

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

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
19 February 2004 (19.02.2004)

PCT

(10) International Publication Number
WO 2004/015060 A2

(51) International Patent Classification ⁷ :	C12N	US	60/373,050 (CIP)
		Filed on	16 April 2002 (16.04.2002)
(21) International Application Number:	PCT/US2002/024492	US	60/380,970 (CIP)
		Filed on	15 May 2002 (15.05.2002)
		US	60/311,979 (CIP)
(22) International Filing Date:	2 August 2002 (02.08.2002)	Filed on	13 August 2001 (13.08.2001)
		US	60/381,030 (CIP)
(25) Filing Language:	English	Filed on	16 May 2002 (16.05.2002)
		US	60/323,944 (CIP)
(26) Publication Language:	English	Filed on	21 September 2001 (21.09.2001)
		US	60/311,571 (CIP)
(30) Priority Data:		Filed on	10 August 2001 (10.08.2001)
60/310,802	8 August 2001 (08.08.2001)	US	60/311,594 (CIP)
60/310,795	8 August 2001 (08.08.2001)	US	Filed on
60/311,292	9 August 2001 (09.08.2001)	US	10 August 2001 (10.08.2001)
60/311,751	10 August 2001 (10.08.2001)	US	60/313,201 (CIP)
60/311,594	10 August 2001 (10.08.2001)	Filed on	17 August 2001 (17.08.2001)
60/311,571	10 August 2001 (10.08.2001)	US	60/359,294 (CIP)
60/311,979	13 August 2001 (13.08.2001)	Filed on	21 February 2002 (21.02.2002)
60/312,892	16 August 2001 (16.08.2001)	US	60/372,998 (CIP)
60/313,201	17 August 2001 (17.08.2001)	Filed on	16 April 2002 (16.04.2002)
60/314,031	21 August 2001 (21.08.2001)	US	60/380,971 (CIP)
60/315,853	29 August 2001 (29.08.2001)	Filed on	15 May 2002 (15.05.2002)
60/322,716	17 September 2001 (17.09.2001)	US	60/312,892 (CIP)
60/323,944	21 September 2001 (21.09.2001)	Filed on	16 August 2001 (16.08.2001)
60/359,294	21 February 2002 (21.02.2002)	US	60/322,716 (CIP)
60/361,159	28 February 2002 (28.02.2002)	Filed on	17 September 2001 (17.09.2001)
60/360,890	28 February 2002 (28.02.2002)	US	60/360,890 (CIP)
60/373,050	16 April 2002 (16.04.2002)	Filed on	28 February 2002 (28.02.2002)
60/372,998	16 April 2002 (16.04.2002)	US	60/314,031 (CIP)
60/380,971	15 May 2002 (15.05.2002)	Filed on	21 August 2001 (21.08.2001)
60/380,970	15 May 2002 (15.05.2002)	US	60/315,853 (CIP)
60/381,030	16 May 2002 (16.05.2002)	Filed on	29 August 2001 (29.08.2001)
10/211,689	1 August 2002 (01.08.2002)	US	10/211,689 (CIP)
		Filed on	1 August 2002 (01.08.2002)
(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:		(71) Applicant (for all designated States except US): CURA-GEN CORPORATION [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CN 06511 (US).	
US	60,311,751 (CIP)	(72) Inventors; and	
Filed on	10 August 2001 (10.08.2001)	(75) Inventors/Applicants (for US only): ANDERSON, David, W. [US/US]; 85 Montoya Drive, Branford, Connecticut 06405, United States of America (US). BOLDOG, Ferenc, L. [HU/US]; 1687 Hartford Turnpike, North Haven, Connecticut 06473, United States of America (US). CASMAN, Stacie, J. [US/US]; 17 Peck Street, North Haven, Connecticut 06473, United States of America (US). EDINGER, Shlomit, R. [US/US]; 766 Edgewood	
US	60/310,802 (CIP)		
Filed on	8 August 2001 (08.08.2001)		
US	60/310,795 (CIP)		
Filed on	8 August 2001 (08.08.2001)		
US	60/311,292 (CIP)		
Filed on	9 August 2001 (09.08.2001)		
US	60/361,159 (CIP)		
Filed on	28 February 2002 (28.02.2002)		

[Continued on next page]

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

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

WO 2004/015060 A2



Avenue, New Haven, Connecticut 06515, United States of America (US). **GANGOLLI, Esha, A.** [US/US]; 31 Strawberry Hill Road, Madison, Connecticut 06443, United States of America (US). **GERLACH, Valerie, L.** [US/US]; 18 Rock Pasture Road, Branford, Connecticut 06405, United States of America (US). **GORMAN, Linda** [US/US]; 329 Monticello Drive, Branford, Connecticut 06405, United States of America (US). **GUO, Xiaojia** [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). **KHRAMTSOV, Nikolai, V.** [RU/US]; 38 Montoya Circle, Branford, Connecticut 06405, United States of America (US). **LI, Li** [CN/US]; 56 Jerimoth Drive, Branford, Connecticut 06405, United States of America (US). **MACDOUGALL, John, R.** [CA/US]; 117 Russel Street, Hamden, Connecticut 06517, United States of America (CA). **PENA, Carol, E., A.** [US/US]; 604 Orange Street, Number 2, New Haven, Connecticut 06511, United States of America (US). **PEYMAN, John, A.** [US/US]; 336 West Rock Avenue, New Haven, Connecticut 06515, United States of America (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apt. 1C, Branford, CT 06405 (US). **SHIMKETS, Richard, A.** [US/US]; 5 Indian Meadows Drive, Guilford, Connecticut 06437, United States of America (US). **SMITHSON, Glennnda** [US/US]; 125 Michael Drive, Guilford, CT 06405 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street, #1, New Haven, CT 06511 (US). **VERNET, Corine, A. M.** [FR/US]; 1739 Foxon Road, Apartment L6, Branford, Connecticut 06471, United States of America (US). **VOSS, Edward, Z.** [US/US]; 123 Knollwood Drive, Wallingford, Connecticut 06492, United States of America (US). **ZHONG, Mei**

[CA/US]; 45 Harrison Avenue, Apartment 1B, Branford, CT 06405 (US).

(74) **Agent: ELRIFI, Ivor, R.;** Mintz, Levin, Cohn, Ferris, Glosky and Popeo PC, One Financial Center, Boston, MA 0211 (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, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

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

5

10

15

FIELD OF THE INVENTION

The present invention relates to novel polypeptides, and the nucleic acids encoding
20 them, having properties related to stimulation of biochemical or physiological responses in
a cell, a tissue, an organ or an organism. More particularly, the novel polypeptides are
gene products of novel genes, or are specified biologically active fragments or derivatives
thereof. Methods of use encompass diagnostic and prognostic assay procedures as well as
methods of treating diverse pathological conditions.

25

BACKGROUND OF THE INVENTION

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

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

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

Many pathological conditions involve dysregulation of expression of important effector proteins. In certain classes of pathologies the dysregulation is manifested as diminished or suppressed level of synthesis and secretion of protein effectors. In other classes of pathologies the dysregulation is manifested as increased or up-regulated level of synthesis and secretion of protein effectors. In a clinical setting a subject may be suspected of suffering from a condition brought on by altered or mis-regulated levels of a

protein effector of interest. Therefore there is a need to assay for the level of the protein effector of interest in a biological sample from such a subject, and to compare the level with that characteristic of a nonpathological condition. There also is a need to provide the protein effector as a product of manufacture. Administration of the effector to a subject in need thereof is useful in treatment of the pathological condition. Accordingly, there is a need for a method of treatment of a pathological condition brought on by a diminished or suppressed levels of the protein effector of interest. In addition, there is a need for a method of treatment of a pathological condition brought on by a increased or up-regulated levels of the protein effector of interest.

10

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of isolated polypeptides including amino acid sequences selected from mature forms of the amino acid sequences selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39. The invention also is based in part upon variants of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. In another embodiment, the invention includes the amino acid sequences selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39. In another embodiment, the invention also comprises variants of the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed. The invention also involves fragments of any of the mature forms of the amino acid sequences selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, or any other amino acid sequence selected from this group. The invention also comprises fragments from these groups in which up to 15% of the residues are changed.

In another embodiment, the invention encompasses polypeptides that are naturally occurring allelic variants of the sequence selected from the group consisting of SEQ ID

NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39. These allelic variants include amino acid sequences that are the translations of nucleic acid sequences differing by a single nucleotide from nucleic acid sequences selected from the group consisting of SEQ ID NOS: 121, 123, 125, 127, 129, 131, and
5 2n-1, wherein n is an integer between 1 and 39. The variant polypeptide where any amino acid changed in the chosen sequence is changed to provide a conservative substitution.

In another embodiment, the invention comprises a pharmaceutical composition involving a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NOS:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer
10 between 1 and 39 and a pharmaceutically acceptable carrier. In another embodiment, the invention involves a kit, including, in one or more containers, this pharmaceutical composition.

In another embodiment, the invention includes the use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease,
15 the disease being selected from a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NOS:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39 wherein said therapeutic is the polypeptide selected from this group.

In another embodiment, the invention comprises a method for determining the
20 presence or amount of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NOS:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39 in a sample, the method involving providing the sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the polypeptide, thereby
25 determining the presence or amount of polypeptide in the sample.

In another embodiment, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NOS:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39 in a first
30 mammalian subject, the method involving measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in this sample to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease, wherein an alteration in the expression level of the polypeptide in the first subject

as compared to the control sample indicates the presence of or predisposition to the disease.

In another embodiment, the invention involves a method of identifying an agent that binds to a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, the method including introducing the polypeptide to the agent; and determining whether the agent binds to the polypeptide. The agent could be a cellular receptor or a downstream effector.

In another embodiment, the invention involves a method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, the method including providing a cell expressing the polypeptide of the invention and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent.

In another embodiment, the invention involves a method for screening for a modulator of activity or of latency or predisposition to a pathology associated with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, the method including administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of the invention, wherein the test animal recombinantly expresses the polypeptide of the invention; measuring the activity of the polypeptide in the test animal after administering the test compound; and comparing the activity of the protein in the test animal with the activity of the polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of, or predisposition to, a pathology associated with the polypeptide of the invention. The recombinant test animal could express a test protein transgene or express the transgene under the control of a promoter at an increased level relative to a

wild-type test animal The promoter may or may not be the native gene promoter of the transgene.

In another embodiment, the invention involves a method for modulating the activity of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer
5 between 1 and 39, the method including introducing a cell sample expressing the polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

In another embodiment, the invention involves a method of treating or preventing a
10 pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, the method including administering the polypeptide to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject. The subject could be human.

15 In another embodiment, the invention involves a method of treating a pathological state in a mammal, the method including administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124,
20 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39 or a biologically active fragment thereof.

In another embodiment, the invention involves an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence
25 given SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39; a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more
30 than 15% of the amino acid residues in the sequence of the mature form are so changed; the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39; a variant of the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1

and 39, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID

5 NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39 or any variant of the polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and the complement of any of the nucleic acid molecules.

10 In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, wherein the nucleic acid molecule comprises the nucleotide
15 sequence of a naturally occurring allelic nucleic acid variant.

In another embodiment, the invention involves an isolated nucleic acid molecule including a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer
20 between 1 and 39 that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.

In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an
25 integer between 1 and 39, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 121, 123, 125, 127, 129, 131, and 2n-1, wherein n is an integer between 1 and 39.

In another embodiment, the invention includes an isolated nucleic acid molecule
30 having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of the nucleotide sequence selected from the group

consisting of SEQ ID NO:121, 123, 125, 127, 129, 131, and 2n-1, wherein n is an integer between 1 and 39; a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NOs:121, 123, 125, 127, 129, 131, and SEQ ID NO:2n-1, wherein n is an integer between 1 and 39, is changed
5 from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NOs:121, 123, 125, 127, 129, 131, and SEQ ID NO:2n-1, wherein n is an integer between 1 and 39; and a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence
10 selected from the group consisting of SEQ ID NOs:121, 123, 125, 127, 129, 131, and SEQ ID NO:2n-1, wherein n is an integer between 1 and 39, is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

In another embodiment, the invention includes an isolated nucleic acid molecule
15 having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, wherein the nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NOs:121, 123,
20 125, 127, 129, 131, and SEQ ID NO:2n-1, wherein n is an integer between 1 and 39, or a complement of the nucleotide sequence.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given
25 SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, wherein the nucleic acid molecule has a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen nucleotide sequence is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides in the chosen coding
30 sequence are so changed, an isolated second polynucleotide that is a complement of the first polynucleotide, or a fragment of any of them.

In another embodiment, the invention includes a vector involving the nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence

given SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39. This vector can have a promoter operably linked to the nucleic acid molecule. This vector can be located within a cell.

5 In another embodiment, the invention involves a method for determining the presence or amount of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39 in a sample, the method including providing the sample; introducing the sample to a probe that binds to the nucleic acid molecule; and determining the presence or amount of the probe bound to the nucleic acid molecule, thereby determining the presence or amount of the nucleic acid molecule in the sample. The presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type. The cell type can be cancerous.

15 In another embodiment, the invention involves a method for determining the presence of or predisposition for a disease associated with altered levels of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39 in a first mammalian subject, the method including measuring the amount of the nucleic acid in a sample from the first mammalian subject; and comparing the amount of the nucleic acid in the sample of step (a) to the amount of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease; wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

25 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

5 The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences, their encoded polypeptides, antibodies, and other related compounds. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX
10 proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table 1 provides a summary of the NOVX nucleic acids and their encoded polypeptides.

TABLE 1. Sequences and Corresponding SEQ ID Numbers

NOVX ASSIGNMENT	INTERNAL IDENTIFICATION	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	HOMOLOGY
1a	CG101926-01	1	2	Osteoglycin
1b	207639361	3	4	Osteoglycin
2a	CG109754-02	5	6	Bioactive Peptide Thymosin beta-10
2b	CG109754-01	121	122	Thymosin beta-10
3a	CG114834-02	7	8	Bioactive Peptide Thymosin alpha
3b	CG114834-01	123	124	Thymosin beta-10
4a	CG124728-02	9	10	Complement-C1q tumor necrosis factor-related protein like homo sapiens
4b	CG124728-03	11	12	Complement-C1q tumor necrosis factor-related protein like homo sapiens
4c	263479529	13	14	Complement-C1Q -like Proteins
4d	271674589	15	16	Complement-C1Q -like Proteins
5a	CG127616-01	17	18	Erythropoietin like homo sapiens
5b	CG127616-02	19	20	Erythropoietin splice variant-like Proteins, derived Peptides, and Nucleic Acids Encoding Same
5c	214374151	21	22	Erythropoietin splice variant-like Proteins, derived Peptides, and Nucleic Acids Encoding Same
5d	219936857	23	24	Erythropoietin
5e	259333914	25	26	Erythropoietin
5f	219936857	27	28	Erythropoietin IFC
6a	CG128348-01	29	30	Apolipoprotein E Precursor like homo sapiens
6b	CG128348-02	31	32	Apolipoprotein E Precursor like

				homo sapiens
6c	CG128348-03	33	34	Apolipoprotein E Precursor like homo sapiens
6d	278480724	35	36	Apolipoprotein E Precursor-like
6e	278480754	37	38	Apolipoprotein E Precursor-like
6f	278480861	39	40	Apolipoprotein E Precursor-like
6g	278876278	41	42	Apolipoprotein E Precursor
6h	278876445	43	44	Apolipoprotein E Precursor
6i	278876664	45	46	Apolipoprotein E Precursor
7a	CG129136-01	47	48	4-1BB Ligand like homo sapiens
7b	CG129136-03	49	50	Novel 4-1BB Ligand-like Proteins and Nucleic Acids Encoding Same
7c	278876545	51	52	IFC- 4-1BB Ligand
7d	278876597	53	54	IFC- 4-1BB Ligand
8a	CG133483-01	55	56	Win4 like homo sapiens
9a	CG135316-01	57	58	splice variant of FGF14 like homo sapiens
10a	CG53473-03	59	60	Bioactive Peptide Neuromedin
10b	CG53473-01	125	126	Neuromedin
11a	CG54725-03	61	62	Bioactive Peptide Thymosin beta-10
11b	CG54725-01	127	128	Thymosin beta-10
12a	CG56983-02	63	64	Bioactive Peptide YY
12b	CG56983-03	65	66	Bioactive Peptide YY
12c	CG56983-04	67	68	Bioactive Peptide YY
12d	CG56983-01	129	130	YY Peptide
13a	CG57666-02	69	70	Bioactive Peptide MHC Class I
14a	CG59651-03	71	72	Growth/differentiation factor-15
14b	207775400	73	74	Growth/differentiation factor-15
14c	207775387	75	76	Prostate differentiation factor-like Proteins, derived Peptides, and Nucleic Acids Encoding Same
15a	CG89614-04	77	78	Bioactive Peptide Oxytocin
15b	CG89614-01	131	132	Oxytocin

Table 1 indicates homology of NOVX nucleic acids to known protein families.

Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table 1 will be useful in therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table 1.

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

Consistent with other known members of the family of proteins, identified in column 5 of Table 1, the NOVX polypeptides of the present invention show homology to, and contain domains that are characteristic of, other members of such protein families.

Details of the sequence relatedness and domain analysis for each NOVX are presented in

5 Example A.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit diseases associated with the

10 protein families listed in Table 1.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are presented in Example C.

Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the

15 detection of a variety of diseases with differential expression in normal vs. diseased tissues, *e.g.* a variety of cancers.

