780
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|
NOV15              578 SYNALEALPEELFFCRKLRTLLLGDNQLSQLSPHVGALRALSRLELKGNRLEALPEELGN 637
gi|13376597|       355 SYNALEALPEELFFCRKLRTLLLGDNQLSQLSPHVGALRALSRLELKGNRLEALPEELGN 414
gi|14150009|       619 TCNKVESLPDELYFCKKLKTLKIGKNSLSVLSPKIGNLLFLSYLDVKGNHFEILPPELGD 678
gi|19343671|       619 TCNKVESLPDELYFCKKLKTLKIGKNSLSVLSPKIGNLLFLSYLDIKGNHFEVLPPELGD 678
gi|7243272|        717 TANRIETLPPELFQCRKLRALHLGNNVLQSLPSRVGELTNLTQIELRGNRLECLPVELGE 776
gi|8922442|        584 TGNKVDILPKQLFKCIKLRTLNLGQNCITSLPEKVGQLSQLTQLELKGNCLDRLPAQLGQ 643

                          790       800       810
                   ....|....|....|....|....|....|....|....
NOV15              638 CGGLKKAGLLVEDTLYQGLPAEVRDKMEEE--------- 667
gi|13376597|       415 CGGLKKAGLLVEDTLYQGLPAEVRDKMEEE--------- 444
gi|14150009|       679 CRALKRAGLVVEDALFETLPSDVREQMKTE--------- 708
gi|19343671|       679 CRALKRARLVVEDALFETLPSDVREQMKAD--------- 708
gi|7243272|        777 CPLLKRSGLVVEEDLFNTLPPEVKERLWRADKEQA---- 811
gi|8922442|        644 CRMLKKSGLVVEDHLFDTLPLEVKEALNQDINIPFANGI 682
```

The NOV15 Clustal W alignment shown in Table 15E was modified to begin at amino residue 121. The data in Table 15E includes all of the regions overlapping with the NOV15 protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 15F lists the domain description from DOMAIN analysis results against NOV15.

Table 15F Domain Analysis of NOV15			
Model	Region of Homology	Score (bits)	E value
LRR	454-476	5.6	1.5e+02
LRR	477-498	10.8	26
LRR	502-524	11.9	16
LRR	525-547	18.8	0.13
LRR	548-570	15.9	0.99
LRR	571-593	13.9	4
LRR	594-616	12.3	12
LRR	617-639	9.4	42

Consistent with other known members of the LRR-like family of proteins, NOV15 has, for example, eight Leucine Rich Repeat (LRR) signature sequences and homology to other members of the LRR-like Protein Family. NOV15 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV15 nucleic acids and polypeptides can be used to identify proteins that are members of the LRR-like Protein Family. The NOV15 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV15 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, *e.g.,* cellular activation, cellular replication, and signal transduction. These molecules can be used to treat, *e.g.*, cancer, trauma, regeneration (*in vitro* and *in vivo*), viral/bacterial/parasitic infections, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation, Hirschsprung's disease , Crohn's Disease, appendicitis, osteoporosis, hypercalceimia, arthritis, ankylosing spondylitis, scoliosis, systemic lupus erythematosus, autoimmune disease, xerostomia as well as other diseases, disorders and conditions.

In addition, various NOV15 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV15 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the LRR-like Protein Family.

LRR Proteins are a family of proteins characterized by a structural motif rich in leucine residues. They are either transmembrane or secreted proteins and are involved in protein-protein interactions. Members of this family have been implicated in extracellular matrix assembly and cellular growth. In addition, several proteins belonging to this family, such as slit, Toll and robo have been shown to mediate key roles in central nervous system

development and organogenesis in *Drosophila*. Vertebrate orthologs of these proteins have also been shown to have similar roles in the CNS as well as other organ systems like kidney..

LRRs are relatively short motifs (22-28 residues in length) found in a variety of cytoplasmic, membrane and extracellular proteins. Although these proteins are associated with widely different functions, a common property involves protein-protein interaction. Little is known about the 3D structure of LRRs, although it is believed that they can form amphipathic structures with hydrophobic surfaces capable of interacting with membranes. *In vitro* studies of a synthetic LRR from *Drosophila* Toll protein have indicated that the peptides form gels by adopting beta-sheet structures that form extended filaments (Packman *et al.* FEBS Lett.1991; 291: 87-91). These results are consistent with the idea that LRRs mediate protein-protein interactions and cellular adhesion. Other functions of LRR-containing proteins include, for example, binding to enzymes and vascular repair.

The NOV15 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of cardiac, immune, and nerve physiology. As such, the NOV15 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to treat cardiovascular, nervous, and immune system disorders, *e.g.,* cancer, trauma, regeneration (*in vitro* and *in vivo*), viral/bacterial/parasitic infections, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation, Hirschsprung's disease, Crohn's Disease, appendicitis, osteoporosis, hypercalceimia, arthritis, ankylosing spondylitis, scoliosis, systemic lupus erythematosus, autoimmune disease, xerostomia as well as other diseases, disorders and conditions.

The NOV15 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV15 nucleic acid is expressed in Coronary Artery, Parotid Salivary glands, Liver, Colon, Bone, Synovium/Synovial membrane, and Brain.

Additional utilities for NOV15 nucleic acids and polypeptides according to the invention are disclosed herein.

NOV16