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

NOVX clones

20 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
25 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
30 described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

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

10 In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) an amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and (e) a fragment of any of (a) through (d).

In another specific embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence given, SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid

residues in the sequence of the mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; (e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NOs:122, 124, 126, 128, 130, 132, and SEQ ID NO:2n, wherein n is an integer between 1 and 39, or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and (f) the complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NOs:121, 123, 125, 127, 129, 131, and SEQ ID NO:2n-1, wherein n is an integer between 1 and 39; (b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NOs:121, 123, 125, 127, 129, 131, and SEQ ID NO:2n-1, wherein n is an integer between 1 and 39, is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NOs:121, 123, 125, 127, 129, 131, and SEQ ID NO:2n-1, wherein n is an integer between 1 and 39; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NOs:121, 123, 125, 127, 129, 131, and SEQ ID NO:2n-1, wherein n is an integer between 1 and 39, is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

30

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify

NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term “nucleic acid molecule” is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

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

The term “probe”, as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), and 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific,

and much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as used herein, is a nucleic acid which
5 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
10 contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, 0.1 kb, or less of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, culture medium, or of chemical precursors or
15 other chemicals.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39, or a complement of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information
20 provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39, 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
25 Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993).

A nucleic acid of the invention can be amplified using cDNA, mRNA or, alternatively, genomic DNA as a template with appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be
30 cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues. A short oligonucleotide sequence may be based on, or designed from, a genomic

or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide
5 comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs: 2n-1, wherein n is an integer between 1 and 39, 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
10 comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs: 2n-1, wherein n is an integer between 1 and 39, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of a NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide
15 sequence shown SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39, is one that is sufficiently complementary to the nucleotide sequence shown SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39, that it can hydrogen bond with few or no mismatches to the nucleotide sequence shown SEQ ID NOS: 121, 123, 125, 127, 129,
20 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39, 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
25 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
30 without other substantial chemical intermediates.

“Fragments” provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, and are at most some portion less than a full length sequence.

Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice.

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

"Derivatives" are nucleic acid sequences or amino acid sequences formed from the native compounds either directly, by modification, or by partial substitution. "Analogues" are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound, *e.g.* they differ from it in respect to certain components
15 or side chains. Analogs may be synthetic or derived from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length. Derivatives or
20 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
25 by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the proteins of the invention under 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.

30 A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences include those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of

RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms.

- 5 Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding a human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see
10 below) in SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

- A NOVX polypeptide is encoded by the open reading frame ("ORF") of a NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be
15 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
20 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
25 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 a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NOS:
30 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39; or an anti-sense strand nucleotide sequence of SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39; or of a naturally occurring mutant of SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe has a detectable label attached, *e.g.* the label can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes

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

"A polypeptide having a biologically-active portion of a NOVX polypeptide"

10 refers to polypeptides exhibiting activity similar, but not necessarily identical, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer

15 between 1 and 39, that encodes a polypeptide having a NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

20 NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide

25 sequences shown in SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39. 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: 122, 124, 126, 128, 130, 132, and SEQ ID NOs:2n, wherein n is an integer between 1 and 39.

30 In addition to the human NOVX nucleotide sequences shown in SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic

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

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39, 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: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 65% homologous to each other typically remain hybridized to each other.

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

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

different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess at T_m, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60 °C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65 °C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50 °C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39, 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: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39, 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.*

- 5 In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39, 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%
10 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,*
15 *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

- 20 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: 121, 123, 125, 127, 129, 131, and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39, thereby leading to changes in the amino acid sequences of the encoded NOVX proteins, without
25 altering the functional ability of the NOVX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence SEQ ID NOS:122, 124, 126, 128, 130, 132 and 2n, wherein n is an integer between 1 and 39. 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
30 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: 121, 123, 125, 127, 129, 131, and SEQ ID NOS:2n-1, wherein n is an integer between 1 and 39, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 40% homologous to the amino acid sequences SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39; more preferably at least about 70% homologous SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39; still more preferably at least about 80% homologous to SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39; even more preferably at least about 90% homologous to SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39; and most preferably at least about 95% homologous to SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39.

An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein of SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:121, 123, 125, 127, 129, 131, and 2n-1, wherein n is an integer between 1 and 39, 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:121, 123, 125, 127, 129, 131, and 2n-1, wherein n is an integer between 1 and 39, 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
5 along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein
10 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,
15 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.

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

25 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
30 molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the

coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39, or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39, are additionally provided.

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

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

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil,

hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, beta-D-mannosylqueosine, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 5 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, 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 10 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 15 interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by 20 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 25 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 30 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. A 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
5 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).

10 **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
15 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
20 Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX cDNA disclosed herein (*i.e.*, SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39).
25 For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. See, e.g., U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See,
30 e.g., Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (e.g., the NOVX promoter and/or enhancers) to form triple helical structures that prevent

transcription of the NOVX gene in target cells. *See, e.g., Helene, 1991. Anticancer Drug Des. 6: 569-84; Helene, et al. 1992. Ann. N.Y. Acad. Sci. 660: 27-36; Maher, 1992. Bioassays 14: 807-15.*

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

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.,* inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.,* PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.,* S₁ nucleases (*See, Hyrup, et al., 1996. supra*); or as probes or primers for DNA sequence and hybridization (*See, Hyrup, et al., 1996, supra*; Perry-O'Keefe, *et al., 1996. supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.,* to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.,* RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see, Hyrup, et al., 1996. supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al., 1996. supra* and Finn, *et al., 1996. Nucl Acids Res 24: 3357-3363.* For

example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and 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.

- 5 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.

- 10 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
15 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.

20

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:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39. The
25 invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

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

deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

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

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

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

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence shown in SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of a NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

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

In an embodiment, the NOVX protein has an amino acid sequence shown SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39, and retains the functional activity of the protein of SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39, 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:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39, and retains the functional activity of the NOVX proteins of SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39.

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
5 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%,
10 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39.

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
15 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
20 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
25 sequence identity as compared to a reference sequence over a comparison region.

CHIMERIC AND FUSION PROTEINS

The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide
30 operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a NOVX protein SEQ ID NOS:122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX

protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically active portion of a NOVX protein. In another
5 embodiment, a NOVX fusion protein comprises at least two biologically active portions of a NOVX protein. In yet another embodiment, a NOVX fusion protein comprises at least three biologically active portions of a NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX
10 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.

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

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

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional

techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional
5 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,
10 many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX AGONISTS AND ANTAGONISTS

15 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
20 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
25 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
30 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
5 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
10 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

15 In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs
20 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 S₁ nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal
25 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
30 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 Youvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 5 6:327-331.

NOVX Antibodies

The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} , F_{ab}' and $F_{(ab)2}$ fragments, and an F_{ab} expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated protein of the invention intended to serve as an antigen, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NOs: 122, 124, 126, 128, 130, 132, and 2n, wherein n is an integer between 1 and 39, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, *e.g.*, a

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

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

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

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*,

lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.),
adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium
parvum, or similar immunostimulatory agents. Additional examples of adjuvants which
can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic
5 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
10 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).

15

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
20 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
25 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.

30 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
5 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
10 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
15 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].

20 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
25 techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

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

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

5 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
10 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
15 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
20 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

25 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
30 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); Verhoeven *et al.*, Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human

antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones *et al.*, 1986; Riechmann *et al.*, 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

Human Antibodies

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

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks *et al.*, J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, *e.g.*, mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks *et al.* (Bio/Technology 10,

779-783 (1992)); Lonberg *et al.* (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild *et al.*, (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

5 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
10 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
15 preferred embodiment of such a nonhuman animal is a mouse, and is termed the XenomouseTM as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized
20 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,
25 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
30 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.

5 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.

10 **F_{ab} Fragments and Single Chain Antibodies**

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

Bispecific Antibodies

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

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the

correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, EMBO J., 10:3655-3659 (1991).

- 5 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
- 10 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).
- 15 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
- 20 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.
- 25 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
- 30 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)

5 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
10 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
15 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
20 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
25 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
30 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
5 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
10 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
15 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

20 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
25 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
30 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).

5 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
10 (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
15 bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as
20 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
25 exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

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

Immunoliposomes

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein *et al.*, Proc. Natl. Acad. Sci. USA, **82**: 3688 (1985); Hwang *et al.*, Proc. Natl. Acad. Sci. USA, **77**: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545.

5 Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

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

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

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

An antibody specific for a protein of the invention can be used to isolate the protein by standard techniques, such as immunoaffinity chromatography or
25 immunoprecipitation. Such an antibody can facilitate the purification of the natural protein antigen from cells and of recombinantly produced antigen expressed in host cells. Moreover, such an antibody can be used to detect the antigenic protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic protein. Antibodies directed against the protein can be used diagnostically to
30 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
5 fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibody Therapeutics

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

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

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases,
30 interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume of the subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or

antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

5 **Pharmaceutical Compositions of Antibodies**

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

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

30 The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes,

albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

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

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

ELISA Assay

- 20 An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, F_{ab} or F_{(ab)2}) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect
25 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,
30 as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection

of an analyte mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of an analyte genomic DNA include Southern hybridizations.

- 5 Procedures for conducting immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulos, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Theory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, *in vivo* techniques
- 10 for detection of an analyte protein include introducing into a subject a labeled anti-analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

15 **NOVX Recombinant Expression Vectors and Host Cells**

- Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of
- 20 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
- 25 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
- 30 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
5 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).

10 The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in
15 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
20 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
25 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
30 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
5 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,
10 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
15 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
20 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
25 techniques.

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

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

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

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

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

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

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

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

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

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

methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector.

Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while
5 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
10 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

15 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
20 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
25 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
30 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: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of SEQ ID NOS: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39), 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: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39, can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is then introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND 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.

- 5 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 G₀ phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species
- 10 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.

15 **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
- 20 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
- 25 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
- 30 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, 5 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 10 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, 15 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. 20 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 25 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 30 which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed

by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and
5 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
10 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
15 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.

20 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
25 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
30 the art.

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

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

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

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

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

Screening and Detection Methods

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

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

Screening Assays

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

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring

deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. See, e.g., Lam, 1997.

5 *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
10 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.*
15 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
20 (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.).

25 In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a NOVX protein determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be
30 accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or

by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase; and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a
5 membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test
10 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
15 compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule
20 with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention. In one
25 embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with
30 NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by

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

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

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

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and

determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

- 5 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
- 10 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).
- 15 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
- 20 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
- 25 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
- 30 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

5 (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
10 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.

15 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
20 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
25 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
30 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
5 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
10 codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor.
15 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.

20

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)
25 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.

30

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: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein

n is an integer between 1 and 39, 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.

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

30

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: 121, 123, 125, 127, 129, 131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of

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

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

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

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

DIAGNOSTIC ASSAYS

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NOS: 121, 123, 125, 127, 129,

131, and SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 39, 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
5 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
10 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
15 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
20 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
25 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
30 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
5 capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

10

PROGNOSTIC ASSAYS

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the
15 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
20 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
25 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
30 example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic

acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

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

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

comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see*,
5 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);
Q β Replicase (*see*, Lizardi, *et al.*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid
amplification method, followed by the detection of the amplified molecules using
techniques well known to those of skill in the art. These detection schemes are especially
10 useful for the detection of nucleic acid molecules if such molecules are present in very low
numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can
be identified by alterations in restriction enzyme cleavage patterns. For example, sample
and control DNA is isolated, amplified (optionally), digested with one or more restriction
15 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.

20 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 oligonucleotide 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
25 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
30 characterization of specific mutations by using smaller, specialized probe arrays
complementary to all variants or mutations detected. Each mutation array is composed of
parallel probe sets, one complementary to the wild-type gene and the other complementary
to the mutant gene.

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

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. See, e.g., Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S₁ nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

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

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

- 5 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
10 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
20 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
25 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
30 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.

- 5 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
10 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
15 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
20 diagnose patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for
25 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
30 can be administered to individuals to treat (prophylactically or therapeutically) disorders (The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X

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

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

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome Pregnancy Zone Protein Precursor enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which

all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its
5 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
10 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
15 efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

MONITORING OF EFFECTS DURING CLINICAL TRIALS

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

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule)

that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

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

30 **Methods of Treatment**

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect

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

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

DISEASES AND DISORDERS

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (i) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to an aforementioned peptide; (iii) nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or

mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

PROPHYLACTIC METHODS

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

Therapeutic Methods

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

to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination
5 of agents that modulates (*e.g.*, up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situations* in which NOVX is
10 abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

15

Determination of the Biological Effect of the Therapeutic

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

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

Prophylactic and Therapeutic Uses of the Compositions of the Invention

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

metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.

- As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.
- Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES**Example A: Polynucleotide And Polypeptide Sequences, And Homology Data****EXAMPLE 1.**

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

Table 1A. NOV1 Sequence Analysis			
	SEQ ID NO: 1	975 bp	
NOV1a, CG101926-01 DNA Sequence	AGCAGGCCACTAGTTTATTAACCTCCAGCCACCTTGATTTTTGCTAAATGAAGACTC TGCAGTCTACACTTCTCCTATTACTGCTTGTGCCCTCTGATAAAGCCAGCACCACCAAC CCAGCAGGACTCACGCATTATCTATGATTATGGAACAGATAATTTGAAGAATCCATA TTTAGCCAAGATTATGAGGATAAATACCTGGATGGAAAAATATTAAGGAAAAAGAAA CTGTGATAATACCCAATGAGAAAAGTCTTCAATTACAAAAAGATGAGGCAATAACACC ATTACCTCCCAAGAAAGAAAATGATGAAATGCCACGTGCTGCTGTGTGTTTGTTTA AGTGGCTCTGTATACTGTGAAGAAGTTGACATTGATGCTGTACCACCTTACCAAAGG AATCAGCCTATCTTTACGCACGATTCAACAAAATTAAGGCTGACTGCCAAAGATTT TGCAGACATACCTAACTTAAGAAGACTCGATTTTACAGGAAATTTGATAGAAGATATA GAAGATGGTACTTTTCAAAACTTTCTCTGTTAGAAGAACTTTCACTTGTGTGAAATC AACTACTAAAACTTCCAGTTCTTCTCCCAAGCTCACTTTATTAAATGCAAAATACAA CAAAATCAAGAGTAGGGGAATCAAAGCAAATGCATTCAAAAACTGAATAACCTCACC TTCTCTGCTTGGACCATAATGCCCTGGAATCCGTGCCTCTTAATTTACCAGAAAGTC TACGTGTAATTCATCTTCAGTTCAACAACATAGCTTCAATTACAGATGACACATTCTG CAAGGCTAATGACACCAAGTTACATCCGGGACCGCATTGAAGAGATACGCCCTGGAGGC AATCCAATCGTCTTGGGAAAGCATCCAAACAGTTTATTGCTTAAAAAGATTACCGA TAGGGTCATACTTTTAACTCTATTGGTACAACATATAAATGAAAGT		
	ORF Start: ATG at 49	ORF Stop: TAA at 943	
	SEQ ID NO: 2	298 aa	MW at 33861.6kD
NOV1a, CG101926-01 Protein Sequence	MKTQLSTLLLLLLVPLIKPAPPTQQDSRIIYDGTDFNFEESIFSQDYEDKYLDGKNIK EKETVIIIPNEKSLQLQKDEAITPLPPKKENDEMPCLLCVCLSGSVYCEEVDIDAVPP LPKESAYLYARFNKIKKLTAKDFADIPNLRLDFTGNLIEDGTFSKLSLLEELSL AENQLKLPLVLPKLTFLNAKYNKIKSRGIKANAFKKLNNLTFLCLDHNALESVPLNL PESLRVIHLQFNNIASITDDTFCKANDTSYIRDRIEIRLEGNPIVLGKHPNSFICLK RLPIGSYF		
	SEQ ID NO: 3	849 bp	
NOV1b, 207639361 DNA Sequence	GGATCCGCACCACCAACCCAGCAGGACTCACGCATTATCTATGATTATGGAACAGATA ATTTTGAAGAATCCATATTTAGCCAAGATTATGAGGATAAATACCTGGATGGAAAAAA TATTAAGGAAAAAGAACTGTGATAATACCCAATGAGAAAAGTCTTCAATTACAAAAA GATGAGGCAATAACACCATTACCTCCCAAGAAAGAAAATGATGAAATGCCACGTGTC TGCTGTGTGTTTGTGTTAAGTGGCTCTGTATACTGTGAAGAAGTTGACATTGATGCTGT ACCACCTTACCAAAGGAATCAGCCTATCTTTACGCACGATTCAACAAAATTAAGAAAG CTGACTGCCAAAGATTTTGCAGACATACCTAACTTAAGAAGACTCGATTTTACAGGAA ATTTGATAGAAGATATAGAAGATGGTACTTTTCAAAACTTTCTCTGTGTTAGAAGAACT TTCACTTGCTGAAAATCAACTACTAAAACTTCCAGTTCTTCTCCCAAGCTCACTTTA TTTAATGCAAAATACAACAAAATCAAGAGTAGGGGAATCAAAGCAAATGCATTCAAAA AACTGAATAACCTCACCTTCTCTGCTTGGACCATAATGCCCTGGAATCCGTGCCTCT TAATTTACCAGAAAGTCTACGTGTAATTCATCTTCAGTTCAACAACATAGCTTCAATT ACAGATGACACATTCTGCAAGGCTAACGACACCAAGTTACATCCGGGACCGCATTGAAG AGATACGCCTGGAGGGCAATCCAATCGTCTTGGGAAAGCATCCAAACAGTTTATTATTG CTTAAAAAGATTACCGATAGGGTCATACTTTCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 4	283 aa	MW at 32131.3kD
NOV1b, 207639361	GSAPPTQQDSRIIYDGTDFNFEESIFSQDYEDKYLDGKNIKEKETVIIIPNEKSLQLQK DEAITPLPPKKENDEMPCLLCVCLSGSVYCEEVDIDAVPPLPKESAYLYARFNKIKK LTAKDFADIPNLRLDFTGNLIEDGTFSKLSLLEELSLAENQLKLPLVLPKLTFL		

Protein Sequence	FNAKYNKIKSRGIKANAFKKLNNLTFLCLDHNALESVPLNLPESLRVIHLQFNNIASI TDDTFCKANDTSYIRDRIEEIRLEGNPIVLGKHPNSFICLKRLPIGSYFLE
-------------------------	---

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

Table 1B. Comparison of NOV1a against NOV1b.		
Protein Sequence	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV1b	20..298	251/279 (89%)
	3..281	251/279 (89%)

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

Table 1C. Protein Sequence Properties NOV1a	
PSort analysis:	0.4610 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 20 and 21

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

Table 1D. Geneseq Results for NOV1a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB90583	Human secreted protein, SEQ ID NO: 121 - Homo sapiens, 298 aa. [WO200121658-A1, 29-MAR-2001]	1..298 1..298	297/298 (99%) 297/298 (99%)	e-171
AAB90549	Human secreted protein, SEQ ID NO: 87 - Homo sapiens, 298 aa. [WO200121658-A1, 29-MAR-2001]	1..298 1..298	297/298 (99%) 297/298 (99%)	e-171
AAE03780	Human gene 17 encoded secreted protein HSDEE58, SEQ ID NO:50 - Homo sapiens, 298 aa. [WO200132837-A1, 10-MAY-2001]	1..298 1..298	297/298 (99%) 297/298 (99%)	e-171
AAM38687	Human polypeptide SEQ ID NO 1832 - Homo sapiens, 298 aa. [WO200153312-A1, 26-JUL-2001]	1..298 1..298	297/298 (99%) 297/298 (99%)	e-171
AAM40473	Human polypeptide SEQ ID NO 5404 - Homo sapiens, 307 aa. [WO200153312-A1, 26-JUL-2001]	1..298 10..307	296/298 (99%) 297/298 (99%)	e-170

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

Table 1E. Public BLASTP Results for NOV1a				
Protein Accession Number	Protein/Organism/Length	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P20774	Osteoinductive factor precursor (OIF) (Osteoglycin) (Mimecan) - Homo sapiens (Human), 298 aa.	1..298 1..298	297/298 (99%) 297/298 (99%)	e-170
P19879	Osteoinductive factor precursor (OIF) (Osteoglycin) - Bos taurus (Bovine), 299 aa.	1..297 1..298	276/298 (92%) 287/298 (95%)	e-160
O18818	MIMECAN - Bos taurus (Bovine), 299 aa.	1..297 1..298	275/298 (92%) 286/298 (95%)	e-159
AAM46865	Osteoglycin - Oryctolagus cuniculus (Rabbit), 298 aa.	1..298 1..298	261/298 (87%) 280/298 (93%)	e-152
Q62000	Osteoglycin precursor - Mus musculus (Mouse), 298 aa.	1..298 1..298	253/298 (84%) 273/298 (90%)	e-145

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

Table 1F. Domain Analysis of NOV1a			
Pfam Domain	NOV1a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LRR	144..167	7/25 (28%) 21/25 (84%)	0.0073
LRR	235..258	6/25 (24%) 20/25 (80%)	0.64

5

EXAMPLE 2.

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

Table 2A. NOV2 Sequence Analysis			
	SEQ ID NO: 5	126 bp	
NOV2a, CG109754-02 DNA Sequence	GCAGACAAACCAGACGTGGGGGAATCGCCAGCTTCAATAGGGCCAAGCTGAAGAAAA CGGAGACGCAGGAGAAGAACACCCCTGCCGACCAAGAGACCACTGGGCAGAAGCGGAG TGAAATTTCC		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 6	42 aa	MW at 4604.1kD

NOV2a, CG109754-02 Protein Sequence	ADKPDVGGIASFNRAKLKKTETQEKNTLPTKETTGQKRSEIS
---	--

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

Table 2B. Protein Sequence Properties NOV2a	
PSort analysis:	0.5600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

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

5

Table 2C. Geneseq Results for NOV2a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY80267	Thymosin beta 4 peptide isoform Tbeta10 - Unidentified, 43 aa. [WO200006190-A1, 10-FEB-2000]	1..42 1..43	36/43 (83%) 39/43 (89%)	4e-12
AAR96932	Thymosin beta 10 - Synthetic, 43 aa. [WO9611016-A1, 18-APR-1996]	1..42 1..43	36/43 (83%) 39/43 (89%)	4e-12
AAR27109	N-terminus of thymosin beta peptide - Synthetic, 43 aa. [JP04234325-A, 24-AUG-1992]	1..42 1..43	35/43 (81%) 39/43 (90%)	9e-12
AAY80270	Thymosin beta 4 peptide isoform Tbeta13 - Unidentified, 41 aa. [WO200006190-A1, 10-FEB-2000]	1..38 1..38	30/38 (78%) 35/38 (91%)	1e-10
AAR96935	Thymosin beta 13 - Synthetic, 41 aa. [WO9611016-A1, 18-APR-1996]	1..38 1..38	30/38 (78%) 35/38 (91%)	1e-10

In a BLAST search of public sequence databases, the NOV2a protein was found to have homology to the proteins shown in the BLASTP data in Table 2D.

Table 2D. Public BLASTP Results for NOV2a				
Protein Accession Number	Protein/Organism/Length	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAM34215	Thymosin beta 10 - Equus caballus (Horse), 44 aa.	1..42 2..44	36/43 (83%) 39/43 (89%)	9e-12
P13472	Thymosin beta-10 - Homo sapiens (Human), 43 aa.	1..42 1..43	36/43 (83%) 39/43 (89%)	9e-12
P21753	Thymosin beta-9 - Sus scrofa (Pig), 41 aa.	1..38 1..38	29/38 (76%) 34/38 (89%)	1e-09
P21752	Thymosin beta-9 (Thymosin beta-10) [Contains: Thymosin beta-8] - Bos taurus (Bovine), 41 aa.	1..38 1..38	29/38 (76%) 34/38 (89%)	1e-09
O76538	Thymosin beta - Strongylocentrotus purpuratus (Purple sea urchin), 41 aa.	1..39 2..40	26/39 (66%) 34/39 (86%)	4e-09

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

Table 2E. Domain Analysis of NOV2a			
Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Thymosin	1..40	29/41 (71%) 36/41 (88%)	1.8e-16

EXAMPLE 3.

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

Table 3A. NOV3 Sequence Analysis			
	SEQ ID NO: 7	84 bp	
NOV3a, CG114834-02 DNA Sequence.	TCAGATGCAGCTGTAGACACCAGCTCTGAAATCATTGCCAAGGACTTAAAGGAGAAGAAGGAAGTTGTGAAAGAGGCGGAAAAT		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 8	28 aa	MW at 3047.3kD
NOV3a, CG114834-02 Protein Sequence	SDAAVDTSSSEIIAKDLKEKKEVVKEAEN		

Further analysis of the NOV3a protein yielded the following properties shown in

5 Table 3B.

Table 3B. Protein Sequence Properties NOV3a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 3C. Geneseq Results for NOV3a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB67858	Amino acid sequence of a human polypeptide designated PTMA-1 - Homo sapiens, 109 aa. [WO200123572-A2, 05-APR-2001]	1..28 2..29	28/28 (100%) 28/28 (100%)	6e-08
AAM03072	Peptide #1754 encoded by probe for measuring breast gene expression - Homo sapiens, 70 aa. [WO200157270-A2, 09-AUG-2001]	1..28 2..29	28/28 (100%) 28/28 (100%)	6e-08
AAM27792	Peptide #1829 encoded by probe for measuring placental gene expression - Homo sapiens, 70 aa. [WO200157272-A2, 09-AUG-2001]	1..28 2..29	28/28 (100%) 28/28 (100%)	6e-08
AAM15317	Peptide #1751 encoded by probe for measuring cervical gene expression - Homo sapiens, 70 aa. [WO200157278-A2, 09-AUG-2001]	1..28 2..29	28/28 (100%) 28/28 (100%)	6e-08
AAM67501	Human bone marrow expressed probe encoded protein SEQ ID NO: 27807 - Homo sapiens, 70 aa. [WO200157276-A2, 09-AUG-2001]	1..28 2..29	28/28 (100%) 28/28 (100%)	6e-08

In a BLAST search of public sequence databases, the NOV3a protein was found to have homology to the proteins shown in the BLASTP data in Table 3D.

Table 3D. Public BLASTP Results for NOV3a				
Protein Accession Number	Protein/Organism/Length	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
S15073	prothymosin alpha - mouse, 111 aa.	1..28 2..29	25/28 (89%) 26/28 (92%)	4e-06
C33356	prothymosin alpha homolog (clone 32) - human, 59 aa (fragment).	1..28 2..29	25/28 (89%) 26/28 (92%)	4e-06
TNRTA	prothymosin alpha - rat, 112 aa.	1..28 2..29	25/28 (89%) 26/28 (92%)	4e-06
AAK30146	Prothymosin alpha - Homo sapiens (Human), 110 aa.	1..28 2..29	25/28 (89%) 26/28 (92%)	4e-06
AAA63238	HUMAN PROTHYMOSIN-ALPHA PSEUDOGENE, COMPLETE SEQUENCE - Homo sapiens (Human), 109 aa.	1..28 2..29	25/28 (89%) 26/28 (92%)	4e-06

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

Table 3E. Domain Analysis of NOV3a			
Pfam Domain	NOV3a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

EXAMPLE 4.

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

Table 4A. NOV4 Sequence Analysis		
	SEQ ID NO: 9	657 bp
NOV4a, CG124728-02 DNA Sequence	CCGGTGCCAGCGCTATGAGGCCACTCCTCGTCTGCTGCTCCTGGGCTGGCGGCCGG CTCGCCCCCACTGGACGACAACAAGATCCCCAGCCTCTGCCCGGGGCACCCCGGCCTT CCAGGACTGCCGGGACCTCGAGGGGACCCCGGGCCGCGAGGAGAGCGGGACCCGCGG GGCCACCGGGCTGCCGGGGAGTGCTCGGTGCCCTCCGCGATCCGCCTTCAGCGCCAA GCGCTCCGAGAGCCGGGTGCCTCCGCGTCTGACGCACCCTTGCCCTTCGACCGCGTG CTGGTGAACGAGCAGGGACATTACGACGCCGTCACCGCAAGTTCACCTGCCAGGTGC CTGGGTCTACTACTTCGCCGTCCATGCCACCGTCTACCGGGCCAGCCTGCAGTTTGA TCTGGTGAAGAATGGCGAATCCATTGCCTCTTTCTTCCAGTTTTTCGGGGGGTGCCCC AAGCCAGCCTCGCTCTCGGGGGGGCCATGGTGAGGCTGGAGCCTGAGGACCAAGTGT GGGTGCAGGTGGGTGTGGGTGACTACATTGGCATCTATGCCAGCATCAAGACAGACAG CACCTTCTCCGGATTTCTGGTGACTCCGACTGGCGCAGCTCCCAGTCTTTGCTTAG	

	<u>TGCCCCACTGCAAAGTGAGC</u>		
	ORF Start: ATG at 15	ORF Stop: TAG at 636	
	SEQ ID NO: 10	207 aa	MW at 21887.6kD
NOV4a, CG124728-02 Protein Sequence	MRPLLVLVLLGLAAGSPPLDDNKIPSLCPGHPGLPGLPGPRGDPGPRGEAGPAGPTGP AGECSVPPRFAFSAKRSESRVPPPSDAPLPFDRVLVNEQGHYDAVTGKFTCQVPGVYY FAVHATVYRASLQFDLVKNESIASFFQFFGGWPKPASLSGGAMVRLEPEDQVWVQVG VGDIYIGIYASIKTDTSTFSGLVYSDWRSSPVFA		
	SEQ ID NO: 11	643 bp	
NOV4b, CG124728-03 DNA Sequence	<u>CACCGGATCCACCATGAGGCCACTCCTCGTCCTGCTGCTCCTGGGCCTGGCGGCCGGC</u> TCGCCCCCACTGGACGACAACAAGATCCCCAGCCTCTGCCCGGGGCACCCCGGCCTTC CAGGACTGCCGGGACCTCGAGGGGACCCCGGGCCGCGAGGAGAGCGGGACCCGCGGG GCCACCGGGCCTGCCGGGGAGTGCTCGGTGCCCTCCGCGATCCGCCTTACGCGCCAAG CGCTCCGAGAGCCGGGTGCCTCCGCCGTCTGACGCACCCCTTGCCCTTCGACCGCGTGC TGGTGAACGAGCAGGGACATTACGACGCCGTACCGGCAAGTTACCTGCCAGGTGCC TGGGTCTACTACTTCGCCGTCCATGCCACCGTCTACCGGGCCAGCCTGCAGTTTGAT CTGGTGAAGAATGGCGAATCCATTGCCTCTTTCTTCCAGTTTTTCGGGGGGTGGCCCA AGCCAGCCTCGCTCTCGGGGGGGGCCATGGTGAGGCTGGAGCCTGAGGACCAAGTGTG GGTGCAGGTGGGTGTGGGTGACTACATTGGCATCTATGCCAGCATCAAGACAGACAGC ACCTTCTCCGATTCTTGGTGTACTCCGACTGGCGCAGCTCCCCAGTCTTTGCTGT <u>TCG</u> <u>ACGGC</u>		
	ORF Start: ATG at 14	ORF Stop: at 635	
	SEQ ID NO: 12	207 aa	MW at 21887.6kD
NOV4b, CG124728-03 Protein Sequence	MRPLLVLVLLGLAAGSPPLDDNKIPSLCPGHPGLPGLPGPRGDPGPRGEAGPAGPTGP AGECSVPPRFAFSAKRSESRVPPPSDAPLPFDRVLVNEQGHYDAVTGKFTCQVPGVYY FAVHATVYRASLQFDLVKNESIASFFQFFGGWPKPASLSGGAMVRLEPEDQVWVQVG VGDIYIGIYASIKTDTSTFSGLVYSDWRSSPVFA		
	SEQ ID NO: 13	643 bp	
NOV4c, 263479529 DNA Sequence	<u>CACCGGATCCACCATGAGGCCACTCCTCGTCCTGCTGCTCCTGGGCCTGGCGGCCGGC</u> TCGCCCCCACTGGACGACAACAAGATCCCCAGCCTCTGCCCGGGGCACCCCGGCCTTC CAGGACTGCCGGGACCTCGAGGGGACCCCGGGCCGCGAGGAGAGCGGGACCCGCGGG GCCACCGGGCCTGCCGGGGAGTGCTCGGTGCCCTCCGCGATCCGCCTTACGCGCCAAG CGCTCCGAGAGCCGGGTGCCTCCGCCGTCTGACGCACCCCTTGCCCTTCGACCGCGTGC TGGTGAACGAGCAGGGACATTACGACGCCGTACCGGCAAGTTACCTGCCAGGTGCC TGGGGTCTACTACTTCGCCGTCCATGCCACCGTCTACCGGGCCAGCCTGCAGTTTGAT CTGGTGAAGAATGGCGAATCCATTGCCTCTTTCTTCCAGTTTTTCGGGGGGTGGCCCA AGCCAGCCTCGCTCTCGGGGGGGGCCATGGTGAGGCTGGAGCCTGAGGACCAAGTGTG GGTGCAGGTGGGTGTGGGTGACTACATTGGCATCTATGCCAGCATCAAGACAGACAGC ACCTTCTCCGATTCTTGGTGTACTCCGACTGGCGCAGCTCCCCAGTCTTTGCTGT <u>TCG</u> <u>ACGGC</u>		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 14	214 aa	MW at 22505.2kD
NOV4c, 263479529 Protein Sequence	TGSTMRLVLVLLGLAAGSPPLDDNKIPSLCPGHPGLPGLPGPRGDPGPRGEAGPAG PTGPAGECSVPPRFAFSAKRSESRVPPPSDAPLPFDRVLVNEQGHYDAVTGKFTCQVP GVYYFAVHATVYRASLQFDLVKNESIASFFQFFGGWPKPASLSGGAMVRLEPEDQVW VQVGVDYIGIYASIKTDTSTFSGLVYSDWRSSPVFAVDG		
	SEQ ID NO: 15	595 bp	
NOV4d, 271674589 DNA Sequence	<u>CACCGGATCCTCGCCCCACTGGACGACAACAAGATCCCCAGCCTCTGCCCGGGGCAC</u> CCCGGCCTTCCAGGACTGCCGGGACCTCGAGGGGACCCCGGGCCGCGAGGAGAGCGGG GACCCCGGGGGCCACCGGGCCCTGCCGGGGAGTGCTCGGTGCCCTCCGCGATCCGCCTT CAGCGCCAAGCGCTCCGAGAGCCGGGTGCCTCCGCCGTCTGACGCACCCCTTGCCCTTC GACCGCGTGTGGTGAACGAGCAGGGACATTACGACGCCGTACCGGCAAGTTACCT GCCAGGTGCCCTGGGTCTACTACTTCGCCGTCCATGCCACCGTCTACCGGGCCAGCCT GCAGTTTGATCTGGTGAAGAATGGCGAATCCATTGCCTCTTTCTTCCAGTTTTTCGGG GGTGGCCCAAGCCAGCCTCGCTCTCGGGGGGGGCCATGGTGAGGCTGGAGCCTGAGG ACCAAGTGTGGGTGCAGGTGGGTGTGGGTGACTACATTGGCATCTATGCCAGCATCAA GACAGACAGCACCTTCTCCGATTCTTGGTGTACTCCGACTGGCGCAGCTCCCCAGTC TTTGCTGTGACGGC		