A NOV16 polypeptide has been identified as a RhoGEF-like protein (also referred to as CG95419-02). The disclosed novel NOV16 nucleic acid (SEQ ID NO:33) of 5372 nucleotides is shown in Table 16A. The novel NOV16 nucleic acid sequences maps to the chromosome 5.

An ORF begins with an ATG initiation codon at nucleotides 61-63 and ends with a TAA codon at nucleotides 5179-5181. A putative untranslated region and/or downstream from the termination codon is underlined in Table 16A, and the start and stop codons are in bold letters.

Table 16A. NOV16 Nucleotide Sequence (SEQ ID NO:33)

CCATGGGGCCTCCTGCAATAACTTCTCTTGTTTATTATTTTCATTGCAGATGCGAAAGCC**ATG**GAGT
TGAGCTGCAGCGAAGCACCTCTTTACCAGGGGCAGATGATGATCTATGCGAAGTTTGACAAAAATGT
GTATCTTCCTGAAGATGCTGAGTTTTACTTTACTTATGACGGATCTCATCAGCGACATGTCATGATT
GCAGAGCGCATCGAGGATAACGTTCTCCAGTCCAGCGTCCCAGGCCATGGGCTTCAGGAGACGGTGA
CGGTATCTGTGTGCCTCTGCTCGGAAGGTTACTCTCCGGTGACCATGGGCTCTGGCTCAGTGACCTA
CGTGGACAACATGGCTTGCAGGCTGGCTCGTCTGCTGGTGACGCAGGCCAATCGCCTCACAGCCTGC
AGCCACCAGACCCTGCTGACCCCATTTGCCTTGACGGCAGGAGCACTGCCTGCCTTGGATGAGGAGC
TCGTGCTGGCTCTGACCCATCTGGAATTGCCTCTAGAGTGGACTGTGTTGGGAAGTTCTTCACTTGA
AGTATCTTCTCACAGAGAATCTCTTCTACACCTGGCTATGAGATGGGGCCTGGCTAAACTTTCCCAG
TTCTTCTTGTGTCTCCCGGGGGGAGTCCAGGCCTTGGCTTTACCCAACGAAGAGGGTGCCACACCAT
TAGACTTAGCTTTACGTGAAGGACACTCCAAGCTGGTGGAAGACGTCACAAGTTTTCAGGGCAGATG
GTCCCCAAGCTTCTCCCGAGTGCAGCTCAGTGAAGAAGCCTCCTTGCATTACATTCACTCATCGGAA
ACGCTGACCCTGACCCTGAACCACACAGCCGAGCATTTGTTGGAGGCAGATATTAAACTCTTCCGGA
AATACTTTTGGGATAGAGCCTTTCTTGTCAAGGCCTTTGAGCAAGAAGCCAGGCCAGAGGAAAGAAC
AGCTATGCCCTCCAGCGGTGCAGAAACTGAAGAAGAGATTAAGAATTCAGTGTCCAGCAGATCAGCA
GCCGAAAAGGAAGATATAAAGCGTGTCAAAAGCCTGGTGGTTCAACACAATGAACATGAAGACCAGC
ACAGCCTAGATTCTAGATCGCTCCTTCGATATCCTAAAAAATCCAAGCCGCCCTCGACATTGCTTGC
TGCAGGCCGGCTTTCAGACATGCTGAATGGAGGTGATGAAGTCTACGCTAACTGTATGGTGATTGAT
CAGGTTGGTGATTTGGATATCAGCTATATTAATATAGAGGGAATCACTGCCACTACCAGCCCTGAAT
CCAGAGGTTGCACTCTGTGGCCTCAGAGCAGCAAACACACCCTTCCTACAGAAACCAGTCCCAGTGT
GTACCCACTTAGTGAAAATGTCGAAGGGACAGCACACACTGAAGCCCAGCAGTCCTTCATGTCACCA
TCAAGTTCGTGTGCTTCCAACTTGAATCTTTCTTTTGGTTGGCATGGATTTGAAAAGGAACAAAGTC
ATCTAAAGAAAAGAAGTTCTAGCCTTGATGCCTTGGACGCCGACAGTGAAGGGGAAGGGCATTCTGA
GCCATCCCACATCTGTTACACTCCAGGGTCTCAGAGCTCCTCAAGAACTGGGATTCCTAGTGGGGAT
GAATTGGACTCTTTTGAGACTAACACTGAACCGGATTTTAATATCTCCAGGGCTGAATCCCTTCCTC
TATCAAGTAATCTACAGTTGAAGGAATCACTGCTTTCTGGAGTTCGCTCACGTTCTTATTCTTGCTC
GTCACCCAAAATTTCTTTAGGAAAAACTCGTTTGGTGCGTGAATTAACAGTATGCAGTTCAAGTGAA
GAGCAAAAAGCTTACAGCTTATCGGAGCCACCAAGAGAAAACAGGATTCAGGAAGAAGAATGGGATA
AATACATCATACCTGCCAAATCAGAGTCTGAAAAATATAAAGTGAGTCGAACTTTCAGTTTCCTCAT
GAATAGGATGACTAGCCCTCGGAATAAATCAAAGACAAAAAGCAAGGATGCCAAAGATAAAGAGAAG
CTGAATCGACATCAGTTTGCCCCAGGAACATTCTCTGGGGTTCTGCAGTGTTTGGTTTGTGATAAAA
CACTCCTGGGGAAAGAGTCACTGCAGTGTTCTAGTTGTAATGCAAATGTGCACAAAGGTTGTAAAGA
TGCTGCGCCTGCATGCACCAAGAAATTCCAAGAGAAATATAACAAGAACAAACCACAGACCATCCTT
GGAAGTTCTTCATTTAGAGACATCCCACAGCCTGGTCTCTCCTTGCACCCTTCTTCCTCCGTGCCTG
TTGGATTGCCGACTGGAAGGAGGGAGACTGTGGGACAGGTCCATCCATTGTCCAGAAGTGTTCCAGG
TACCACCTTGGAAAGCTTCAGGAGGTCAGCCACATCCTTGGAGTCTGAGAGTGACAATAACAGCTGC
AGAAGCAGGTCTCATTCTGATGAGCTGCTACAGTCCATGGGCTCTTCTCCCTCTACAGAGTCTTTCA
TAATGGAAGATGTTGTGGATTCTTCTCTGTGGAGTGACCTCAGCAGTGATGCCCAGGAGTTTGAAGC
AGAATCTTGGAGTCTTGTGGTGGATCCCTCATTTTGTAATAGGCAGGAGAAGGATGTCATCAAAAGA

CAGGATGTCATTTTTGAGCTAATGCAAACAGAGATGCATCACATCCAGACCCTGTTCATCATGTCTG
AGATCTTCAGGAAAGGCATGAAAGAGGAGCTGCAGCTGGACCACAGCACCGTGGATAAAATTTTCCC
CTGTTTAGATGAGTTGCTTGAAATCCACAGGCATTTCTTCTACAGTATGAAGGAACGAAGGCAGGAA
TCAAGTGCTGGCAGCGACAGGAATTTTGTGATCGACCGAATTGGAGATATTTTGGTACAACAGTTTT
CAGAAGAAAATGCAAGTAAAATGAAGAAAATATATGGAGAATTCTGTTGCCATCATAAAGAAGCTGT
TAACCTCTTTAAAGAACTCCAGCAGAATAAAAAGTTTCAGAATTTTATTAAGCTCCGAAATAGTAAT
CTTTTGGCTCGACGCCGAGGAATTCCAGAATGCATTCTGTTGGTCACTCAGCGTATTACAAAATACC
CTGTCTTGGTGGAAAGGATATTGCAGTACACAAAGGAAAGAACTGAGGAACATAAAGACTTACGCAA
AGCCCTTTGCTTAATTAAAGACATGATTGCAACAGTGGATTTAAAAGTCAATGAATATGAGAAAAAC
CAAAAATGGCTTGAGATCCTAAATAAGATTGAAAACAAAACATACACGAAGCTCAAAAATGGACATG
TGTTTAGGAAGCAGGCACTGATGAGTGAAGAAAGGACTCTGTTATATGATGGCCTTGTTTACTGGAA
AACTGCTACAGGTCGTTTCAAAGATATCCTAGCTCTACTTCTAACTGATGTGCTGCTCTTTTTACAA
GAAAAAGACCAGAAATACATCTTTGCAGCCGTTGATCAGAAGCCATCAGTTATTTCCCTTCAAAAGC
TTATTGCTAGAGAAGTTGCTAATGAGGAGAGAGGAATGTTTCTGATCAGTGCTTCATCTGCTGGTCC
TGAGATGTATGAAATTCACACCAATTCCAAGGAGGAACGCAATAACTGGATGAGACGGATCCAGCAG
GCTGTAGAAAGTTGTCCTGAAGAAAAAGGGGGAAGGACAAGTGAATCTGATGAAGACAAGAGGAAAG
CTGAAGCCAGAGTGGCCAAAATTCAGCAATGTCAAGAAATACTCACTAACCAAGACCAACAAATTTG
TGCGTATTTGGAGGAGAAGCTGCATATCTATGCTGAACTTGGAGAACTGAGCGGATTTGAGGACGTC
CATCTAGAGCCCCACCTCCTTATTAAACCTGACCCAGGCGAGCCTCCCCAGGCAGCCTCATTACTGG
CAGCAGCACTGAAAGAAGCTGAGAGCCTACAAGTTGCAGTGAAGGCCTCACAGATGGGCGCCGTGAG
TCAATCATGTGAGGACAGTTGTGGAGACTCTGTCTTGGCGGACACACTCAGTTCTCATGATGTACCA
GGATCACCGACTGCCTCATTAGTCACAGGAGGGAGAGAAGGAAGAGGCTGTTCGGATGTGGATCCCG
GGATCCAGGGTGTGGTAACCGACTTGGCCGTCTCTGATGCAGGGGAGAAGGTGGAATGTAGAAATTT
TCCAGGTTCTTCACAATCAGAGATTATACAAGCCATACAGAATTTAACCCGTCTCTTATACAGCCTT
CAGGCCGCCTTGACCATTCAGGACAGCCACATTGAGATCCACAGGCTGGTTCTCCAGCAGCAGGAGG
GCCTGTCTCTCGGCCACTCTATCCTCCGAGGCGGCCCCTTGCAGGACCAGAAGTCTCGCGACGCGGA
CAGGCAGCATGAGGAGCTGGCCAATGTGCACCAGCTTCAGCACCAGCTCCAGCAGGAGCAGCGGCGC
TGGCTGCGCAGGTGTGAGCAGCAGCAGCGGGCGCAGGCGACCAGGGAGAGCTGGCTGCAGGAGCGGG
AGCGGGAGTGCCAGTCGCAGGAGGAGCTGCTGCTGCGGAGCCGGGGCGAGCTGGACCTCCAGCTCCA
GGAGTACCAGCACAGCCTGGAGCGGCTGAGGGAGGGCCAGCGCCTGGTGGAGAGGGAGCAGGCGAGG
ATGCGGGCCCAGCAGAGCCTGCTGGGCCACTGGAAGCACGGCCGGCAGAGGAGCCTGCCCGCGGTGC
TCCTTCCGGGTGGCCCCGAGGTAATGGAACTTAATCGATCTGAGAGTTTATGTCATGAAAACTCATT
CTTCATCAATGAAGCTTTAGTACAAATGTCATTTAACACTTTCAACAAACTGAATCCGTCAGTTATC
CATCAGGATGCCACTTACCCTACAACTCAATCTCATTCTGACTTGGTGAGGACTAGTGAACATCAAG