	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 16	198 aa	MW at 20872.2kD
NOV4d, 271674589 Protein Sequence	TGSSPPLDDNKIPSLCPGHPGLPGLPGPRGDPGPRGEAGPAGPTGPAGECSVPPPSAF SAKRSESRVPPPSDAPLPFDRVLVNEQGHYDAVTGKFTCQVPGVYFVAVHATVYRASL QFDLVKNGESIASFQFFGGWPKPASLSGGAMVRLEPEDQVWVQVGVDYIGIYASIK TDSTFSGFLVYSDWRSSEPVFAVDG		

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

Table 4B. Comparison of NOV4a against NOV4b through NOV4d.		
Protein Sequence	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV4b	61..207	147/147 (100%)
	61..207	147/147 (100%)
NOV4c	61..207	147/147 (100%)
	65..211	147/147 (100%)
NOV4d	16..207	156/192 (81%)
	4..195	156/192 (81%)

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

Table 4C. Protein Sequence Properties NOV4a	
PSort analysis:	0.6042 probability located in outside; 0.2159 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 16 and 17

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

Table 4D. Geneseq Results for NOV4a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB49599	Human adipocyte complement related protein homolog zsig39 - Homo sapiens, 243 aa. [WO200073446-A2, 07-DEC-2000]	1..207 1..243	206/243 (84%) 206/243 (84%)	e-116
AAB49593	Human adipocyte complement related protein homolog zsig39 - Homo sapiens, 243 aa. [WO200073444-A1, 07-DEC-2000]	1..207 1..243	206/243 (84%) 206/243 (84%)	e-116
AAB65888	Human secreted protein related protein SEQ ID NO: 102 - Homo sapiens, 243 aa. [WO200078808-A1, 28-DEC-2000]	1..207 1..243	206/243 (84%) 206/243 (84%)	e-116
AAB65815	Human TANGO 253 SEQ ID NO: 3 - Homo sapiens, 243 aa. [WO200078808-A1, 28-DEC-2000]	1..207 1..243	206/243 (84%) 206/243 (84%)	e-116
AAB65816	Human mature TANGO 253 SEQ ID NO: 4 - Homo sapiens, 228 aa. [WO200078808-A1, 28-DEC-2000]	16..207 1..228	191/228 (83%) 191/228 (83%)	e-108

In a BLAST search of public sequence databases, the NOV4a protein was found to have homology to the proteins shown in the BLASTP data in Table 4E.

Table 4E. Public BLASTP Results for NOV4a				
Protein Accession Number	Protein/Organism/Length	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAH29485	C1q and tumor necrosis factor related protein 5 - Homo sapiens (Human), 243 aa.	1..207 1..243	206/243 (84%) 206/243 (84%)	e-116
Q9BXJ0	Complement-c1q tumor necrosis factor-related protein 5 precursor - Homo sapiens (Human), 243 aa.	1..207 1..243	206/243 (84%) 206/243 (84%)	e-116
Q8R002	Similar to DKFZP586B0621 protein (Hypothetical 25.4 kDa protein) - Mus musculus (Mouse), 243 aa.	1..207 1..243	190/243 (78%) 197/243 (80%)	e-107
T14782	hypothetical protein DKFZp586B0621.1 - human, 219 aa (fragment).	25..207 1..219	182/219 (83%) 182/219 (83%)	e-102
BAB84561	Otolin-1 - Oncorhynchus keta (Chum salmon), 508 aa.	30..199 338..505	74/171 (43%) 104/171 (60%)	2e-31

PFam analysis predicts that the NOV4a protein contains the domains shown in the Table 4F.

Table 4F. Domain Analysis of NOV4a			
Pfam Domain	NOV4a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Collagen	15..74	24/60 (40%) 37/60 (62%)	0.012
C1q	69..196	50/138 (36%) 90/138 (65%)	2.1e-34

5

EXAMPLE 5.

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

Table 5A. NOV5 Sequence Analysis		
	SEQ ID NO: 17	500 bp
NOV5a, CG127616-01	CCGGAGCCGGACCGGGGCCACCGCGCCCGCTCTGCTCCGACACCGCGCCCGCTGGAC AGCCGCCCTCTCCTCCAGGCCCGTGGGGCTGGCCCTGCACCGCCGAGCTTCCCGGGAT GAGGGCCCCCGGTGTGGTCACCGCGCGCCCGAGGTCGCTGAGGGACCCGGCCAGG	

DNA Sequence	CGCGGAGATGGGGGTGCACGAATGTCCTGCCTGGCTGTGGCTTCTCCTGTCCCTGCTGTCGCTCCCTCTGGGCCTCCAGTCTCTGGGCGCCCCACCACGCCTCATCTGTGACAGCCGAGTCCTGGAGAGGTACCTCTTGGAGGCCAAGGAGGCCGAGAATATCACGAAGGAAGCATCTCCCCCTCCAGATGCGGCCCTCAGCTGCTCCACTCCGAACAATCACTGCTGACACTTTCCGCAAACCTCTCCGAGTCTACTCCAATTTCTCCGGGGAAAGCTGAAGCTGTACACAGGGGAGGCCTGCAGGACAGGGGACAGATGACCAG		
	ORF Start: ATG at 182	ORF Stop: TGA at 494	
	SEQ ID NO: 18	104 aa	MW at 11567.4kD
NOV5a, CG127616-01 Protein Sequence	MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRILICDSRVLERYLLEAKEAENITKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLYTGEACRTGDR		
	SEQ ID NO: 19	324 bp	
NOV5b, CG127616-02 DNA Sequence	CCTGGCTATCTGTTCTAGAATGTCCTGCCTGGCTGTGGCTTCTCCTGTCCCTGCTGTCGCTCCCTCTGGGCCTCCAGTCTCTGGGCGCCCCACCACGCCTCATCTGTGACAGCCGATCCTGGAGAGGTACCTCTTGGAGGCCAAGGAGGCCGAGAATATCACGAAGGAAGCCATCTCCCCCTCCAGATGCGGCCCTCAGCTGCTCCACTCCGAACAATCACTGCTGACACTTTCCGAAAACCTCTCCGAGTCTACTCCAATTTCTCCGGGGAAAGCTGAAGCTGTACACAGGGGAGGCCTGCAGGACAGGGGACAGATGACCAG		
	ORF Start: at 3	ORF Stop: TGA at 318	
	SEQ ID NO: 20	105 aa	MW at 11741.6kD
NOV5b, CG127616-02 Protein Sequence	WLSVLECPAWLWLLLSLLSLPLGLPVLGAPPRILICDSRVLERYLLEAKEAENITKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLYTGEACRTGDR		
	SEQ ID NO: 21	282 bp	
NOV5c, 214374151 DNA Sequence	GGATCCTGGCTTCTCCTGTCCCTGCTGTGCTCCCTCTGGGCCTCCAGTCTCTGGGCGCCCCACCACGCCTCATCTGTGACAGCCGAGTCCTGGAGAGGTACCTCTTGGAGGCCAAGGAGGCCGAGAATATCACGAAGGAAGCCATCTCCCCCTCCAGATGCGGCCCTCAGCTGCTCCACTCCGAACAATCACTGCTGACACTTTCCGCAAACCTCTCCGAGTCTACTCCAATTTCTCCGGGGAAAGCTGAAGCTGTACACAGGGGAGGCCTGCAGGCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 22	94 aa	MW at 10400.0kD
NOV5c, 214374151 Protein Sequence	GSWLLLSLLSLPLGLPVLGAPPRILICDSRVLERYLLEAKEAENITKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLYTGEACRL		
	SEQ ID NO: 23	333 bp	
NOV5d, 219936857 DNA Sequence	CGCGGATCCATGGGGGTGCACGAATGTCCTGCCTGGCTGTGGCTTCTCCTGTCCCTGCTGTCGCTCCCTCTGGGCCTCCAGTCTCTGGGCGCCCCACCACGCCTCATCTGTGACAGCCGAGTCCTGGAGAGGTACCTCTTGGAGGCCAAGGAGGCCGAGAATATCACGAAGGAAGCCATCTCCCCCTCCAGATGCGGCCCTCAGCTGCTCCACTCCGAACAATCACTGCTGACACTTTCCGCAAACCTCTCCGAGTCTACTCCAATTTCTCCGGGGAAAGCTGAAGCTGTACACAGGGGAGGCCTGCAGGACAGGGGACAGATGACTCGAGCGG		
	ORF Start: at 1	ORF Stop: TGA at 322	
	SEQ ID NO: 24	107 aa	MW at 11867.7kD
NOV5d, 219936857 Protein Sequence	RSGMGVHECPAWLWLLLSLLSLPLGLPVLGAPPRILICDSRVLERYLLEAKEAENITKEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLYTGEACRTGDR		
	SEQ ID NO: 25	330 bp	
NOV5e, 259333914 DNA Sequence	CGCGGATCCATGGGGGTGCACGAATGTCCTGCCTGGCTGTGGCTTCTCCTGTCCCTGCTGTCGCTCCCTCTGGGCCTCCAGTCTCTGGGCGCCCCACCACGCCTCATCTGTGACAGCCGAGTCCTGGAGAGGTACCTCTTGGAGGCCAAGGAGGCCGAGAATATCACGAAGGAAGCCATCTCCCCCTCCAGATGCGGCCCTCAGCTGCTCCACTCCGAACAATCACTGCTGACACTTTCCGCAAACCTCTCCGAGTCTACTCCAATTTCTCCGGGGAAAGCTGAAGCTGTACACAGGGGAGGCCTGCAGGACAGGGGACAGATGACTCGAGCGG		

Sequence	CTTTCGCGAAACTCTTCCGAGTCTACTCCAATTTCTCTCCGGGGAAAGCTGAAGCTGTA CACAGGGGAGGCCTGCAGGACAGGGGACAGACTCGAGCGG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 26	110 aa	MW at 12266.1kD
NOV5e, 259333914 Protein Sequence	RGSMDVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITKE AISPPDAASAAPLRTTITADTFRKLFRVYSNFLRGKLLKLYTGEACRTGDR		
	SEQ ID NO: 27	333 bp	
NOV5f, 219936857 DNA Sequence	CGCGGATCCATGGGGGTGCACGAATGTCCTGCCTGGCTGTGGCTTCTCCTGTCCCTGC TGTCGCTCCCTCTGGGCCTCCCAGTCTCTGGGCGCCACCACGCCTCATCTGTGACAG CCGAGTCTTGGAGAGGTACCTCTTGGAGGCCAAGGAGGCCGAGAATATCACGAAGGAA GCCATCTCCCTCCAGATGCGGCCTCAGCTGCTCCACTCCGAACAATCACTGCTGACA CTTTCGCAAACCTCTCCGAGTCTACTCCAATTTCTCTCCGGGAAAGCTGAAGCTGTA CACAGGGGAGGCCTGCAGGACAGGGGACAGATGACTCGAGCGG		
	ORF Start: at 1	ORF Stop: TGA at 322	
	SEQ ID NO: 28	107 aa	MW at 11867.7kD
NOV5f, 219936857 Protein Sequence	RGSMDVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITKE AISPPDAASAAPLRTTITADTFRKLFRVYSNFLRGKLLKLYTGEACRTGDR		

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

Table 5B. Comparison of NOV5a against NOV5b through NOV5f.		
Protein Sequence	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV5b	33..104 34..105	60/72 (83%) 60/72 (83%)
NOV5c	33..100 25..92	56/68 (82%) 56/68 (82%)
NOV5d	1..104 4..107	66/104 (63%) 66/104 (63%)
NOV5e	1..104 4..107	66/104 (63%) 66/104 (63%)
NOV5f	1..104 4..107	66/104 (63%) 66/104 (63%)

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

Table 5C. Protein Sequence Properties NOV5a	
PSort analysis:	0.8096 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 28 and 29

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

Table 5D. Geneseq Results for NOV5a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW14143	Erythropoietin variant JM - Homo sapiens, 193 aa. [WO9708307-A1, 06-MAR-1997]	1..54 1..54	54/54 (100%) 54/54 (100%)	1e-25
AAE15348	Human erythropoietin (Epo) N47-Fc fusion protein - Homo sapiens, 420 aa. [WO200181405-A2, 01-NOV-2001]	1..53 1..53	53/53 (100%) 53/53 (100%)	5e-25
AAE15341	Human erythropoietin (Epo) protein - Homo sapiens, 193 aa. [WO200181405-A2, 01-NOV-2001]	1..53 1..53	53/53 (100%) 53/53 (100%)	5e-25
AAB35017	Chimpanzee erythropoietin fragment SEQ ID NO: 49 - Pan sp, 193 aa. [WO200068376-A1, 16-NOV-2000]	1..53 1..53	53/53 (100%) 53/53 (100%)	5e-25
AAB35016	Chimpanzee erythropoietin fragment SEQ ID NO: 48 - Pan sp, 193 aa. [WO200068376-A1, 16-NOV-2000]	1..53 1..53	53/53 (100%) 53/53 (100%)	5e-25

In a BLAST search of public sequence databases, the NOV5a protein was found to have homology to the proteins shown in the BLASTP data in Table 5E.

Table 5E. Public BLASTP Results for NOV5a				
Protein Accession Number	Protein/Organism/Length	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P01588	Erythropoietin precursor (Epoetin) - Homo sapiens (Human), 193 aa.	1..53 1..53	53/53 (100%) 53/53 (100%)	1e-24
P07865	Erythropoietin precursor - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 192 aa.	1..53 1..53	50/53 (94%) 52/53 (97%)	2e-23
Q28513	Erythropoietin precursor - Macaca mulatta (Rhesus macaque), 192 aa.	1..53 1..53	49/53 (92%) 52/53 (97%)	3e-23
CAC41224	Sequence 1 from Patent WO0136489 - Homo sapiens (Human), 166 aa (fragment).	41..104 103..166	54/64 (84%) 54/64 (84%)	1e-22
P29676	Erythropoietin precursor - Rattus norvegicus (Rat), 192 aa.	54..104 142..192	44/51 (86%) 46/51 (89%)	9e-19

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

Table 5F. Domain Analysis of NOV5a			
Pfam Domain	NOV5a Match Region	Identities/ Similarities for the Matched Region	Expect Value
EPO_TPO	11..100	50/180 (28%) 89/180 (49%)	2.3e-10

5 EXAMPLE 6.

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

Table 6A. NOV6 Sequence Analysis		
	SEQ ID NO: 29	684 bp
NOV6a, CG128348-01 DNA Sequence	CAATCACAGGCAGGAAGATGAAGGTTCTGTGGGCTGCGTTGCTGGTCACATTTCCTGGC AGGATGCCAGGCCAAGGTGGAGCAAGCGGTGGAGACAGAGCCGGAGCCCGAGCTGCGC CAGCAGACCGAGTGGCAGAGCGGCCAGCGCTGGGAAC TGCCACTGGGTTCGCTTTTGGG ATTACCTGCGCTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGTGTCTCAGCTC CCAGGTCACCCAGGAAC TGAGGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCC TACAAATCGGAAC TGAGGAACAAC TGACCCCGGTGGCGGAGGAGACGCGGGCACGGC TGTC AAGGAGCTGCAGGCGGCGCAGGCCCGGTGGGCGCGGACGAGGTGAAGGGCA GGTGGCGGAGGTGCGCGCCAAGCTGGAGGAGCAGGCCAGCAGATACGCCTGCAGGCC GAGGCCTTCCAGGCCCGCCTCAAGAGCTGGTTTCGAGCCCTGGTGGAAGACATGCAGC GCCAGTGGGCGGGCTGGTGGAGAAGGTGCAGGCTGCCGTGGGCACCAAGCGCGCCC TGTGCCCAGCGACAATCACTGAACGCCGAAGCCTGCAGCCATGCGACCCACGCGACC	