TAGACCTCAAGGTGGACCCTTCTCAGCCTTCGAATGTCAGTCACAAACTGTGGACAGCCGCTGGTTC
CGGCCATCAGATACTTCCTTTCCAAGAAAGCAGCAAGGATTCTTGTAAAAATGATTTGGACACCTCC
CACACTGAGTCCCCAACCCCCCATGACTCAAATTCACACCGCCCTCAACTGCAGGCGTTTATAACAG
AAGCAAAGCTAAATCTACCGACAAGGACAATGACCAGACAAGATGGGGAAACTGGAGATGGAGCCAA
AGAAAATATTGTTTACCTC**TAA**TTGTGTTGTCATTTTTCCAAACAAAACAAAACACTGGCACTTTTG
GGAGAAACTTTTTGTCTCCATTCCTTATGTATGTGTGATTGTCTGTGTCCAAATTGCTTTAAGAATA
ATATTTAATATTTCCTGGAAGCTCATTTTTTTGGCATGAGTCTAATTAAATTATTGAAAGCCAAAAA
AAAAAAAAAAAA

The NOV16 protein (SEQ ID NO:34) encoded by SEQ ID NO:33 is 1706 amino acid residues in length and is presented using the one-letter amino acid code in Table 16B. Psort analysis predicts the NOV16 protein of the invention to be localized in the cytoplasm with a certainty of 0.4500.

Table 16B. Encoded NOV16 protein sequence (SEQ ID NO:34)

MELSCSEAPLYQGQMMIYAKFDKNVYLPEDAEFYFTYDGSHQRHVMIAERIEDNVLQSSVPGHGL
QETVTVSVCLCSEGYSPVTMGSGSVTYVDNMACRLARLLVTQANRLTACSHQTLLTPFALTAGAL
PALDEELVLALTHLELPLEWTVLGSSSLEVSSHRESLLHLAMRWGLAKLSQFFLCLPGGVQALAL
PNEEGATPLDLALREGHSKLVEDVTSFQGRWSPSFSRVQLSEEASLHYIHSSETLTLTLNHTAEH
LLEADIKLFRKYFWDRAFLVKAFEQEARPEERTAMPSSGAETEEEIKNSVSSRSAAEKEDIKRVK
SLVVQHNEHEDQHSLDSRSLLRYPKKSKPPSTLLAAGRLSDMLNGGDEVYANCMVIDQVGDLDIS

YINIEGITATTSPESRGCTLWPQSSKHTLPTETSPSVYPLSENVEGTAHTEAQQSFMSPSSSCAS
NLNLSFGWHGFEKEQSHLKKRSSSLDALDADSEGEGHSEPSHICYTPGSQSSSRTGIPSGDELDS
FETNTEPDFNISRAESLPLSSNLQLKESLLSGVRSRSYSCSSPKISLGKTRLVRELTVCSSSEEQ
KAYSLSEPPRENRIQEEEWDKYIIPAKSESEKYKVSRTFSFLMNRMTSPRNKSKTKSKDAKDKEK
LNRHQFAPGTFSGVLQCLVCDKTLLGKESLQCSSCNANVHKGCKDAAPACTKKFQEKYNKNKPQT
ILGSSSFRDIPQPGLSLHPSSSVPVGLPTGRRETVGQVHPLSRSVPGTTLESFRRSATSLESESD
NNSCRSRSHSDELLQSMGSSPSTESFIMEDVVDSSLWSDLSSDAQEFEAESWSLVVDPSFCNRQE
KDVIKRQDVIFELMQTEMHHIQTLFIMSEIFRKGMKEELQLDHSTVDKIFPCLDELLEIHRHFFY
SMKERRQESSAGSDRNFVIDRIGDILVQQFSEENASKMKKIYGEFCCHHKEAVNLFKELQQNKKF
QNFIKLRNSNLLARRRGIPECILLVTQRITKYPVLVERILQYTKERTEEHKDLRKALCLIKDMIA
TVDLKVNEYEKNQKWLEILNKIENKTYTKLKNGHVFRKQALMSEERTLLYDGLVYWKTATGRFKD
ILALLLTDVLLFLQEKDQKYIFAAVDQKPSVISLQKLIAREVANEERGMFLISASSAGPEMYEIH
TNSKEERNNWMRRIQQAVESCPEEKGGRTSESDEDKRKAEARVAKIQQCQEILTNQDQQICAYLE
EKLHIYAELGELSGFEDVHLEPHLLIKPDPGEPPQAASLLAAALKEAESLQVAVKASQMGAVSQS
CEDSCGDSVLADTLSSHDVPGSPTASLVTGGREGRGCSDVDPGIQGVVTDLAVSDAGEKVECRNF
PGSSQSEIIQAIQNLTRLLYSLQAALTIQDSHIEIHRLVLQQQEGLSLGHSILRGGPLQDQKSRD
ADRQHEELANVHQLQHQLQQEQRRWLRRCEQQQRAQATRESWLQERERECQSQEELLLRSRGELD
LQLQEYQHSLERLREGQRLVEREQARMRAQQSLLGHWKHGRQRSLPAVLLPGGPEVMELNRSESL
CHENSFFINEALVQMSFNTFNKLNPSVIHQDATYPTTQSHSDLVRTSEHQVDLKVDPSQPSNVSH
KLWTAAGSGHQILPFQESSKDSCKNDLDTSHTESPTPHDSNSHRPQLQAFITEAKLNLPTRTMTR
QDGETGDGAKENIVYL

A search against the Patp database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 16C.

Table 16C. Patp results for NOV16

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:ABB44551 Human wound healing related polypeptide	+1	1470	2.8e-150
>patp:AAW93941 Human brx protein	+1	1436	1.5e-148
>patp:ABG05537 Novel human diagnostic protein #5528	+1	1436	1.5e-148
>patp:ABG05537 Novel human diagnostic protein #5528	+1	1436	1.5e-148
>patp:ABG15870 Novel human diagnostic protein #15861	+1	1447	7.5e-148

In a BLAST search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 4339 of 5274 bases (82%) identical to a gb:GENBANK-ID:MMU73199|acc:U73199.1 mRNA from *Mus musculus* (Rho-guanine nucleotide exchange factor mRNA, complete cds). The full amino acid sequence of the protein of the invention was found to have 1350 of 1670 amino acid residues (80%) identical to, and 1460 of 1670 amino acid residues (87%) similar to, the 1693 amino acid residue ptnr:SWISSPROT-ACC:P97433 protein from *Mus musculus* (RHO-GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHOGEF) (RIP2)).

NOV16 also has homology to the proteins shown in the BLASTP data in Table 16D.

Table 16D. BLAST results for NOV16

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|7106395|ref|NP_036156.1|(NM_012026)	Rho interacting protein 2; Rho specific exchange factor [*Mus musculus*]	1693	1350/1674 (80%)	1460/1674 (86%)	0.0
gi|18602674|ref|XP_016989.5|(XM_016989)	hypothetical protein FLJ21817 similar to Rhoip2 [*Homo sapiens*]	669	668/669 (99%)	668/669 (99%)	0.0
gi|10438441|dbj|BAB15243.1|(AK025816)	unnamed protein product [*Homo sapiens*]	669	669/669 (100%)	669/669 (100%)	0.0
gi|15341761|gb|AAH12946.1|AAH12946(BC012946)	hypothetical protein FLJ21817 similar to Rhoip2 [*Homo sapiens*]	615	609/613 (99%)	609/613 (99%)	0.0
gi|17437752|ref|XP_068710.1|(XM_068710)	similar to Rho interacting protein 2; Rho specific exchange factor [*Homo sapiens*]	590	290/292 (99%)	291/292 (99%)	e-170

A multiple sequence alignment is given in Table 16E, with the NOV16 protein being shown on line 1 in Table 16E in a ClustalW analysis, and comparing the NOV16 protein with the related protein sequences shown in Table 16D. This BLASTP data is displayed graphically in the ClustalW in Table 16E.

Table 16E. ClustalW Analysis of NOV16

1) > NOV16; SEQ ID NO:34
2) > gi|7106395|/ Rho interacting protein 2; Rho specific exchange factor [*Mus musculus*]; SEQ ID NO:107
3) > gi|18602674|/ hypothetical protein FLJ21817 similar to Rhoip2 [*Homo sapiens*]; SEQ ID NO:108
4) > gi|10438441|/ unnamed protein product [*Homo sapiens*]; SEQ ID NO:109
5) > gi|15341761|/ hypothetical protein FLJ21817 similar to Rhoip2 [*Homo sapiens*]; SEQ ID NO:110
6) > gi|17437752|/ similar to Rho interacting protein 2; Rho specific exchange factor [*Homo sapiens*]; SEQ ID NO:111