	<u>CCGTGCCTCCTGCCTCCGCGCAGCCTGCAGCGGGAGACCCTGTCCC</u>		
	ORF Start: ATG at 18	ORF Stop: TGA at 600	
	SEQ ID NO: 30	194 aa	MW at 22147.9kD
NOV6a, CG128348-01 Protein Sequence	MKVLWAALLVTFLAGCQAKVEQAVETEPEPELRQQTWQSGQRWELALGRFWDYLRWV QTLSEQVQEELLSSQVTQELRALMDETMKELKAYKSELEEQLTPVAEETRARLSKELQ AAQARLGADDEVKGVQAEVRAKLEEQAQQIRLQAEAFQARLKSWEFLVEDMQRQWAGL VEKVQAAVGTSAAPVPSDNH		
	SEQ ID NO: 31	802 bp	
NOV6b, CG128348-02 DNA Sequence	<u>CAATCACAGGCAGGAAGATGAAGGTTCTGTGGGCTGCGTTGCTGGTCACATTCTTGGC</u> AGGATGCCAGGCCAAGGTGGAGCAAGCGGTGGAGACAGAGCCGAGCCCGAGCTGCGC CAGCAGACCGAGTGGCAGAGCGGCCAGCGCTGGGAAGTGGCACTGGGTGCGTTTGGG ATTACCTGCGCTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCAGCTC CCAGGTACCCAGGAAGTGAAGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCC TACAAATCGGAAGTGGAGGAACAAGTGAACCCCGGTGGCGGAGGAGACGCGGGCCAGCG TGTCCAAGGAGCTGGGGCCCCCTCAAGAGCTGGTTGAGCCCTTGGTGAAGACATGCAG CTTCTTGGCCGGCCAGCCGCTACAGGAGCGGGCCAGGCCCTGGGGCGAGCGGCTGCGC GCGCGGATGGAGGAGATGGGCGAGCCGACCCGCGACCCGCTGGACGAGGTGAAGGAGC AGGTGGCGGAGGTGCGGCCAAGCTGGAGGAGCAGGCCAGCAGATACGCTGCAAGGC CGAGGCCCTTCCAGGCCCGCTCAAGAGCTGGTTGAGCCCTTGGTGAAGACATGCAG CGCCAGTGGGCCGGGCTGGTGGAGAAGGTGCAGGCTGCCGTGGGCACCAAGCGCCGCC CTGTGCCAGCGACAATCACTGAACGCCGAAGCCTGCAGCCATCGGACCCACGCCAC CCCGTGCCTCCTGCCTCCGCGCAGCCTGCAGCGGGAGACCCTGTCCCT		
	ORF Start: ATG at 18	ORF Stop: TGA at 717	
	SEQ ID NO: 32	233 aa	MW at 26743.1kD
NOV6b, CG128348-02 Protein Sequence	MKVLWAALLVTFLAGCQAKVEQAVETEPEPELRQQTWQSGQRWELALGRFWDYLRWV QTLSEQVQEELLSSQVTQELRALMDETMKELKAYKSELEEQLTPVAEETRARLSKELG PLVEQGRVRAATVGFLAGQPLQERAQAWGERLRARMEEMGSRTRDRLDEVKEQVAEVR AKLEEQAQQIRLQAEAFQARLKSWEFLVEDMQRQWAGLVEKVQAAVGTSAAPVPSDN H		
	SEQ ID NO: 33	788 bp	
NOV6c, CG128348-03 DNA Sequence	<u>CAATCACAGGCAGGAAGATGAAGGTTCTGTGGGCTGCGTTGCTGGTCACATTCTTGGC</u> AGGATGCCAGGCCAAGGTGGAGCAAGCGGTGGAGACAGAGCCGAGCCGAGCTGCGC CAGCAGACCGAGTGGCAGAGCGGCCAGCGCTGGGAAGTGGCACTGGGTGCGTTTGGG ATTACCTGCGCTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCAGCTC CCAGGTACCCAGGAAGTGAAGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCC TACAAATCGGAAGTGGAGGAACAAGTGAACCCCGGTGGCGGAGGAGACGCGGGCCAGCG TGTCCAAGGAGCTGCAGGCGGCGCAGGCCCGGCTGGGCGCGGACATGGAGGACGTGCG CGGCCGCTTGGTGCAGTACCGCGCGCAGGTGCAGGCCATGCTCGGCCAGAGCACCAG GAGTGCAGGCTGCGCTCGCTCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCG ATGCCGATGACCTGGAGGAGCAGGCCAGCAGATACGCTGCAAGGCCGAGGCCCTTCCA GGCCCGCTCAAGAGCTGGTTGAGCCCTTGGTGAAGACATGCAGCGCCAGTGGGCC GGGCTGGTGGAGAAGGTGCAGGCTGCCGTGGGCACCAAGCGCCGCCCTGTGCCAGCG ACAATCACTGAACGCCGAAGCCTGCAGCCATGCGACCCACGCCACCCCGTGCCTCCT GCCTCCGCGCAGCCTGCAGCGGGAGACCCTGTCC		
	ORF Start: ATG at 18	ORF Stop: TGA at 705	
	SEQ ID NO: 34	229 aa	MW at 26399.7kD
NOV6c, CG128348-03 Protein Sequence	MKVLWAALLVTFLAGCQAKVEQAVETEPEPELRQQTWQSGQRWELALGRFWDYLRWV QTLSEQVQEELLSSQVTQELRALMDETMKELKAYKSELEEQLTPVAEETRARLSKELQ AAQARLGADMEVGRGLVQYRGVQAMLGQSTEELRVRLASHLRKLRKRLRDADDLE EQAQQIRLQAEAFQARLKSWEFLVEDMQRQWAGLVEKVQAAVGTSAAPVPSDNH		
	SEQ ID NO: 35	718 bp	
NOV6d, 278480724 DNA Sequence	<u>CACCGGATCCATGAAGGTTCTGTGGGCTGCGTTGCTGGTCACATTCTTGGCAGGATGC</u> CAGGCCAAGGTGGAGCAAGCGGTGGAGACAGAGCCGAGCCCGAGCTGCGCCAGCAGCA CCGAGTGGCAGAGCGGCCAGCGCTGGGAAGTGGCACTGGGTGCGTTTGGGATTACCT GCGCTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCAGCTCCCAGGTC ACCCAGGAAGTGAAGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCCTACAAAT CGGAAGTGGAGGAACAAGTGAACCCCGGTGGCGGAGGAGACGCGGGCACGCTGTCCAA GGAGCTGGGGCCCCCTGGTGAACAGGGCCGCTGCGGGCCGCCACTGTGGGCTTCCTG		

	GCCGGCCAGCCGCTACAGGAGCGGGCCAGGCCTGGGGCGAGCGGCTGCGCGCCGCGGA TGGAGGAGATGGGCAGCCGACCCGCGACCGCCTGGACGAGGTGAAGGAGCAGGTGGC GGAGGTGCGCGCCAAGCTGGAGGAGCAGGCCAGCAGATACGCCTGCAGGCCGAGGCC TTCCAGGCCCGCCTCAAGAGCTGGTTCGAGCCCCCTGGTGAAGACATGCAGCGCCAGT GGCCCGGGCTGGTGGAGAAGGTGCAGGCTGCCGTGGGCACCAGCGCCGCCCTGTGCC CAGCGACAATCACCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 36	239 aa	MW at 27287.7kD
NOV6d, 278480724 Protein Sequence	TGSMLVLAALLVTFLLAGCQAKVEQAVETEPEPELRQQTWQSGQRWELALGRFWDYL RWVQTLSEQVQEELLSSQVTQELRALMDETMKELKAYKSELEEQLTPVAEETRRLSK ELGPLVEQGRVRAATVGFLAQPLQERAQAWGERLRARMEEMGSRTRDRLDEVKEQVA EVRAKLEEQAQQIRLQAEAFQARLKSWEPLVEDMQRWAGLVEKVQAAVGTSAAPVP SDNHLEG		
	SEQ ID NO: 37	664 bp	
NOV6e, 278480754 DNA Sequence	CACC GGATCCAAGGTGGAGCAAGCGGTGGAGACAGAGCCGAGCCGAGCTGCGCCAG CAGACCGAGTGGCAGAGCGGCCAGCGCTGGGAACCTGGCACTGGGTGCGCTTTTGGGATT ACCTGCGCTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCAGCTCCCA GGTCAACCCAGGAAGTGGAGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCTTAC AAATCGGAAGTGGAGGAACAACCTGACCCCGGTGGCGGAGGAGACGCGGGCACGGCTGT CCAAGGAGCTGGGGCCCCCTGGTGGAAACAGGGCCGCGTGCGGGGCCCACTGTGGGCTT CCTGGCCGGCCAGCCGCTACAGGAGCGGGCCAGGCTGGGGCGAGCGGCTGCGCGCGC CGGATGGAGGAGATGGGCAGCCGACCCGCGACCCGCTGGACGAGGTGAAGGAGCAGG TGGCGGAGGTGCGCGCCAAGCTGGAGGAGCAGGCCAGCAGATACGCTGCAGGCCGA GGCCTTCCAGGCCCGCTCAAGAGCTGGTTCGAGCCCCCTGGTGAAGACATGCAGCGC CAGTGGGCCGGGCTGGTGGAGAAGGTGCAGGCTGCCGTGGGCACCAGCGCCGCCCTG TGCCAGCGACAATCACCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 38	221 aa	MW at 25370.3kD
NOV6e, 278480754 Protein Sequence	TGSKEVQAVETEPEPELRQQTWQSGQRWELALGRFWDYLRWVQTLSEQVQEELLSSQ VTQELRALMDETMKELKAYKSELEEQLTPVAEETRRLSKELGPLVEQGRVRAATVGFL LAGQPLQERAQAWGERLRARMEEMGSRTRDRLDEVKEQVAEVRAKLEEQAQQIRLQAE AFQARLKSWEPLVEDMQRWAGLVEKVQAAVGTSAAPVPSDNHLEG		
	SEQ ID NO: 39	616 bp	
NOV6f, 278480861 DNA Sequence	CACC GGATCCAAGGTCTGTGGGCTGCGTTGCTGGTCACATTCCTGGCAGGATGCCAG GCCAAGGTGGAGCAAGCGGTGGAGACAGAGCCGAGCCGAGCTGCGCCAGCAGACCG AGTGGCAGAGCGGCCAGCGCTGGGAACCTGGCACTGGGTGCGCTTTTGGGATTACCTGCG CTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCAGCTCCCAGGTCACC CAGGAAGTGGGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCTTACAAATCGG AACTGGAGGAACAACCTGACCCCGGTGGCGGAGGAGACGCGGGCACGGCTGTCCAAGGA GCTGGGGCCCCCTGGTGGAAACAGGGCCGCGTGCGGGCCGCCACTGTGGGCTTCCTGGCC GGCCAGCCGCTACAGGAGCGGGCCAGGCTGGGGCGAGCGGCTGCGCGCGCGGATGG AGGAGATGGGCAGCCGACCCGCGACCCGCTGGACGAGGTGAAGGAGCAGGTGGCGGA GGTGGCGGCCAAGCTGGAGGAGCAGGCCAGCAGATACGCTGCAGGCCGAGGCCCTTC CAGGCCCGCTCAAGAGCTGGTTCGAGCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 40	205 aa	MW at 23672.6kD
NOV6f, 278480861 Protein Sequence	TGSKVLWALLVTFLLAGCQAKVEQAVETEPEPELRQQTWQSGQRWELALGRFWDYLR WVQTLSEQVQEELLSSQVTQELRALMDETMKELKAYKSELEEQLTPVAEETRRLSKE LGPLVEQGRVRAATVGFLAQPLQERAQAWGERLRARMEEMGSRTRDRLDEVKEQVAE VRAKLEEQAQQIRLQAEAFQARLKSWELEG		
	SEQ ID NO: 41	709 bp	
NOV6g, 278876278 DNA Sequence	CACC GGATCCACCATGAAGGTCTGTGGGCTGCGTTGCTGGTCACATTCCTGGCAGGA TGCCAGGCCAAGGTGGAGCAAGCGGTGGAGACAGAGCCGAGCCGAGCTGCGCCAGC AGACCGAGTGGCAGAGCGGCCAGCGCTGGGAACCTGGCACTGGGTGCGCTTTTGGGATTA CCTGCGCTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCAGCTCCAG GTCACCCAGGAAGTGGAGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCTTACA AATCGGAAGTGGAGGAACAACCTGACCCCGGTGGCGGAGGAGACGCGGGCACGGCTGTC CAAGGAGCTGCAGGCGGCGCAGGCCCGGCTGGGCGCGGACATGGAGGACGTGCGCGGC		

	CGCCTGGTGCAGTACCGCGCGAGGTGCAGGCCATGCTCGGCCAGAGCACCGAGGAGC TGCGGGTGC GCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATGC CGATGACCTGGAGGAGCAGGCCAGCAGATACGCCTGCAGGCCGAGGCCCTCCAGGCC CGCCTCAAGAGCTGGTTCGAGCCCCCTGGTGAAGACATGCAGCGCCAGTGGGCCGGC TGGTGGAGAAGGTGCAGGCTGCCGTGGGCACCAGCGCCGCCCTGTGCCAGCGACAA TCACCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 42	236 aa	MW at 27045.4kD
NOV6g, 278876278 Protein Sequence	TGSTMKVLW AALLVTFLAGCQAKVEQAVETEPEPELRQQTEWQSGQRWELALGRFWDY LRWVQTLSEQVQEELLSSQVTQELRALMDETMKELKAYKSELEEQLTPVAEETRARLS KELQAAQARLGADMEDVRGRLVQYRGEVQAMLGQSTEELRVRLASHLRKLRKRLLRDA DDLEEQAQQIRLQAEAFQARLKSWFEPLVEDMQRQWAGLVEKVQAAVGTSAAPVPSDN HLEG		
	SEQ ID NO: 43	652 bp	
NOV6h, 278876445 DNA Sequence	CACCGGATCCAAGGTGGAGCAAGCGGTGGAGACAGAGCCGGAGCCGAGCTGCGCCAG CAGACCGAGTGGCAGAGCGGCCAGCGCTGGGAACCTGGCACTGGGTGCGCTTTTGGGATT ACCTGCGCTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCAGCTCCCA GGTCACCCAGGAACCTGAGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGCCCTAC AAATCGGAACCTGGAGGAACAACCTGACCCCGGTGGCGGAGGAGACGCGGGCAGCGCTGT CCAAGGAGCTGCAGCGCGCGCAGGCCCGGCTGGGCGCGGACATGGAGGACGTGCGCGG CCGCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCGGCCAGAGCACCGAGGAG CTGCGGGTGCCTCGCCTCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATG CCGATGACCTGGAGGAGCAGGCCAGCAGATACGCTGCAGGCCGAGGCCCTTCCAGGC CCGCCTCAAGAGCTGGTTCGAGCCCCCTGGTGAAGACATGCAGCGCCAGTGGGCCGGG CTGGTGGAGAAGGTGCAGGCTGCCGTGGGCACCAGCGCCGCCCTGTGCCAGCGACA ATCACCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 44	217 aa	MW at 25026.9kD
NOV6h, 278876445 Protein Sequence	TGSKVEQAVETEPEPELRQQTEWQSGQRWELALGRFWDYLRWVQTLSEQVQEELLSSQ VTQELRALMDETMKELKAYKSELEEQLTPVAEETRARLSKELQAAQARLGADMEDVRG RLVQYRGEVQAMLGQSTEELRVRLASHLRKLRKRLLRDADLEEQAQQIRLQAEAFQA RLKSWFEPLVEDMQRQWAGLVEKVQAAVGTSAAPVPSDNHLEG		
	SEQ ID NO: 45	691 bp	
NOV6i, 278876664 DNA Sequence	CACCGGATCCAAGGTTCTGTGGGCTGCGTTGCTGGTCACATTCTGGCAGGATGCCAG GCCAAGGTGGAGCAAGCGGTGGAGACAGAGCCGGAGCCGAGCTGCGCCAGCAGACCG AGTGGCAGAGCGGCCAGCGCTGGGAACCTGGCACTGGGTGCGCTTTTGGGATTACCTGCG CTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCAGCTCCCAGGTCACC CAGGAACCTGAGGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCCCTACAAATCGG AACTGGAGGAACAACCTGACCCCGGTGGCGGAGGAGACGCGGGCAGCGCTGTCCAAGGA GCTGCAGCGCGCGCAGGCCCGGCTGGGCGCGGACATGGAGGACGTGCGCGGCCGCTG GTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCGGCCAGAGCACCGAGGAGCTGCGGG TGCGCCTCGCCTCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATGCCGATGA CCTGGAGGAGCAGGCCAGCAGATACGCTGCAGGCCGAGGCCCTTCCAGGCCCGCCTC AAGAGCTGGTTCGAGCCCCCTGGTGAAGACATGCAGCGCCAGTGGGCCGGGCTGGTGG AGAAGGTGCAGGCTGCCGTGGGCACCAGCGCCGCCCTGTGCCCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 46	230 aa	MW at 26359.7kD
NOV6i, 278876664 Protein Sequence	TGSKVLW AALLVTFLAGCQAKVEQAVETEPEPELRQQTEWQSGQRWELALGRFWDYLR WVQTLSEQVQEELLSSQVTQELRALMDETMKELKAYKSELEEQLTPVAEETRARLSKE LQAAQARLGADMEDVRGRLVQYRGEVQAMLGQSTEELRVRLASHLRKLRKRLLRDADD LEEQAQQIRLQAEAFQARLKSWFEPLVEDMQRQWAGLVEKVQAAVGTSAAPVPLEG		

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

Table 6B. Comparison of NOV6a against NOV6b through NOV6i.		
Protein Sequence	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV6b	1..194 1..233	189/233 (81%) 189/233 (81%)
NOV6c	1..194 1..229	185/229 (80%) 191/229 (82%)
NOV6d	1..194 4..236	189/233 (81%) 189/233 (81%)
NOV6e	18..194 3..218	171/216 (79%) 172/216 (79%)
NOV6f	2..161 4..202	145/199 (72%) 150/199 (74%)
NOV6g	1..194 5..233	185/229 (80%) 191/229 (82%)
NOV6h	18..194 3..214	167/212 (78%) 174/212 (81%)
NOV6i	2..190 4..227	180/224 (80%) 186/224 (82%)

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

Table 6C. Protein Sequence Properties NOV6a	
PSort analysis:	0.3700 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 19 and 20

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

Table 6D. Geneseq Results for NOV6a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE13293	Human apolipoprotein E (apoE) isoprotein, apoE4 - Homo sapiens, 317 aa. [WO200177136-A1, 18-OCT-2001]	1..172 1..178	136/178 (76%) 154/178 (86%)	2e-69
AAB80998	Human ApoE4 - Homo sapiens, 317 aa. [JP2001017028-A, 23-JAN-2001]	1..172 1..178	136/178 (76%) 154/178 (86%)	2e-69
AAE13296	Human apolipoprotein E (apoE) allele, apoE1 - Homo sapiens, 317 aa. [WO200177136-A1, 18-OCT-2001]	1..169 1..175	135/175 (77%) 152/175 (86%)	7e-69
AAE15158	Human apolipoprotein E (APOE) protein - Homo sapiens, 317 aa. [WO200179234-A2, 25-OCT-2001]	1..172 1..178	136/178 (76%) 153/178 (85%)	9e-69
AAE13298	Human apolipoprotein E (apoE) allele, apoE2** - Homo sapiens, 317 aa. [WO200177136-A1, 18-OCT-2001]	1..172 1..178	136/178 (76%) 153/178 (85%)	9e-69

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

Table 6E. Public BLASTP Results for NOV6a				
Protein Accession Number	Protein/Organism/Length	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P02649	Apolipoprotein E precursor (Apo-E) - Homo sapiens (Human), 317 aa.	1..172 1..178	136/178 (76%) 153/178 (85%)	2e-68
CAA03641	SEQUENCE 28 FROM PATENT WO9712992 - unidentified, 220 aa (fragment).	1..172 21..198	135/178 (75%) 152/178 (84%)	1e-67
Q9GLM8	Apolipoprotein E - Gorilla gorilla (gorilla), 317 aa.	1..172 1..178	134/178 (75%) 151/178 (84%)	2e-67
Q9GLM7	Apolipoprotein E - Pongo pygmaeus (Orangutan), 317 aa.	1..172 1..178	133/178 (74%) 150/178 (83%)	6e-67
Q9GJU3	Apolipoprotein E (Apolipoprotein-E) - Pan troglodytes (Chimpanzee), 317 aa.	1..172 1..178	131/178 (73%) 148/178 (82%)	8e-66

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

Table 6F. Domain Analysis of NOV6a			
Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Apolipoprotein	2..193	83/283 (29%) 170/283 (60%)	4.4e-27

5

EXAMPLE 7.