```
                                    10        20        30        40        50        60
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16                 1    MELSCSEAPLYQGQMMIYAKFDKNVYLPEDAEFYFTYDGSHQRHVMIAERIEDNVLQSSV 60
gi|7106395|           1    -MELSCSEVPLYGQKTVYAKFGKNVYLPEDAEFYFVYGGSHQRHVVIADRVQDNVLQSSI 59
gi|18602674|          1    ------------------------------------------------------------ 1
gi|10438441|          1    ------------------------------------------------------------ 1
gi|15341761|          1    ------------------------------------------------------------ 1
gi|17437752|          1    ------------------------------------------------------------ 1

                                        70        80        90       100       110       120
                                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16                61    PGHGLQETVTVSVCLCSEGYSPVTMGSGSVTYVDNMACRLARLLVTQANRLTACSHQTLL 120
gi|7106395|          60    PGHWLQETVTVSVCLCSEGYSPVTMGSGSVTYVDNMACRLARLLVTQADRLTACSHQTLL 119
gi|18602674|         1    ------------------------------------------------------------ 1
gi|10438441|         1    ------------------------------------------------------------ 1
gi|15341761|         1    ------------------------------------------------------------ 1
gi|17437752|         1    ------------------------------------------------------------ 1
```

```
                                130       140       150       160       170       180
                           ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              121  TPFALTAGALPALDEELVLALTHLELPLEWTVLGSSSLEVSSHRESLLHLAMRWGLAKLS 180
gi|7106395|        120  TPFALTVEALPALDEELVLALTQLELPLGWTVLGNSSLEVSLHRESLLHLAVRWALPKLF 179
gi|18602674|       1    ------------------------------------------------------------ 1
gi|10438441|       1    ------------------------------------------------------------ 1
gi|15341761|       1    ------------------------------------------------------------ 1
gi|17437752|       1    ------------------------------------------------------------ 1

                                190       200       210       220       230       240
                           ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              181  QFFLCLPGGVQALALPNEEGATPLDLALREGHSKLVEDVTSFQGRWSPSFSRVQLSEEAS 240
gi|7106395|        180  HFLLCLPGGVKALKLPNEEATTPLDLALQGGHSTLVEDITNFQGSHSPGFSRLRLNEEAT 239
gi|18602674|       1    ------------------------------------------------------------ 1
gi|10438441|       1    ------------------------------------------------------------ 1
gi|15341761|       1    ------------------------------------------------------------ 1
gi|17437752|       1    ------------------------------------------------------------ 1

                                250       260       270       280       290       300
                           ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              241  LHYIHSSETLTLTLNHTAEHLLEADIKLFRKYFWDRAFLVKAFEQEARPEERTAMPSSGA 300
gi|7106395|        240  LQFVHSSETLTLTVNHTAEHLLEADIKLFRKYFWDRAFLVKALEQEAKTEKATMPSGAAE 299
gi|18602674|       1    ------------------------------------------------------------ 1
gi|10438441|       1    ------------------------------------------------------------ 1
gi|15341761|       1    ------------------------------------------------------------ 1
gi|17437752|       1    ------------------------------------------------------------ 1

                                310       320       330       340       350       360
                           ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              301  ETEEEIKNSVSSRSAAEKEDIKRVKSLVVQHNEHEDQHSLDSRSLLRYPKKSKPPSTLLA 360
gi|7106395|        300  TEEEVRNLESGRSPSEEEEDAKSIKSQVDGPSEHEDQDRLALDRSFDGLKKSKHVPASLA 359
gi|18602674|       1    ------------------------------------------------------------ 1
gi|10438441|       1    ------------------------------------------------------------ 1
gi|15341761|       1    ------------------------------------------------------------ 1
gi|17437752|       1    ------------------------------------------------------------ 1

                                370       380       390       400       410       420
                           ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              361  AGRLSDMLNGGDEVYANCMVIDQVGDLDISYINIEGITATTSPESRGCTLWPQSSKHTLP 420
gi|7106395|        360  AGQLSDVLNGGDEVYANCMVIDQVGDLDINYYINLEGLSTHTSPESGRSMLGPQACMHTLP 419
gi|18602674|       1    ------------------------------------------------------------ 1
gi|10438441|       1    ------------------------------------------------------------ 1
gi|15341761|       1    ------------------------------------------------------------ 1
gi|17437752|       1    ------------------------------------------------------------ 1

                                430       440       450       460       470       480
                           ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              421  TETSPSVYPLSENVEGTAHTEAQQSFMSPSSSCASNLNLSFGWHGFEKEQSHLKKRSSSL 480
gi|7106395|        420  PDTSPCGRPLIENSEGTLDAAASQSFVTPSSSRTSNLNLSFGLHGFEKEQSHLKKRSSSL 479
gi|18602674|       1    ------------------------------------------------------------ 1
gi|10438441|       1    ------------------------------------------------------------ 1
gi|15341761|       1    ------------------------------------------------------------ 1
gi|17437752|       1    ------------------------------------------------------------ 1

                                490       500       510       520       530       540
                           ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              481  DALDADSEGEGHSEPSHICYTPGSQSSSRTGIPSGDELDSFETNTEPDFNISRAESLPLS 540
gi|7106395|        480  DALVADSEGEGGSEPPICYAVGSQSSPRTG-LPSGDELDSFETNTEPDCNISRTESLSLS 538
gi|18602674|       1    ------------------------------------------------------------ 1
gi|10438441|       1    ------------------------------------------------------------ 1
gi|15341761|       1    ------------------------------------------------------------ 1
gi|17437752|       1    ------------------------------------------------------------ 1

                                550       560       570       580       590       600
                           ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              541  SNLQLKESLLSGVRSRSYSCSSPKISLGKTRLVRELTVCSSSEEQKAYSLSEPPRENRIQ 600
gi|7106395|        539  STLHSKESLLSGIRSRSYSCSSPKISSGKSRLVRDFTVCSTSEEQRSYSFQEPPGEKRIQ 598
gi|18602674|       1    ------------------------------------------------------------ 1
gi|10438441|       1    ------------------------------------------------------------ 1
gi|15341761|       1    ------------------------------------------------------------ 1
```

```
gi|17437752|   1    ------------------------------------------------------------ 1

                          610       620       630       640       650       660
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              601   EEEWDKYIIPAKSESEKYKVSRTFSFLMNRMTSPRNKSKTKSKDAKDKEKLNRHQFAPGT 660
gi|7106395|        599   EEEWDEYVIPAKSESEKYKVSRTFSFLMNRMTSPRNKSKMKNKDTKEKEKMNRHQFVPGT 658
gi|18602674|       1     ------------------------------------------------------------ 1
gi|10438441|       1     ------------------------------------------------------------ 1
gi|15341761|       1     ------------------------------------------------------------ 1
gi|17437752|       1     ------------------------------------------------------------ 1

                          670       680       690       700       710       720
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              661   FSGVLQCLVCDKTLLGKESLQCSSCNANVHKGCKDAAPACTKKFQEKYNKNKPQTILGSS 720
gi|7106395|        659   FSGVLQCSGCDKTLLGKESLQCANCKANTHKGCKDAVPPCTKKFQEKYNKNKPQSILGSS 718
gi|18602674|       1     ------------------------------------------------------------ 1
gi|10438441|       1     ------------------------------------------------------------ 1
gi|15341761|       1     ------------------------------------------------------------ 1
gi|17437752|       1     ------------------------------------------------------------ 1

                          730       740       750       760       770       780
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              721   SFRDIPQPGLSLHPSSSVPVGLPTGRRETVGQVHPLSRSVPGTTLESFRRSATSLESESD 780
gi|7106395|        719   SVRDVPAPGLSLHPSSSMPIGLPAGRKEFAAQVHPLSRSVPGTTLESFRR--AVTSLESE 776
gi|18602674|       1     ------------------------------------------------------------ 1
gi|10438441|       1     ------------------------------------------------------------ 1
gi|15341761|       1     ------------------------------------------------------------ 1
gi|17437752|       1     ------------------------------------------------------------ 1

                          790       800       810       820       830       840
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              781   NNSCRSRSHSDELLQSMGSSPSTESFIMEDVVDSSLWSDLSSDAQEFEAESWSLVVDPSF 840
gi|7106395|        777   GDSWRSRSHSDELFQSMGSSPSTESFMMEDVVDSSLWIDLSSDAQEFEAESWSLVVDPSF 836
gi|18602674|       1     ------------------------------------------------------------ 1
gi|10438441|       1     ------------------------------------------------------------ 1
gi|15341761|       1     ------------------------------------------------------------ 1
gi|17437752|       1     ------------------------------------------------------------ 1

                          850       860       870       880       890       900
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              841   CNRQEKDVIKRQDVIFELMQTEMHHIQTLFIMSEIFRKGMKEELQLDHSTVDKIFPCLDE 900
gi|7106395|        837   CSRQEKDVIKRQDVIFELMQTEVHHIQTLLIMSEVFRKGMKEELQLDHSTVDKIFPCLDE 896
gi|18602674|       1     ------------------------------------------------------------ 1
gi|10438441|       1     ------------------------------------------------------------ 1
gi|15341761|       1     ------------------------------------------------------------ 1
gi|17437752|       1     ------------------------------------------------------------ 1

                          910       920       930       940       950       960
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              901   LLEIHRHFFYSMKERRQESSAGSDRNFVIDRIGDILVQQFSEENASKMKKIYGEFCCHHK 960
gi|7106395|        897   LLETHRHFFFSMKERRQESCAGSDRNFVINQIGDILVQQFSEENASKMKRIYGEFCSHHK 956
gi|18602674|       1     ------------------------------------------------------------ 1
gi|10438441|       1     ------------------------------------------------------------ 1
gi|15341761|       1     ------------------------------------------------------------ 1
gi|17437752|       1     ------------------------------------------------------------ 1

                          970       980       990       1000      1010      1020
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              961   EAVNLFKELQQNKKFQNFIKLRNSNLLARRRGIPECILLVTQRITKYPVLVERILQYTKE 1020
gi|7106395|        957   EAMSLFKELQQNKKFQNFIKIRNSNLLARRRGIPECILLVTQRITKYPVLVERILQYTKE 1016
gi|18602674|       1     ------------------------------------------------------------ 1
gi|10438441|       1     ------------------------------------------------------------ 1
gi|15341761|       1     ------------------------------------------------------------ 1
gi|17437752|       1     ------------------------------------------------------------ 1

                          1030      1040 .    1050      1060      1070      1080
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16              1021  RTEEHKDLRKALCLIKDMIATVDLKVNEYE---KNQKWLEILNKIENKTYTKLKNGHVFR 1077
gi|7106395|        1017  RTEEHRDLCKALGLIKDMIAAVDLKVSEYE---KNQKWLEILNKIENKTYTKLKNGHVFR 1073
gi|18602674|       1     ----------------MIATVDLKVNEYE---KNQKWLEILNKIENKTYTKLKNGHVFR 40
```

```
gi|10438441|   1    ----------------MIATVDLKVNEYE---KNQKWLEILNKIENKTYTKLKNGHVFR 40
gi|15341761|   1    ----------------MIATVDLKVNEYE---KNQKWLEILNKIENKTYTKLKNGHVFR 40
gi|17437752|   1    ----------------MPDNQNPINVCSGETRKEAWPEALEGLS----TPASNGRCES 38

                                1090      1100      1110      1120      1130      1140
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16          1078 KQALMSEERTLLYDGLVYWKTATGRFKDILALLLTDVLLFLQEKDQKYIFAAVDQKPSVI 1137
gi|7106395|    1074 KQALLSQERALLHDGLVYWKTATGRFKDILALLLTDVLLFLQEKDQKYIFAAVDQKPSVI 1133
gi|18602674|   41   KQALMSEERTLLYDGLVYWKTATGRFKDILALLLTDVLLFLQEKDQKYIFAAVDQKPSVI 100
gi|10438441|   41   KQALMSEERTLLYDGLVYWKTATGRFKDILALLLTDVLLFLQEKDQKYIFAAVDQKPSVI 100
gi|15341761|   41   KQALMSEERTLLYDGLVYWKTATGRFKDILALLLTDVLLFLQEKDQKYIFAAVDQKPSVI 100
gi|17437752|   39   HGVELQRSTSLRRAGGQGPSCLDGLFPSSGKQLLIEVVSDLVALFWGHPAGAFFR----I 94

                                1150      1160      1170      1180      1190      1200
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16          1138 SLQKLIAREVANEERGMFLISASSAGPEMYEIHTNSKEERNNWMRRIQQAVESCPEEKGG 1197
gi|7106395|    1134 SLQKLIAREVANEERGMFLISASSAGPEMYEIHTNSKEERNNWMRRIQQAVESCPEEEGG 1193
gi|18602674|   101  SLQKLIAREVANEERGMFLISASSAGPEMYEIHTNSKEERNNWMRRIQQAVESCPEEKGG 160
gi|10438441|   101  SLQKLIAREVANEERGMFLISASSAGPEMYEIHTNSKEERNNWMRRIQQAVESCPEEKGG 160
gi|15341761|   101  SLQKLIAREVANEERGMFLISASSAGPEMYEIHTNSKEERNNWMRRIQQAVESCPEEKGG 160
gi|17437752|   95   PADSLWLLGPVEGNPVLQTLPFSLGGPPS------PTARRDREMVQQEGAVIIIFSKFGG 148

                                1210      1220      1230      1240      1250      1260
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16          1198 RTSESDEDKR-KAEARVAKIQQCQEILTNQDQQICAYLEEKLHIYAELGELSGFEDVHLE 1256
gi|7106395|    1194 RTSESDEERR-KAEARVAKIQQCQEILSNQDQQICTYLEEKLHIYAELGELSGFEDVHLE 1252
gi|18602674|   161  RTSESDEDKR-KAEARVAKIQQCQEILTNQDQQICAYLEEKLHIYAELGELSGFEDVHLE 219
gi|10438441|   161  RTSESDEDKR-KAEARVAKIQQCQEILTNQDQQICAYLEEKLHIYAELGELSGFEDVHLE 219
gi|15341761|   161  RTSESDEDKR-KAEARVAKIQQCQEILTNQDQQICAYLEEKLHIYAELGELSGFEDVHLE 219
gi|17437752|   149  LGNMSQSLKLNIFKLPVTRAERQIQKFGMFGRHVC------FVTIVMVSVCKVWKEIHLA 202