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

Table 7A. NOV7 Sequence Analysis		
	SEQ ID NO: 47	747 bp
NOV7a, CG129136-01 DNA Sequence	<u>CATGGAATACGCCTCTGACGCTTCACTGGACCCGAAGCCCCGTGGCCTCCCGCGCCC</u> CGCGCTCGCGCCTGCCGCGTACTGCCTTGGGCCCTGGTCGCGGGGCTGCTGCTGCTGC TGCTGCTCACTGCGCGCTGCGCGTCTTCCTGCGCTGCCCTGGGCCGTGCCGGGGC TCGCGCCTCGCCCGGCTCCGCGCCAGCCCGAGACTCCGCGAGGGTCCCGAGCTTTTCG CCGACGATCCCGCGGCCCTCTGGACCTGCGGCAGGGCATGTTTGCGCAGCTGGTGG CCAAGGCTGGAGTCTACTATGTCTTTTCAACTAGAGCTGCGGCGCGTGGTGGCCGG CGAGGGCTCAGGCTCCGTTTCACTTGGCTGCACCTGCAGCCACTGCGCTCTGCTGCT GGGCGCGCGCCCTGGCTTTGACCGTGGACCTGCCACCCGCTCCTCCGAGGCTCGGA	

	ACTCGGCCCTTCGGTTTCCAGGGCCGCTTGCTGCACCTGAGTGCCGGCCAGCGCCTGGGCGTCCATCTTACACACTGAGGCCAGGCCATGCCTGGCAGCTTACCCAGGGCGCCACAGTCTTGGGACTCTTCCGGGTGACCCCGAAATCCCAGCCGGACTCCCTTACCCGAGTCCGGAATAACGCCAGCCTGGGTGCGGGCCACCTGGACAGAGTCCGAATCCTACTCATCCTTCATGGAGACCCCTGGTGTGCTGGGTCCCTGCTGCTTTCTCTACCTC		
	ORF Start: ATG at 2	ORF Stop: TAA at 647	
	SEQ ID NO: 48	215 aa	MW at 22565.7kD
NOV7a, CG129136-01 Protein Sequence	MEYASDASLDPEAPWPPAPRARACRVLPWALVAGLLLLLLTAACAVFLACPWAVSGARASPGSAASPRLREGPELSPDDPAGLLDLRQGMFAQLVAKAGVYVFFQLELRRVVGEGSGSVSLALHLQPLRSAAGAAALALTVDLPASSEARNSAFGFQGRLLHLSAGQRLGVHLHTEARARHAWQLTQGATVLGLFRVTPEIPAGLPSRSE		
	SEQ ID NO: 49	442 bp	
NOV7b, CG129136-03 DNA Sequence	GTCATGGAATACGCCTCTGACGCTGTCAAGGAATACGCCTCTGACGCTGTATGGAATACGCCTCTGACGCTGTCAAGGAATACGCCTCTGACGCTTCACTGGACCCCGAAGCCCCGTGGCCTCCCGCGCCCCGCGCTCGCGCCTGCCGCTACTGCCCTGGGCCCTGGTCGCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCGGCCAGCGCCTGGGCGTCCATCTTACACTGAGGCCAGGGCACGCCATGCCTGGCAGCTTACCCAGGGCGCCACAGTCTTGGGACTCTTCCGGGTGACCCCGAAATCCCAGCCGGACTCCCTTCACCGAGGTCCGAATAACGCCCAGCCTGGGTGCAGCCACCTGGACAGAGTCCGAATCCTACTCCATCCTTCATGGAGACCCCTGGTGCCGGGTCCCTGCTGCTTTCTCTACCTC		
	ORF Start: ATG at 75	ORF Stop: TAA at 342	
	SEQ ID NO: 50	89 aa	MW at 9594.0kD
NOV7b, CG129136-03 Protein Sequence	MEYASDASLDPEAPWPPAPRARACRVLPWALVAGLLLLLLAAGQRLGVHLHTEARARHAWQLTQGATVLGLFRVTPEIPAGLPSRSE		
	SEQ ID NO: 51	289 bp	
NOV7c, 278876545 DNA Sequence	CACCGGATCCACCATGGAATACGCCTCTGACGCTTCACTGGACCCCGAAGCCCCGTGGCCTCCCGCGCCCCGCGCTCGCGCCTGCCGCTACTGCCCTGGGCCCTGGTCGCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCGGCCAGCGCCTGGGCGTCCATCTTACACTGAGGCCAGGGCACGCCATGCCTGGCAGCTTACCCAGGGCGCCACAGTCTTGGGACTCTTCGGGTGACCCCGAAATCCCAGCCGGACTCCCTTCACCGAGGTCCGAACCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 52	96 aa	MW at 10239.7kD
NOV7c, 278876545 Protein Sequence	TGSTMEYASDASLDPEAPWPPAPRARACRVLPWALVAGLLLLLLAAGQRLGVHLHTEARARHAWQLTQGATVLGLFRVTPEIPAGLPSRSELEG		
	SEQ ID NO: 53	178 bp	
NOV7d, 278876597 DNA Sequence	CACCGGATCCGCCTGCCGCTACTGCCCTGGGCCCTGGTCGCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCCGGCCAGCGCCTGGGCGTCCATCTTACACTGAGGCCAGGGCACGCCATGCCCTGGCAGCTTACCCAGGGCGCCACAGTCTTGGGACTCTTCCGGGTGCTCGAGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 54	59 aa	MW at 6296.4kD
NOV7d, 278876597 Protein Sequence	TGSACRVLPWALVAGLLLLLLAAGQRLGVHLHTEARARHAWQLTQGATVLGLFRVLE G		

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

Table 7B. Comparison of NOV7a against NOV7b through NOV7d.		
Protein Sequence	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV7b	176..215 50..89	40/40 (100%) 40/40 (100%)
NOV7c	176..215 54..93	40/40 (100%) 40/40 (100%)
NOV7d	176..201 31..56	26/26 (100%) 26/26 (100%)

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

Table 7C. Protein Sequence Properties NOV7a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane); 0.1732 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 58 and 59

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

Table 7D. Geneseq Results for NOV7a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB75953	Human cytokine 4-1BB ligand - Homo sapiens, 254 aa. [US6355779-B1, 12-MAR-2002]	1..215 1..254	214/254 (84%) 214/254 (84%)	e-114
AAW26657	Human 4-1BB ligand - Homo sapiens, 254 aa. [US5674704-A, 07-OCT-1997]	1..215 1..254	214/254 (84%) 214/254 (84%)	e-114
AAR64190	Human 4-1BB-L polypeptide - Homo sapiens, 254 aa. [WO9426290-A, 24-NOV-1994]	1..215 1..254	214/254 (84%) 214/254 (84%)	e-114
ABB75952	Murine cytokine 4-1BB ligand - Mus sp, 309 aa. [US6355779-B1, 12-MAR-2002]	10..206 65..307	70/248 (28%) 93/248 (37%)	2e-06
AAW26656	Murine 4-1BB ligand - Mus musculus, 309 aa. [US5674704-A, 07-OCT-1997]	10..206 65..307	70/248 (28%) 93/248 (37%)	2e-06

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

Table 7E. Public BLASTP Results for NOV7a				
Protein Accession Number	Protein/Organism/Length	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P41273	Tumor necrosis factor ligand superfamily member 9 (4-1BB ligand) (4- 1BBL) - Homo sapiens (Human), 254 aa.	1..215 1..254	214/254 (84%) 214/254 (84%)	e-113
P41274	Tumor necrosis factor ligand superfamily member 9 (4-1BB ligand) (4- 1BBL) - Mus musculus (Mouse), 309 aa.	10..206 65..307	70/248 (28%) 93/248 (37%)	5e-06
Q93QG0	Putative regulator - Brevibacterium sp. HCU, 263 aa.	13..84 53..112	29/72 (40%) 32/72 (44%)	0.064
P42514	Hypothetical protein PA4219 - Pseudomonas aeruginosa, 394 aa.	77..202 257..389	45/140 (32%) 59/140 (42%)	0.94
Q9PEB0	UDP-N-acetylmuramoylalanine-D-glutamate ligase - Xylella fastidiosa, 468 aa.	127..185 387..447	27/62 (43%) 33/62 (52%)	2.8

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

Table 7F. Domain Analysis of NOV7a			
Pfam Domain	NOV7a Match Region	Identities/ Similarities for the Matched Region	Expect Value
TNF	95..201	33/115 (29%) 102/115 (89%)	1.3e-27

5

EXAMPLE 8.

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

Table 8A. NOV8 Sequence Analysis		
	SEQ ID NO: 55	598 bp
NOV8a, CG133483-01 DNA Sequence	GCGGCCGCGAGCCGCTGCCCCGGGCGGGCGCCCGCGGCGGCACCATGAGTCCCCGCTC GTGCCTGCGTTTCGCTGCGCCTCCTCGTCTTCGCCGTCTTCTCAGCCGCCGCGAGCAAC TGGCTGTACCTGGCCAAGCTGTCGTCGGTGGGAGCATCTCAGAGGAGGAGACGTGCG AGAAACTCAAGGGCTGATCCAGAGGCAGGTGCAGATGTGCAAGCGGAACCTGGAAGT CATGGACTCGGTGCGCCGCGGTGCCAGCTGGCCATTGAGGAGTGCCAGTACCAGTTC CGGAACCGGCGCTGGAAGTGTCCACACTCGACTCCTTGCCCGTCTTCGGCAAGGTGG TGACGCAAGGTGGCAAAGCCTGGCATGGAGGAGACAGTGGATGTCATCTACTGACAGA	

	AAGCAGGTCAGGGGAGAGGACTCAAAGGCGCAAGGAAGGAGGTAGCCTGGGGCAATGC TGGCCTGAAGCCATCGTGGGTACTCAGGACCCCATGAGAAGCCCCCTTACCTTTGGG GGAGGAACCAAGGGGGCACCCACACCGGCTCAGCATCTGAGGGACACTGAGATTCCCTA CTAGGCCCTGGGTGGTACC		
	ORF Start: ATG at 45	ORF Stop: TGA at 570	
	SEQ ID NO: 56	175 aa	MW at 19416.8kD
NOV8a, CG133483-01 Protein Sequence	MSPRSLRSLRLLVFAVFSAAASNWLYLAKLSSVGSISEEETCEKLGKLIQRQVQMCK RNLEVMDSVRRGAQLAIEECQYQFRNRRWNCSTLDSLVPFGKVVTQGGKAWHGGDSGC HLLTESRSGERTQRRKEGGS LGQCWPEAIVGTQDPMRSPPYLWGRNQGGTHTGSASEG H		

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

Table 8B. Protein Sequence Properties NOV8a	
PSort analysis:	0.4753 probability located in outside; 0.1425 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 23 and 24

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

Table 8C. Geneseq Results for NOV8a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB90437	Human polypeptide SEQ ID NO 2813 - Homo sapiens, 185 aa. [WO200190304-A2, 29-NOV-2001]	1..105 1..105	105/105 (100%) 105/105 (100%)	8e-55
AAE09707	Human gene 3 encoding novel protein HE8QR01, SEQ ID NO:54 - Homo sapiens, 365 aa. [WO200155202-A1, 02-AUG-2001]	1..105 15..119	105/105 (100%) 105/105 (100%)	8e-55
AAM40675	Human polypeptide SEQ ID NO 5606 - Homo sapiens, 365 aa. [WO200153312-A1, 26-JUL-2001]	1..105 15..119	105/105 (100%) 105/105 (100%)	8e-55
AAM38889	Human polypeptide SEQ ID NO 2034 - Homo sapiens, 351 aa. [WO200153312-A1, 26-JUL-2001]	1..105 1..105	105/105 (100%) 105/105 (100%)	8e-55
AAU29063	Human PRO polypeptide sequence #40 - Homo sapiens, 351 aa. [WO200168848-A2, 20-SEP-2001]	1..105 1..105	105/105 (100%) 105/105 (100%)	8e-55

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

Table 8D. Public BLASTP Results for NOV8a				
Protein Accession Number	Protein/Organism/Length	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P56705	WNT-4 protein precursor - Homo sapiens (Human), 351 aa.	1..105 1..105	105/105 (100%) 105/105 (100%)	2e-54
P22724	WNT-4 protein precursor - Mus musculus (Mouse), 351 aa.	1..105 1..105	105/105 (100%) 105/105 (100%)	2e-54
Q9QXQ5	WNT-4 protein precursor - Rattus norvegicus (Rat), 351 aa.	1..105 1..105	104/105 (99%) 104/105 (99%)	7e-54
JC2451	Cwnt-4 protein precursor - chicken, 351 aa.	1..105 1..105	95/105 (90%) 98/105 (92%)	2e-48
P49337	WNT-4 protein precursor - Gallus gallus (Chicken), 351 aa.	1..105 1..105	95/105 (90%) 98/105 (92%)	2e-48

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

Table 8E. Domain Analysis of NOV8a			
Pfam Domain	NOV8a Match Region	Identities/ Similarities for the Matched Region	Expect Value
wnt	42..105	43/67 (64%) 64/67 (96%)	5.1e-47

5

EXAMPLE 9.