                                1270      1280      1290      1300      1310      1320
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16          1257 PHLLIKPDPGEPPQAASLLAAALKEAESLQVAVKASQMGAVSQSCEDSCGDSVLADTLSS 1316
gi|7106395|    1253 PHLLIKPDPGEPPQAASLLAAALREAESLQVAVKASKMGDVSQSSEESPGGTVLMDTPST 1312
gi|18602674|   220  PHLLIKPDPGEPPQAASLLAAALKEAESLQVAVKASQMGAVSQSCEDSCGDSVLADTLSS 279
gi|10438441|   220  PHLLIKPDPGEPPQAASLLAAALKEAESLQVAVKASQMGAVSQSCEDSCGDSVLADTLSS 279
gi|15341761|   220  PHLLIKPDPGEPPQAASLLAAALKEAESLQVAVKASQMGAVSQSCEDSCGDSVLADTLSS 279
gi|17437752|   203  GQMMIYAKFDKN--------VYLPEDAEFYFTYDGSHQRHVMIA--ERIEDNVLQSSVPG 252

                                1330      1340      1350      1360      1370      1380
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16          1317 HDVPGSPTASLVTGGREGRGCSDVDPGIQGVVTDLAVSDAGEKVECRN-FPGSSQSEIIQ 1375
gi|7106395|    1313 QDVPASPTASLVTEGTEGRGCWDVDPGLQGVVTDLAVSDAGEKVEYRS-FSGSSQSEIIQ 1371
gi|18602674|   280  HDVPGSPTASLVTGGREGRGCSDVDPGIQGVVTDLAVSDAGEKVECRN-FPGSSQSEIIQ 338
gi|10438441|   280  HDVPGSPTASLVTGGREGRGCSDVDPGIQGVVTDLAVSDAGEKVECRN-FPGSSQSEIIQ 338
gi|15341761|   280  HDVPGSPTASLVTGGREGRGCSDVDPGIQGVVTDLAVSDAGEKVECRN-FPGSSQSEIIQ 338
gi|17437752|   253  HGLQETVTVSVCLCS-EGYSPVTMGSGSVTYVDNMACRLARLLVTQANRLTACSHQTLLT 311

                                1390      1400      1410      1420      1430      1440
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16          1376 AIQNLTRLLYSLQAALTIQDSHIEIHRLVLQQQEGLSLGHSILRGGPLQDQ-KSRDADRQ 1434
gi|7106395|    1372 AIQNLTRLLYSLQAALTIQDSHIEIHKLVLQQRESLAPSHS-FRGGPLQDQEKSRYLEKQ 1430
gi|18602674|   339  AIQNLTRLLYSLQAALTIQDSHIEIHRLVLQQQEGLSLGHSILRGGPLQDQ-KSRDADRQ 397
gi|10438441|   339  AIQNLTRLLYSLQAALTIQDSHIEIHRLVLQQQEGLSLGHSILRGGPLQDQ-KSRDADRQ 397
gi|15341761|   339  AIQNLTRLLYSLQAALTIQDSHIEIHRLVLQQQEGLSLGHSILRGGPLQDQ-KSRDADRQ 397
gi|17437752|   312  PFALTAGALPALDEELVLALTHLELP----------LEWTVLGSSSLEVS--------- 351

                                1450      1460      1470      1480      1490      1500
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16          1435 HEELANVHQLQHQLQQEQRRWLRRCEQQQRAQATRESWLQERERECQSQEELLLRSRGEL 1494
gi|7106395|    1431 REELANIHKLQHQFQQEQRRWHRTCDQQQREQEAQESWLQARERECQSQEELLLRHRSEL 1490
gi|18602674|   398  HEELANVHQLQHQLQQEQRRWLRRCEQQQRAQATRESWLQERERECQSQEELLLRSRGEL 457
gi|10438441|   398  HEELANVHQLQHQLQQEQRRWLRRCEQQQRAQATRESWLQERERECQSQEELLLRSRGEL 457
gi|15341761|   398  HEELANVHQLQHQLQQEQRRWLRRCEQQQRAQATRESWLQERERECQSQEELLLRSRGEL 457
gi|17437752|   351  ------SHR-ESLLHLAMRWGLAKLSQFFLCLPGGVQALALPNEEGATPLDLALR----- 399

                                1510      1520      1530      1540      1550      1560
                    ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16          1495 DLQLQEYQHSLERLREGQRLVEREQARMRAQQSLLGHWKHGRQRSLPAVLLPGGPEVMEL 1554
```

```
gi|7106395|    1491 DHQLQEYQQSLERLREGQRMVERERQKMRVQQGLLGHCKHSRQRSLPAVFSPGSKEVTEL 1550
gi|18602674|   458  DLQLQEYQHSLERLREGQRLVEREQARMRAQQSLLGHWKHGRQRSLPAVLLPGGPEVMEL 517
gi|10438441|   458  DLQLQEYQHSLERLREGQRLVEREQARMRAQQSLLGHWKHGRQRSLPAVLLPGGPEVMEL 517
gi|15341761|   458  DLQLQEYQHSLERLREGQRLVEREQARMRAQQSLLGHWKHGRQRSLPAVLLPGGPEVMEL 517
gi|17437752|   399  -EGHSKLVEDVTNFQG--RWSPSFSRVQLSEEASLHYIHSSETLTLTLNHTAEHLLEADI 456

                          1570      1580      1590      1600      1610      1620
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16          1555 NRSESLCHENSFFINEALVQMSFNTFNKLNPSVIHQDAT--YPTTQSHSDLVRTSEH-QV 1611
gi|7106395|    1551 NRAESLCHENSFFINEAFGHMSLNTSNKPNPSGVPWDA---HPLEGSHFDLARTSES-PT 1606
gi|18602674|   518  NRSESLCHENSFFINEALVQMSFNTFNKLNPSVIHQDAT--YPTTQSHSDLVRTSEH-QV 574
gi|10438441|   518  NRSESLCHENSFFINEALVQMSFNTFNKLNPSVIHQDAT--YPTTQSHSDLVRTSEH-QV 574
gi|15341761|   518  NRSESLCHENSFFINEALVQMSFNTFNKLNPSVIHQDAT--YPTTQFHSDLVRTSEH-QV 574
gi|17437752|   457  KLFRKYFWDRAFLVKAFEPEARPEERTAMPSSGAETEEARLKIVLKKQFLVIRLGAHPDA 516

                          1630      1640      1650      1660      1670      1680
                     ....|....|....|....|....|....|....|....|....|....|....|....|
NOV16          1612 DLKVDPSQPSNVS-HKLWTAAGSGHQILPFQESSKDSCKN--DLDTSHTESPTPHDSNSH 1668
gi|7106395|    1607 ELKIDISQPPDVN-SELWTTG-PGHQRPALQENSKESYKNVADLDSFQSESSSPQDSNQR 1664
gi|18602674|   575  DLKVDPSQPSNVS-HKLWTAAGSGHQILPFQESSKDSCKN--DLDTPHTESPTPHDSNSH 631
gi|10438441|   575  DLKVDPSQPSNVS-HKLWTAAGSGHQILPFQESSKDSCKN--DLDTSHTESPTPHDSNSH 631
gi|15341761|   575  DLLVDPSQPLNVS-HKLWTAAGSGHQILPFHESSKDSCKN--GN---------------- 615
gi|17437752|   517  NVLVKPTLDLPLQTALLWSPREFAFTILPLWVTLATGCVDGATLYPAAQEVYILTTTVSR 576

                          1690      1700      1710
                     ....|....|....|....|....|....|....|...
NOV16          1669 RPQLQAFITEAKLNLPTRTMTRQDGETGDGAKENIVYL 1706
gi|7106395|    1665 WPPATDTHNRSKTKSSDGGWTRWRCWRRG---------- 1693
gi|18602674|   632  RPQLQAFITEAKLNLPTRTMTRQDGETGDGAKENIVYL 669
gi|10438441|   632  RPQLQAFITEAKLNLPTRTMTRQDGETGDGAKENIVYL 669
gi|15341761|   615  ------------------------------------- 615
gi|17437752|   577  KPFIQKEDTRCARG------------------------ 590
```

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). Table 16F lists the domain description from DOMAIN analysis results against NOV16.

Table 16F Domain Analysis of NOV16

Model	Region of Homology	Score (bits)	E value
Phorbol esters/diacylglycerol binding dom (DAG_PE-bind)	654-700	40.1	4.9e-08
PHD-finger (PHD)	666-703	1.4	0.06
Phorbol esters/diacylglycerol binding domain (DAG PE-bind)	654-698	46.2	2.0e-05
C1, Protein kinase C conserved region 1	654-698	45.8	2.0e-05
RhoGEF domain (RhoGEF)	854-1044	78.4	1.5e-19
bZIP transcription factor (bZIP)	1022-1048	5.0	7.1
PH domain (PH)	1088-1189	40.9	4.5e-10

Consistent with other known members of the subunit family of proteins, NOV16 has, for example, a RhoGEF signature sequence and homology to other members of the RhoGEF-like Protein Family. NOV16 nucleic acids, and the encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, NOV16 nucleic acids and polypeptides can be used to identify proteins that are members of the RhoGEF-like Protein Family. The NOV16 nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOV16 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, *e.g.,* cellular activation, cellular replication, and signal transduction. These molecules can be used to treat, *e.g.*, cancer, trauma, regeneration (*in vitro* and *in vivo*), viral/bacterial/parasitic infections, diabetes, Von Hippel-Lindau (VHL) syndrome, pancreatitis, obesity, anemia , bleeding disorders, scleroderma, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, cirrhosis, transplantation, adrenoleukodystrophy , congenital adrenal hyperplasia, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease, lymphedema , allergies, immunodeficiencies, osteoporosis, hypercalceimia, arthritis, ankylosing spondylitis, scoliosis, tendinitis, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, ARDS, endometriosis, fertility, hyperthyroidism, hypothyroidism, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalceimia, psoriasis, actinic keratosis, tuberous sclerosis, acne, hair growth/loss, alopecia, pigmentation disorders, endocrine disorders as well as other diseases, disorders and conditions.

In addition, various NOV16 nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. For example, the NOV16 nucleic acids and their encoded polypeptides include structural motifs that are characteristic of proteins belonging to the RhoGEF-like Protein Family.

GEF (Guanine nucleotide exchange factor) for Rho/Rac/Cdc42-like GTPases is also called Dbl-homologous (DH) domain. It appears that PH (pleckstrin homology) domains

invariably occur C-terminal to RhoGEF/DH domains. Although the exact function of PH domains is unclear, several choices include binding to the beta/gamma subunit of heterotrimeric G proteins, binding to lipids, e.g. phosphatidylinositol-4,5-bisphosphate, binding to phosphorylated Ser/Thr residues, attachment to membranes by an unknown mechanism. The DAG_PE-binding domain binds two zinc ions; the ligands of these metal ions are probably the six cysteines and two histidines that are conserved in this domain and can regulate signal transduction by the PKC family of kinases.

NOV16 belongs to the guanine nucleotide exchange factor family of proteins which play a significant role in signal transduction. The guanine nucleotide exchange factor (GEF) domain that regulates GTP binding protein signaling. The GEF domain regulates positively the signaling cascades that utilize GTP-binding proteins (such as those of the ras superfamily) that function as molecular switches in fundamental events such as signal transduction, cytoskeleton dynamics and intracellular trafficking. An example of a protein containing GEF and PH domains is FGD1 (faciogenital dyplasia protein) Experiments have shown that the GEF and (PH) domains of FGD1 can bind specifically to the Rho family GTPase Cdc42Hs and stimulates the GDP-GTP exchange of the isoprenylated form of Cdc42Hs. The GEF domain of FGD1 has also been shown to activate 2 kinases involved in cell proliferation; the Jun NH2-terminal kinase and the p70 S6 kinase (Zheng *et al.*; J. Biol. Chem 1996 Dec 27;271(52):33169-72). Thus, NOV16 polypeptide may play an important role in normal development as well as disease. This class of molecules (GEFs) is also being considered as a good drug target as the guanine nucleotide exchange factor RasGRP is a high -affinity target for diacylglycerol and phorbol esters and is bound by bryostatin 1, a compound currently in clinical trials (Lorenzo *et al.*; Mol. Pharmacol 2000 May; 57(5):840-6). The homolog of RhoGEF, DRhoGEF2 fail to gastrulate due to a defect in cell shape changes required for tissue invagination and the mRNA is found throughout oogenesis and embryogenesis (Barrett *et al.*; Cell 1997; 91(7):905-15; Werner *et al.*; Gene 1997; 187(1):107-14). RhoGEF also interacts with c-Jun amino-terminal kinase (JNK) interacting protein-1 (JIP-1). JIP-1 might function as a scaffold protein by complexing specific components of the JNK signaling pathway, namely JNK, mitogen-activated protein kinase kinase 7, and mixed lineage kinase 3 (Meyer *et al.*; J Biol Chem 1999; 274(49):35113-8).

The NOV16 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in the mediation of blood and nerve physiology. As such, the NOV16 nucleic acids and polypeptides, antibodies and related compounds according to the invention may be used to

treat blood and nervous system disorders, *e.g.,* cancer, trauma, regeneration (*in vitro* and *in vivo*), viral/bacterial/parasitic infections, diabetes, Von Hippel-Lindau (VHL) syndrome, pancreatitis, obesity, anemia , bleeding disorders, scleroderma, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, cirrhosis, transplantation, adrenoleukodystrophy , congenital adrenal hyperplasia, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease, lymphedema , allergies, immunodeficiencies, osteoporosis, hypercalceimia, arthritis, ankylosing spondylitis, scoliosis, tendinitis, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, ARDS, endometriosis, fertility, hyperthyroidism, hypothyroidism, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalceimia, psoriasis, actinic keratosis, tuberous sclerosis, acne, hair growth/loss, alopecia, pigmentation disorders, endocrine disorders as well as other diseases, disorders and conditions.

The NOV16 nucleic acids and polypeptides are useful for detecting specific cell types. For example, expression analysis has demonstrated that a NOV16 nucleic acid is expressed in Adipose, Umbilical Vein, Pancreas, Thymus, Brain, Lung, Kidney, Adrenal Gland/Suprarenal gland, Peripheral Blood, Lymph node, Cartilage, Mammary gland/Breast, Uterus, Prostate, Trachea, Cochlea, Dermis, Heart, Aorta, Coronary Artery, Thyroid, Liver, Bone, Bone Marrow, Spinal Cord, Cervix, and Retina.

Additional utilities for NOV16 nucleic acids and polypeptides according to the invention are disclosed herein.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term

"nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

An 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, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

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 SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, 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 NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 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 SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, or 33, 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 an NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, or 33 is one that is sufficiently complementary to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, or 33 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, or 33, 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. 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 an 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 NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

An NOVX polypeptide is encoded by the open reading frame ("ORF") of an 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 SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, or 33; or an anti-sense strand nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, or 33; or of a naturally occurring mutant of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33.

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 an NOVX protein, such as by measuring a level of an 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 an 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 SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, or 33, that encodes a polypeptide having an NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34.

In addition to the human NOVX nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, 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 an NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 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 NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33. 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 the sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, 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 NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well-known within the art. *See*, *e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See*, *e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* **78**: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, 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 SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34. A "non-essential" amino acid residue is a residue that can be altered

from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous to the amino acid sequences SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34; more preferably at least about 70% homologous SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34; still more preferably at least about 80% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34; even more preferably at least about 90% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34; and most preferably at least about 95% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34.

An isolated nucleic acid molecule encoding an NOVX protein homologous to the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 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 an 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 SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, 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 an 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 NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, 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 an NOVX protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34, or antisense nucleic acids complementary to an NOVX nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, 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 an 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 an 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 an NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of an NOVX cDNA disclosed herein (*i.e.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33). 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 an NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

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

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

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

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, 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 nucleobases, and orientation (*see,* Hyrup, *et al.*, 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.,* 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and 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 SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34 while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, an 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, an 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 shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of an 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 an NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length. Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34, and retains the functional activity of the protein of SEQ ID

NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below.

Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34, and retains the functional activity of the NOVX proteins of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34.

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 shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33.

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, an NOVX "chimeric protein" or "fusion protein" comprises an NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an NOVX protein SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, or 34), 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 an NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of an NOVX protein. In one embodiment, an NOVX fusion protein comprises at least one biologically-active portion of an NOVX protein. In another embodiment, an NOVX fusion protein comprises at least two biologically-active portions of an NOVX protein. In yet another embodiment, an NOVX fusion protein comprises at least three biologically-active portions of an 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 an 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 an 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 an NOVX ligand and an 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 an 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 an NOVX ligand.

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

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

Anti-NOVX Antibodies

Also included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule 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 NOVX-related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally 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 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-related protein that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX-related protein sequence will indicate which regions of a NOVX-related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

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

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow and Lane, 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). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. 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 relate to antibody molecules in which essentially the entire sequences 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.*, 1991 *EMBO J.*, 10:3655-3659.

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.

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

Anti-NOVX antibodies may be used in methods known within the art relating to the localization and/or quantitation of an NOVX protein (*e.g.*, for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for NOVX proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody

derived binding domain, are utilized as pharmacologically-active compounds (hereinafter "Therapeutics").

An anti-NOVX antibody (*e.g.*, monoclonal antibody) can be used to isolate an NOVX polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-NOVX antibody can facilitate the purification of natural NOVX polypeptide from cells and of recombinantly-produced NOVX polypeptide expressed in host cells. Moreover, an anti-NOVX antibody can be used to detect NOVX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the NOVX protein. Anti-NOVX antibodies 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.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an 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 □-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 SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27,

29, 31, and 33 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 an NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 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.*, an 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 an 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 an 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 an 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 an NOVX protein, wherein determining the ability of the test compound to interact with an 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 an NOVX target

molecule. As used herein, a "target molecule" is a molecule with which an NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses an 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. An NOVX target molecule can be a non-NOVX molecule or an NOVX protein or polypeptide of the invention. In one embodiment, an 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 an 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 an 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 an 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 an 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 an NOVX protein, wherein determining the ability of the test compound to interact with an 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 an 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 an 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 an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of an 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 likely to be involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway. The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an 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, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, 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 predicted coding sequences, such as those in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 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 an 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 NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, 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 an 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 an 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 an NOVX gene; (*ii*) an addition of one or more nucleotides to an NOVX gene; (*iii*) a substitution of one or more nucleotides of an NOVX gene, (*iv*) a chromosomal rearrangement of an NOVX gene; (*v*) an alteration in the level of a messenger RNA transcript of an NOVX gene, (*vi*) aberrant modification of an 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 an NOVX gene, (*viii*) a non-wild-type level of an NOVX protein, (*ix*) allelic loss of an NOVX gene, and (*x*) inappropriate post-translational modification of an 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 an 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 an 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 an 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 an 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 an 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 P450 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 an 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

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 192

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 192

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

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

WHAT IS CLAIMED IS:

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

(a) a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34;

(b) a variant of a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of the amino acid residues from the amino acid sequence of said mature form;

(c) an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34; and

(d) a variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34 wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence.

2. The polypeptide of claim 1, wherein said polypeptide comprises the amino acid sequence of a naturally-occurring allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34.

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 NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33.

4. The polypeptide of claim 1, wherein the amino acid sequence of said variant comprises a conservative amino acid substitution.

5. 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 an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34;

(b) a variant of a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of the amino acid residues from the amino acid sequence of said mature form;

(c) an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34;

(d) a variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence;

(e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising an amino acid sequence chosen from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34, or a variant of said polypeptide, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence; and

(f) a nucleic acid molecule comprising the complement of (a), (b), (c), (d) or (e).

6. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally-occurring allelic nucleic acid variant.

7. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule encodes a polypeptide comprising the amino acid sequence of a naturally-occurring polypeptide variant.

8. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33.

9. The nucleic acid molecule of claim 5, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of

(a) a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33;

(b) a nucleotide sequence differing by one or more nucleotides from a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, provided that no more than 20% of the nucleotides differ from said nucleotide sequence;

(c) a nucleic acid fragment of (a); and

(d) a nucleic acid fragment of (b).

10. The nucleic acid molecule of claim 5, wherein said nucleic acid molecule hybridizes under stringent conditions to a nucleotide sequence chosen from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, or a complement of said nucleotide sequence.

11. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of

(a) a first nucleotide sequence comprising a coding sequence differing by one or more nucleotide sequences from a coding sequence encoding said amino acid sequence, provided that no more than 20% of the nucleotides in the coding sequence in said first nucleotide sequence differ from said coding sequence;

(b) an isolated second polynucleotide that is a complement of the first polynucleotide; and

(c) a nucleic acid fragment of (a) or (b).

12. A vector comprising the nucleic acid molecule of claim 11.

13. The vector of claim 12, further comprising a promoter operably-linked to said nucleic acid molecule.

14. A cell comprising the vector of claim 12.

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

16. The antibody of claim 15, wherein said antibody is a monoclonal antibody.

17. The antibody of claim 15, wherein the antibody is a humanized antibody.

18. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

(a) providing the sample;

(b) contacting the sample with 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.

19. A method for determining the presence or amount of the nucleic acid molecule of claim 5 in a sample, the method comprising:

(a) providing the sample;

(b) contacting the sample with a probe that binds to said nucleic acid molecule; and

(c) determining the presence or amount of the probe bound to said nucleic acid molecule,

thereby determining the presence or amount of the nucleic acid molecule in said sample.

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

(a) contacting said polypeptide with said agent; and

(b) determining whether said agent binds to said polypeptide.

21. A method for identifying an agent that modulates the expression or activity of the polypeptide of claim 1, the method comprising:

(a) providing a cell expressing said polypeptide;

(b) contacting the cell with said agent; and

(c) determining whether the agent modulates expression or activity of said polypeptide,

whereby an alteration in expression or activity of said peptide indicates said agent modulates expression or activity of said polypeptide.

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

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

24. The method of claim 23, wherein said subject is a human.

25. A method of treating or preventing a NOVX-associated disorder, said method comprising administering to a subject in which such treatment or prevention is desired the nucleic acid of claim 5 in an amount sufficient to treat or prevent said NOVX-associated disorder in said subject.

26. The method of claim 25, wherein said subject is a human.

27. A method of treating or preventing a NOVX-associated disorder, said method comprising administering to a subject in which such treatment or prevention is desired the antibody of claim 15 in an amount sufficient to treat or prevent said NOVX-associated disorder in said subject.

28. The method of claim 27, wherein the subject is a human.

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

30. A pharmaceutical composition comprising the nucleic acid molecule of claim 5 and a pharmaceutically-acceptable carrier.

31. A pharmaceutical composition comprising the antibody of claim 15 and a pharmaceutically-acceptable carrier.

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

33. A kit comprising in one or more containers, the pharmaceutical composition of claim 30.

34. A kit comprising in one or more containers, the pharmaceutical composition of claim 31.

35. 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 NOVX-associated disorder, wherein said therapeutic is selected from the group consisting of a NOVX polypeptide, a NOVX nucleic acid, and a NOVX antibody.

36. A method for screening for a modulator of activity or of latency or predisposition to a NOVX-associated disorder, said method comprising:

(a) administering a test compound to a test animal at increased risk for a NOVX-associated disorder, wherein said test animal recombinantly expresses the polypeptide of claim 1;

(b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a);

(c) comparing the activity of said protein in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator of latency of or predisposition to a NOVX-associated disorder.

37. The method of claim 36, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

38. 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.

39. A method for determining the presence of or predisposition to a disease associated with altered levels of the nucleic acid molecule of claim 5 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.

40. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising an amino acid sequence of at least one of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34, or a biologically active fragment thereof.

41. A method of treating a pathological state in a mammal, the method comprising administering to the mammal the antibody of claim 15 in an amount sufficient to alleviate the pathological state.