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

Table 9A. NOV9 Sequence Analysis		
	SEQ ID NO: 57	1047 bp
NOV9a, CG135316-01 DNA Sequence	ATGGCCGCGCCATCGCTAGCGGCTTGATCCGCCAGAAGCGGCAGGCGCGGGAGCAGC ACTGGGACCGCCGCTCTGCCAGCAGGAGGCGGAGCAGCCCCAGCAGAACC GCGGGCT CTGCAACGGCAACCTGGTGGATATCTTCTCCAAAGTGC GCATCTTCGGCCTCAAGAAG CGCAGGTTGCGGCGCCAAGATCCCCAGCTCAAGGGTATAGTGACCAGGTTATATTGCA GGCAAGGCTACTACTTGCAAATGCACCCCGATGGAGCTCTCGATGGAACCAAGGATGA CAGCACTAATTCTACTCTTCAACCTCATACCAAGTGGGACTACGTGTTGTTGCCATC CAGGGAGTGAAAACAGGGTTGTATATAGCCATGAATGGAGAAGGTACCTCTACCCAT CAGAACTTTTTACCCCTGAATGCAAGTTTAAAGAATCTGTTTTTGA AATTATATGT AATCTACTCATCCATGTTGTACAGACAACAGGAATCTGGTAGAGCTGGTTTTTGGGA TTAAATAAGGAAGGGCAAGCTATGAAAGGGAACAGAGTAAAGAAAACCAACCAGCAG CTCATTTTCTACCCAAGCCATTGGAAGTTGCCATGTACCGAGAACCATCTTTGCATGA TGTGTTGGGAAACGGTCCCGAAGCTGGGGTGACGCCAAGTAAAGCACAAGTGCGTCT AAATCCATTTTCAGATATACTCCGTCCTGTTTTTAAATGAACCAAACCTTAACGCCATCCC CGTTTCTGGCTGCGTTCCTCCCTCATACTCAGCAGAGCATGGGCAAGACGGCTGTTGTGT	

	TCTTTCGTGGTCCGTAAAGTTTAACTTCTGATCCTTAATAGGAGGATAAGCGCCGTG ATAGAGAAATCCAAAGGTCATTTGTATTACGATGGCTAGATAATGTAATGAATTCCAA TGCTGTGCATCAGCGAATACGTCATCAAAATTGCTACAAAACAATAATAATAGGTTG TTCACAGCTTAAATGTTTAGGTAGTGAAGAGGAAAGAATATAACCTACATTATTAT TGA		
	ORF Start: ATG at 1	ORF Stop: TAG at 907	
	SEQ ID NO: 58	302 aa	MW at 34001.7kD
NOV9a, CG135316-01 Protein Sequence	MAAAIASGLIRQKRQAREQHWDPSASRRRSSPSKNRGLCNGNLVDIFSKVRIIFGLKK RRLRRQDPQLKGIVTRLYCRQGYLQMHPDGDGTDKDDSTNSTLFNLIIPVGLRVVAI QGVKTGLYIAMNREGYLYPSELTPECKFKESVFENYYVIYSSMLYRQQESGRAWFLG LNKEGQAMKGNRVKTKPAAHFLPKPLEVAMYREPSLHDVGETVPKPGVTPSKSTSAS KSISDILRPVFNEPNLTSPFLAAFPSSAHEGQDCCVLSWSVKFNFLILNRRISAV IEKSKGHLYYDG		

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

Table 9B. Protein Sequence Properties NOV9a	
PSort analysis:	0.4539 probability located in mitochondrial matrix space; 0.4027 probability located in mitochondrial intermembrane space; 0.3000 probability located in nucleus; 0.1467 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 9C. Geneseq Results for NOV9a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE18819	Human FGF-14 protein - Homo sapiens, 247 aa. [US2002001825-A1, 03-JAN-2002]	1..232 1..232	232/232 (100%) 232/232 (100%)	e-134
AAB65291	Human PRO185 protein sequence SEQ ID NO:499 - Homo sapiens, 247 aa. [WO200073454-A1, 07-DEC-2000]	1..232 1..232	232/232 (100%) 232/232 (100%)	e-134
AAB31182	Amino acid sequence of human polypeptide PRO185 - Homo sapiens, 247 aa. [WO200077037-A2, 21-DEC-2000]	1..232 1..232	232/232 (100%) 232/232 (100%)	e-134
AAB47288	PRO185 polypeptide - Homo sapiens, 247 aa. [WO200140464-A1, 07-JUN-2001]	1..232 1..232	232/232 (100%) 232/232 (100%)	e-134
AAE04406	Human fibroblast growth factor homologous factor 4 (FHF-4) - Homo sapiens, 247 aa. [WO200147957-A2, 05-JUL-2001]	1..232 1..232	232/232 (100%) 232/232 (100%)	e-134

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

Table 9D. Public BLASTP Results for NOV9a				
Protein Accession Number	Protein/Organism/Length	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q92915	Fibroblast growth factor-14 (FGF-14) (Fibroblast growth factor homologous factor 4) (FHF-4) - Homo sapiens (Human), 247 aa.	1..232 1..232	232/232 (100%) 232/232 (100%)	e-133
Q8R5L7	Fibroblast growth factor14 - Rattus norvegicus (Rat), 247 aa.	1..232 1..232	230/232 (99%) 230/232 (99%)	e-132
P70379	Fibroblast growth factor-14 (FGF-14) (Fibroblast growth factor homologous factor 4) (FHF-4) - Mus musculus (Mouse), 247 aa.	1..232 1..232	229/232 (98%) 229/232 (98%)	e-130
Q9IAI6	Fibroblast growth factor homologous factor 4 isoform 1A - Gallus gallus (Chicken), 237 aa (fragment).	12..242 1..231	213/231 (92%) 221/231 (95%)	e-123
Q8R4X0	Fibroblast growth factor-like factor-4D - Rattus norvegicus (Rat), 211 aa (fragment).	58..232 22..196	167/175 (95%) 168/175 (95%)	3e-94

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

Table 9E. Domain Analysis of NOV9a			
Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
FGF	71..200	54/147 (37%) 100/147 (68%)	2.7e-43

5 **EXAMPLE 10.**

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

Table 10A. NOV10 Sequence Analysis	
	SEQ ID NO: 59 30 bp
NOV10a, CG53473-03 DNA Sequence	GGCAAGCTCTGGGCCATCGGTCACTTCATG

	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 60	10 aa	MW at 1159.4kD
NOV10a, CG53473-03 Protein Sequence	GKLWAI GHFM		

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

Table 10B. Protein Sequence Properties NOV10a	
PSort analysis:	
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 10C. Geneseq Results for NOV10a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

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

Table 10D. Public BLASTP Results for NOV10a				
Protein Accession Number	Protein/Organism/Length	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

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

Table 10E. Domain Analysis of NOV10a			
Pfam Domain	NOV10a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Bombesin	1..10	7/10 (70%) 10/10 (100%)	0.41

10 **EXAMPLE 11.**

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

Table 11A. NOV11 Sequence Analysis			
	SEQ ID NO: 61	126 bp	
NOV11a, CG54725-03 DNA Sequence	GCACACAACTAGACCTGGAAGAAATTGCCAGCTTGGATAAGGCCAAGCTGAAGGCCA CAGAGATGCAGAAGAACAACCTCTGATGACCAAAGAGACCACAGAGCAGGAGAAGTGGAG TGAAATTTC		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 62	42 aa	MW at 4848.5kD
NOV11a, CG54725-03 Protein Sequence	AHKLDLEEIASLDKAKLKATEMQKNLTMTKETTEQEKWSEIS		

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

Table 11B. Protein Sequence Properties NOV11a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 11C. Geneseq Results for NOV11a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB05036	Human NOV2c protein SEQ ID NO:8 - Homo sapiens, 43 aa. [WO200190155-A2, 29-NOV-2001]	1..42 2..43	42/42 (100%) 42/42 (100%)	5e-17
ABB05035	Human NOV2b protein SEQ ID NO:6 - Homo sapiens, 43 aa. [WO200190155-A2, 29-NOV-2001]	1..42 2..43	42/42 (100%) 42/42 (100%)	5e-17
ABB05034	Human NOV2a protein SEQ ID NO:4 - Homo sapiens, 43 aa. [WO200190155-A2, 29-NOV-2001]	1..42 2..43	42/42 (100%) 42/42 (100%)	5e-17
AAY80267	Thymosin beta 4 peptide isoform Tbeta10 - Unidentified, 43 aa. [WO200006190-A1, 10-FEB-2000]	1..42 1..43	32/43 (74%) 33/43 (76%)	9e-08
AAR96932	Thymosin beta 10 - Synthetic, 43 aa. [WO9611016-A1, 18-APR-1996]	1..42 1..43	32/43 (74%) 33/43 (76%)	9e-08

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

Table 11D. Public BLASTP Results for NOV11a				
Protein Accession Number	Protein/Organism/Length	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAM34215	Thymosin beta 10 - Equus caballus (Horse), 44 aa.	1..42 2..44	32/43 (74%) 33/43 (76%)	2e-07
P13472	Thymosin beta-10 - Homo sapiens (Human), 43 aa.	1..42 1..43	32/43 (74%) 33/43 (76%)	2e-07
CAC38717	Sequence 1 from Patent WO0129217 - Homo sapiens (Human), 58 aa.	1..37 2..38	26/37 (70%) 30/37 (80%)	1e-06
P21752	Thymosin beta-9 (Thymosin beta-10) [Contains: Thymosin beta-8] - Bos taurus (Bovine), 41 aa.	1..37 1..38	28/38 (73%) 28/38 (73%)	2e-05
P21753	Thymosin beta-9 - Sus scrofa (Pig), 41 aa.	1..37 1..38	27/38 (71%) 28/38 (73%)	4e-05

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

Table 11E. Domain Analysis of NOV11a			
Pfam Domain	NOV11a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Thymosin	1..40	29/41 (71%) 32/41 (78%)	1.8e-11

5

EXAMPLE 12.

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

Table 12A. NOV12 Sequence Analysis			
	SEQ ID NO: 63	108 bp	
NOV12a, CG56983-02 DNA Sequence	TGCCCCATCAAACCCGAGGCTCCTGGCGAAGACGAGTCCCTGGAGGAGCTGAGCCACT ATTATGCTTCCCTGTGCCACTACCTCAACGTGGTCACCAGACAGTTAATT		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 64	36 aa	MW at 4118.6kD
NOV12a,	CPIKPEAPGEDESLEELSHYYASLCHYLNVTTRQLI		

CG56983-02 Protein Sequence			
	SEQ ID NO: 65	45 bp	
NOV12b, CG56983-03 DNA Sequence	GCTTCCCTGTGCCACTACCTCAACGTGGTCACCAGACAGTTAATT		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 66	15 aa	MW at 1730.0kD
NOV12b, CG56983-03 Protein Sequence	ASLCHYLNVVTRQLI		
	SEQ ID NO: 67	69 bp	
NOV12c, CG56983-04 DNA Sequence	CTGGAGGAGCTGAGCCACTATTATGCTTCCCTGTGCCACTACCTCAACGTGGTCACCA GACAGTTAATT		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 68	23 aa	MW at 2765.1kD
NOV12c, CG56983-04 Protein Sequence	LEELSHYYASLCHYLNVVTRQLI		

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

Table 12B. Comparison of NOV12a against NOV12b and NOV12c.		
Protein Sequence	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV12b	22..36 1..15	15/15 (100%) 15/15 (100%)
NOV12c	14..36 1..23	23/23 (100%) 23/23 (100%)

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

Table 12C. Protein Sequence Properties NOV12a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 12D. Geneseq Results for NOV12a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE09439	Human sbghPYYa protein - Homo sapiens, 70 aa. [WO200160850-A1, 23-AUG-2001]	1..34 29..62	34/34 (100%) 34/34 (100%)	1e-14
AAU85987	Modified human peptide YY - Homo sapiens, 36 aa. [WO200210195-A2, 07-FEB-2002]	2..34 2..34	27/33 (81%) 29/33 (87%)	6e-09
AAU06188	Human peptide tyrosine-tyrosine (PYY) - Homo sapiens, 36 aa. [WO200158409-A2, 16-AUG-2001]	2..34 2..34	27/33 (81%) 29/33 (87%)	6e-09
AAB91223	Peptide YY SEQ ID NO:397 - Homo sapiens, 36 aa. [WO200069900-A2, 23-NOV-2000]	2..34 2..34	27/33 (81%) 29/33 (87%)	6e-09
AAG75364	Human colon cancer antigen protein SEQ ID NO:6128 - Homo sapiens, 176 aa. [WO200122920-A2, 05-APR-2001]	2..34 54..86	27/33 (81%) 29/33 (87%)	6e-09

In a BLAST search of public sequence databases, the NOV12a protein was found to have homology to the proteins shown in the BLASTP data in Table 12E.

Table 12E. Public BLASTP Results for NOV12a				
Protein Accession Number	Protein/Organism/Length	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
S34568	peptide YY precursor (clone L1) - human, 90 aa (fragment).	2..34 30..62	27/33 (81%) 29/33 (87%)	1e-08
S34569	peptide YY precursor (clone L2) - human, 90 aa (fragment).	2..34 30..62	27/33 (81%) 29/33 (87%)	1e-08
A31358	peptide YY - human, 36 aa.	2..34 2..34	27/33 (81%) 29/33 (87%)	1e-08
P10082	Peptide YY precursor (PYY) (Peptide tyrosine tyrosine) - Homo sapiens (Human), 97 aa.	2..34 30..62	27/33 (81%) 29/33 (87%)	1e-08
Q91XD0	Unknown (protein for MGC:19143) - Mus musculus (Mouse), 98 aa.	2..34 30..62	27/33 (81%) 28/33 (84%)	2e-08

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

Table 12F. Domain Analysis of NOV12a			
Pfam Domain	NOV12a Match Region	Identities/ Similarities for the Matched Region	Expect Value
hormone3	1..34	17/37 (46%) 25/37 (68%)	3.1e-11

EXAMPLE 13.

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

Table 13A. NOV13 Sequence Analysis			
	SEQ ID NO: 69	72 bp	
NOV13a, CG57666-02 DNA Sequence	CGGAGCACACCGAACATCAGGCCCGCGCACAGACTGACAAGAGTGAACCTGCCCATGC CGCGCCGCTACTAC		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 70	24 aa	MW at 2965.4kD
NOV13a, CG57666-02 Protein Sequence	RSTPNIRPAHRLTRVNLMPRRYY		

Further analysis of the NOV13a protein yielded the following properties shown in

5 Table 13B.

Table 13B. Protein Sequence Properties NOV13a	
PSort analysis:	0.8500 probability located in lysosome (lumen); 0.7500 probability located in mitochondrial intermembrane space; 0.6850 probability located in nucleus; 0.3600 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 13C. Geneseq Results for NOV13a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG07914	Novel human diagnostic protein #7905 - Homo sapiens, 286 aa. [WO200175067-A2, 11-OCT-2001]	1..24 80..103	23/24 (95%) 23/24 (95%)	7e-07
ABG07914	Novel human diagnostic protein #7905 - Homo sapiens, 286 aa. [WO200175067-A2, 11-OCT-2001]	1..24 80..103	23/24 (95%) 23/24 (95%)	7e-07

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

Table 13D. Public BLASTP Results for NOV13a				
Protein Accession Number	Protein/Organism/Length	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

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

Table 13E. Domain Analysis of NOV13a			
Pfam Domain	NOV13a Match Region	Identities/ Similarities for the Matched Region	Expect Value

EXAMPLE 14.

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

Table 14A. NOV14 Sequence Analysis			
	SEQ ID NO: 71	934 bp	
NOV14a, CG59651-03 DNA Sequence	CATGCCCCGGGCAAGA ACTCAGGACGCTGAATGGCTCTCAGATGCTCCTGGTGTGTGCTG GTGCTCTCTCGTGGCTGCCGCATGGGGGCGCCCTGTCTCTGGCCGAGGCGAGCCGCGCAA GTTTCCCGGGGACCCCTCAGAGTTGCACACCGAAGACTCCAGATTCCGAGAGTTGCCGAA ACGCTACGAGGCTGCGTAAACCAGGCTCGGGCCAACTCAGAGCTGGGAAGATTGCAAC ACCGACCTCGTCCCGGCCCTGCAGTCCGGATACTCACGCCAGAAGTGCGGCCTGGGAT CCGGCGGCCACCTGCACCTGCGTATCTCTCGGGCGCCCTTCCCGAGGGGCTCCCCGA GGCCTCCCGCCTTACC GGCTCTGTTCGGCTGTCCCCGACGGCGTCAAGGTCGTGG GACGTGACACGACCTCTGCGGCGTCAGCTCAGCCTTGCAAGACCCAGGCGCCCGCGC TGCACCTGCGACTGTGCGCGCCGCCGTGCGAGTCGGACCAACTGCTGGCAGAACTTTC GTCCGCACGGCCCCAGCTGGAGTTGCACCTGCGGCCGCAAGCCGCCAGGGGGCGCCGC AGAGCGCGTGGCGCAACGGGGACCACTGTCCGCTCGGGCCCGGGCGTTGCTGCCGCT TGACACACGGTCCGCGCGTCGCTGGAAGACCTGGGCTGGGCGGATTTGGGTGTTTGC ACGGGAGGTGCAAGTGACCATGTGCATCGGCGCGTGCCCGAGCCAGTTCCGGGCGGCA AACATGCACGCGCAGATCAAGACGAGCCTGCACCGCCTGAAGCCCGACACGGTGCCAG CGTCTGCTGCGTGCCCGCCAGCTACAATCCCATGGTGCTCATTCAAAAGACCGACAC CGGGGTGTCGCTCCAGACCTATGATGACTTGTTAGCCAAAGACTGCCACTGCATATGA <u>ACTAGT</u>		
	ORF Start: ATG at 2	ORF Stop: TGA at 926	
	SEQ ID NO: 72	308 aa	MW at 34157.7kD
NOV14a, CG59651-03 Protein Sequence	MPGQELRTLNGSQMLLVLLVLSWLP HGGALSLAEASRASFPGPSELHTEDSRFRELRK RYEDLLTRLRANQSWEDSNTDLVPAPAVRILTPPEVRLGSGGHLHLRISRAALPEGLPE ASRLHRLFRLSPTASRSWDVTRPLRRQLSLARPQAPALHLRLSPPPSQSDQLLAESS SARPQLELHLRPQAARGRRRARARNGDHCPLGPGRCRLHTVRASLEDLGWADWVLS REVQVTMCIGACPSQFRAANMHAQIKTSLHRLKPDTPASCCVPASYNPMVLIQKTD GVSLQTYDDLAKDCHCI		
	SEQ ID NO: 73	306 bp	
NOV14b, 207775400 DNA Sequence	AGATCTTGCCGCTGCACACGGTCCGCGCGTCGCTGGAAGACCTGGGCTGGGCCGATT GGGTGCTGTGCGCCACGGGAGGTGCAAGTGACCATGTGCATCGGCGCGTGCCCGAGCCA GTTCCGGGCGGCAACATGCACGCGCAGATCAAGACGAGCCTGCACCGCCTGAAGCCC GACACGGTGCCAGCGCCCTGCTGCTGCGTCCCCGACGCTACAATCCCATGGTGCTCATT AAAAGACCGACACCGGGGTGTCGCTCCAGACCTATGATGACTTGTTAGCCAAAGACTG		

	CCACTGCATACTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 74	102 aa	MW at 11338.0kD
NOV14b, 207775400 Protein Sequence	RSCRLHTVRASLEDLGWADWVLSPREVQVTMCIGACPSQFRAANMHAQIKTSLHRLKPD DTVPAPCCVPASYNPMVLIQKTDGTGVS LQTYDDLAKDCHCILE		
	SEQ ID NO: 75	849 bp	
NOV14c, 207775387 DNA Sequence	AGATCTCTGTCTCTGGCCGAGGCGAGCCGCGCAAGTTTCCCGGGACCTCAGAGTTGC ACTCCGAAGACTCCAGATTCCGAGAGTTGCGGAAACGCTACGAGGACCTGCTAACCAG GCTGCGGGCCAACCAGAGCTGGGAAGATTGGAACACCGACCTCGTCCCGGGCCCTGCA GTCCGGATACTACGCCAGAAGTGCGGCTGGGATCCGGCGGCCACCTGCACCTGCGTA TCTCTCGGGCCGCCCTTCCCGAGGGGCTCCCCGAGGCCTCCCGCCTTACCGGGCTCT GTTCCGGCTGTCCCCGACGGCGTCAAGGTCTGTGGGACGTGACACGACCGCTGCGGCGT CAGCTCAGCCTTGCAAGACCCAGGCGCCCGCGTGCACCTGCGACTGTGCGCCGCCG CGTCGCAGTCGGACCAACTGCTGGCAGAATCTTCGTCGCGACGGCCCCAGCTGGAGTT GCACCTGCGGCGCGCAAGCCGCCAGGGGGCGCCGCAGAGCGCGTGCAGCGCAACGGGGAC CACTGTCCGCTCGGGCCCGGGCGTTGCTGCCGTCTGCACACGGTCCGCGCGTCTGCTGG AAGACCTGGGCTGGGCGGATTGGGTGCTGTGCGCACGGGAGGTGCAAGTGACCATGTG CATCGGCGCGTGGCCGAGCCAGTTCCGGGCGGCAACATGCACGCGCAGATCAAGACG AGCCTGCACCGCTGAAGCCCGACACGGTGCCAGCGCCCTGCTGCGTGCCCGCCAGCT ACAATCCCATGGTGCTCATTTCAAAGACCGACACCGGGGTGTCGCTCCAGACCTATGA TGACTTGTTAGCCAAAGACTGCCACTGCATACTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 76	283 aa	MW at 31525.5kD
NOV14c, 207775387 Protein Sequence	RSLSLAEASRASFPGPSELHSEDSRFRELKRYEDLLTRLRANQSWEDSNTDLVPAPA VRILTPEVRLGSGGHLHLRISRAALPEGLPEASRLHRLFRLSPTASRSWDVTRPLRR QLSLARPQAPALHLRLSPPPSQSDQLLAESSSARPQLEHLRPAARGRRRARARNGD HCPLGPGRCRLHTVRASLEDLGWADWVLSPREVQVTMCIGACPSQFRAANMHAQIKT SLHRLKPDTPAPCCVPASYNPMVLIQKTDGTGVS LQTYDDLAKDCHCILE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 14B.

Table 14B. Comparison of NOV14a against NOV14b and NOV14c.		
Protein Sequence	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV14b	209..308 1..100	98/100 (98%) 98/100 (98%)
NOV14c	40..308 13..281	256/269 (95%) 257/269 (95%)

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

Table 14C. Protein Sequence Properties NOV14a	
PSort analysis:	0.5851 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 30 and 31

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

Table 14D. Geneseq Results for NOV14a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB09174	Human NAG-1 protein SEQ ID NO:2 - Homo sapiens, 308 aa. [WO200220759-A2, 14-MAR-2002]	1..308 1..308	307/308 (99%) 307/308 (99%)	e-180
AAE13538	Human macrophage inhibitory cytokine 1 (MIC-1) wild-type protein - Homo sapiens, 308 aa. [WO200181928-A1, 01-NOV-2001]	1..308 1..308	307/308 (99%) 307/308 (99%)	e-180
AAB26122	Human TGF-beta superfamily protein PCIGF - Homo sapiens, 308 aa. [WO200056352-A2, 28-SEP-2000]	1..308 1..308	307/308 (99%) 307/308 (99%)	e-180
AAW10662	Human TGF-beta-like cytokine pCL13 - Homo sapiens, 308 aa. [WO9700958-A1, 09-JAN-1997]	1..308 1..308	307/308 (99%) 307/308 (99%)	e-180
AAR77097	Human TGF-beta superfamily protein - Homo sapiens, 308 aa. [JP07250688-A, 03-OCT-1995]	1..308 1..308	307/308 (99%) 307/308 (99%)	e-180

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

Table 14E. Public BLASTP Results for NOV14a				
Protein Accession Number	Protein/Organism/Length	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q99988	Growth/differentiation factor 15 precursor (GDF-15) (Placental bone morphogenic protein) (Placental TGF-beta) (Macrophage inhibitory cytokine-1) (MIC-1) (Prostate differentiation factor) (NSAID-regulated protein 1) (NRG-1) - Homo sapiens (Human), 308 aa.	1..308 1..308	307/308 (99%) 307/308 (99%)	e-180
Q9BWA0	Prostate differentiation factor (Unknown) (Protein for MGC:4145) - Homo sapiens (Human), 308 aa.	1..308 1..308	305/308 (99%) 307/308 (99%)	e-179
JC5697	placental transforming growth factor-beta homolog - human, 309 aa.	1..308 1..309	305/309 (98%) 305/309 (98%)	e-176
Q9Z0J7	Growth/differentiation factor 15 precursor (GDF-15) - Mus musculus (Mouse), 303 aa.	3..307 13..302	182/305 (59%) 224/305 (72%)	7e-96
Q9Z0J6	Growth/differentiation factor 15 precursor (GDF-15) - Rattus norvegicus (Rat), 303 aa.	14..307 22..302	174/294 (59%) 218/294 (73%)	3e-93

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

Table 14F. Domain Analysis of NOV14a			
Pfam Domain	NOV14a Match Region	Identities/ Similarities for the Matched Region	Expect Value
TGF-beta	208..308	32/112 (29%) 63/112 (56%)	2.9e-15

5

EXAMPLE 15.

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

Table 15A. NOV15 Sequence Analysis		
	SEQ ID NO: 77	27 bp
NOV15a.	TGCTACTTCCAGAACTGCCCGAGGGGC	

CG89614-04 DNA Sequence			
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 78	9 aa	MW at 1087.2kD
NOV15a, CG89614-04 Protein Sequence	CYFQNCPRG		

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

Table 15B. Protein Sequence Properties NOV15a	
PSort analysis:	
SignalP analysis:	No Known Signal Sequence Predicted

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

Table 15C. Geneseq Results for NOV15a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

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

Table 15D. Public BLASTP Results for NOV15a				
Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

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

Table 15E. Domain Analysis of NOV15a			
Pfam Domain	NOV15a Match Region	Identities/ Similarities for the Matched Region	Expect Value
hormone4	1..9	7/9 (78%) 9/9 (100%)	0.34

Example B: Sequencing Methodology and Identification of NOVX Clones

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

in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other
5 sequence variations.

3. PathCalling™ Technology:

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

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

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corporation proprietary library of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain
30 reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen

Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

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

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

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

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

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

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

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

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an

emphasis on metabolic diseases), AI_comprehensive_panel (containing normal tissue and samples from autoimmune diseases), Panel CNSD.01 (containing central nervous system samples from normal and diseased brains) and CNS_neurodegeneration_panel (containing samples from normal and Alzheimer's diseased brains).

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

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

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

- Probes and primers were designed for each assay according to Applied Biosystems
25 Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (T_m) range = 58°-60°C, primer optimal T_m = 59°C, maximum primer difference = 2°C, probe does not have 5'G, probe
30 T_m must be 10°C greater than primer T_m , amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by SyntheGen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the

probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

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

When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using 1X TaqMan® Universal Master mix 20 (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were analyzed and processed as described previously.

Panels 1, 1.1, 1.2, and 1.3D

25 The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, 30 gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the

following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea,

5 breast, ovary, uterus, placenta, prostate, testis and adipose.

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

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

10 s cell var = small cell variant,

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

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

15 astro = astrocytoma, and

neuro = neuroblastoma.

General_screening_panel_v1.4 and General_screening_panel_v1.5

The plates for Panels 1.4 and 1.5 include 2 control wells (genomic DNA control
20 and chemistry control) and 94 wells containing cDNA from various samples. The samples in Panels 1.4 and 1.5 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer,
25 gastric cancer and pancreatic cancer. Cell lines used in Panel 1.4 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on Panels 1.4 and 1.5 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or fetuses. These samples are
30 derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose. Abbreviations
35 are as described for Panels 1, 1.1, 1.2, and 1.3D.

Panels 2D and 2.2

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

Panel 3D

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

Panels 4D, 4R, and 4.1D

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal

tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease
5 and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells,
10 human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately
15 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM
20 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2 μ g/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were
25 cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5 μ g/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two
30 donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately 2x10⁶cells/ml in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol (5.5x10⁻⁵M) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

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

Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 μ g/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco) and plated at 10⁶cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5 μ g/ml anti-CD28 (Pharmingen) and 3 μ g/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco),

mercaptoethanol $5.5 \times 10^{-5} \text{M}$ (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

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

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

30 The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at 5×10^5 cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5×10^5 cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), $100 \mu\text{M}$ non

essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol
5.5x10⁻⁵M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or
cells activated with PMA at 10ng/ml and ionomycin at 1μg/ml for 6 and 14 hours.
Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also
5 obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100μM non
essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol
5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14
hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells
were activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9,
10 5ng/ml IL-13 and 25ng/ml IFN gamma.

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

AI_comprehensive panel_v1.0

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

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

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

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

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

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

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

35 **Panels 5D and 5I**

 The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases.

Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

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

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

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

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

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

the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

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

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

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

Panel CNSD.01

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

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

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

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

10 **Panel CNS_Neurodegeneration_V1.0**

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

Disease diagnoses are taken from patient records. The panel contains six brains from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no evidence of dementia prior to death. The eight normal control brains are divided into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the following regions are represented: hippocampus, temporal cortex (Brodman Area 21), parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus; the parietal cortex shows moderate neuronal death in the late stages of the disease; the occipital cortex is spared in AD and therefore acts as a "control" region within AD patients. Not all brain regions are represented in all cases.

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

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

A. CG124728-02 and CG124728-03: COMPLEMENT-C1Q

- 10 Expression of gene CG124728-02 and CG124728-03 was assessed using the primer-probe sets Ag918, Ag6167 and Ag6492, described in Tables AA, AB and AC. Results of the RTQ-PCR runs are shown in Tables AD, AE, AF, AG and AH. CG124728-02 and CG124728-03 represent full-length physical clones.

Table AA. Probe Name Ag918

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gtgaacgagcagggacattac-3'	21	294	79
Probe	TET-5'-caagttcacctgccaggtgctg-3'-TAMRA	23	329	80
Reverse	5'-atggacggcgaagtagtagac-3'	21	354	81

15 Table AB. Probe Name Ag6167

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ctcgtcctgctgctcct-3'	17	27	82
Probe	TET-5'-acaacaagatccccagcctctgc-3'-TAMRA	23	76	83
Reverse	5'-ccggcagtcctggaag-3'	16	114	84

Table AC. Probe Name Ag6492

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttccaggactgccgg-3'	15	115	85
Probe	TET-5'-ctcggcgctccgcgatc-3'-TAMRA	18	200	86
Reverse	5'-cgcttgccgctgaag-3'	15	221	87

Table AD. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag918, Run 206989724	Tissue Name	Rel. Exp.(%) Ag918, Run 206989724
AD 1 Hippo	23.3	Control (Path) 3 Temporal Ctx	5.0
AD 2 Hippo	28.9	Control (Path) 4 Temporal Ctx	13.0
AD 3 Hippo	10.7	AD 1 Occipital	12.4

		Ctx	
AD 4 Hippo	18.9	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	18.7	AD 3 Occipital Ctx	8.8
AD 6 Hippo	39.0	AD 4 Occipital Ctx	17.4
Control 2 Hippo	17.6	AD 5 Occipital Ctx	4.3
Control 4 Hippo	82.9	AD 6 Occipital Ctx	12.1
Control (Path) 3 Hippo	76.8	Control 1 Occipital Ctx	9.5
AD 1 Temporal Ctx	13.3	Control 2 Occipital Ctx	24.7
AD 2 Temporal Ctx	21.2	Control 3 Occipital Ctx	10.4
AD 3 Temporal Ctx	6.4	Control 4 Occipital Ctx	17.7
AD 4 Temporal Ctx	23.0	Control (Path) 1 Occipital Ctx	36.6
AD 5 Inf Temporal Ctx	24.0	Control (Path) 2 Occipital Ctx	9.1
AD 5 Sup Temporal Ctx	100.0	Control (Path) 3 Occipital Ctx	4.3
AD 6 Inf Temporal Ctx	26.2	Control (Path) 4 Occipital Ctx	8.8
AD 6 Sup Temporal Ctx	20.7	Control 1 Parietal Ctx	11.4
Control 1 Temporal Ctx	11.4	Control 2 Parietal Ctx	23.0
Control 2 Temporal Ctx	12.2	Control 3 Parietal Ctx	3.5
Control 3 Temporal Ctx	8.1	Control (Path) 1 Parietal Ctx	17.4
Control 4 Temporal Ctx	13.2	Control (Path) 2 Parietal Ctx	7.2
Control (Path) 1 Temporal Ctx	11.5	Control (Path) 3 Parietal Ctx	6.5
Control (Path) 2 Temporal Ctx	12.5	Control (Path) 4 Parietal Ctx	16.8

Table AE. General_screening_panel_v1.5

Tissue Name	Rel. Exp.(%) Ag6167, Run 254397160	Tissue Name	Rel. Exp.(%) Ag6167, Run 254397160
-------------	--	-------------	--

Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	5.6	Bladder	0.0
Melanoma* Hs688(B).T	37.4	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	100.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK- OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	4.4	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	19.9
Ovarian ca. OVCAR-8	4.6	Bone Marrow Pool	0.0
Ovary	9.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA- MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0

Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	5.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	4.2
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	34.9
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	10.1
Kidney Pool	0.0	Adrenal Gland	9.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	6.5	Pancreas Pool	0.0

Table AF. General_screening_panel_v1.6

Tissue Name	Rel. Exp.(%) Ag6492, Run 277251131	Tissue Name	Rel. Exp.(%) Ag6492, Run 277251131
Adipose	14.8	Renal ca. TK-10	5.3
Melanoma* Hs688(A).T	69.7	Bladder	29.7
Melanoma* Hs688(B).T	27.7	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0

Testis Pool	1.3	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	1.7	Colon ca. CaCo-2	0.0
Placenta	7.2	Colon cancer tissue	100.0
Uterus Pool	9.3	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	25.2
Ovarian ca. OVCAR-5	3.9	Small Intestine Pool	10.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	18.3
Ovarian ca. OVCAR-8	20.3	Bone Marrow Pool	6.7
Ovary	12.1	Fetal Heart	19.9
Breast ca. MCF-7	0.0	Heart Pool	12.2
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	60.3
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	3.8
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	24.0
Breast Pool	23.5	Thymus Pool	17.1
Trachea	2.8	CNS cancer (glio/astro) U87-MG	0.0
Lung	2.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	41.2	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	2.7
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	9.7
Lung ca. NCI-H526	0.0	Brain (cerebellum)	3.8
Lung ca. NCI-H23	0.0	Brain (fetal)	0.0

Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	16.6
Lung ca. HOP-62	0.9	Cerebral Cortex Pool	1.2
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	39.2
Liver	1.6	Brain (Thalamus) Pool	1.7
Fetal Liver	2.4	Brain (whole)	6.9
Liver ca. HepG2	0.0	Spinal Cord Pool	28.7
Kidney Pool	47.0	Adrenal Gland	4.7
Fetal Kidney	11.8	Pituitary gland Pool	4.6
Renal ca. 786-0	0.0	Salivary Gland	2.9
Renal ca. A498	0.0	Thyroid (female)	7.1
Renal ca. ACHN	1.6	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	10.8	Pancreas Pool	0.0

Table AG. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag918, Run 152842616	Tissue Name	Rel. Exp.(%) Ag918, Run 152842616
Liver adenocarcinoma	0.0	Kidney (fetal)	16.4
Pancreas	12.2	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	1.3
Adrenal gland	9.7	Renal ca. RXF 393	0.8
Thyroid	28.3	Renal ca. ACHN	1.4
Salivary gland	3.5	Renal ca. UO-31	2.7
Pituitary gland	8.7	Renal ca. TK-10	1.2
Brain (fetal)	1.2	Liver	4.5
Brain (whole)	19.1	Liver (fetal)	9.0
Brain (amygdala)	18.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	3.1	Lung	72.2
Brain (hippocampus)	100.0	Lung (fetal)	72.7
Brain (substantia nigra)	9.9	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	32.3	Lung ca. (small cell) NCI-H69	0.4
Cerebral Cortex	5.1	Lung ca. (s.cell var.) SHP-77	0.2
Spinal cord	16.2	Lung ca. (large cell) NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	0.3	Lung ca. (non-s.cell)	0.8

		NCI-H23	
astrocytoma SW1783	0.3	Lung ca. (non-s.cell) HOP-62	2.6
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.5
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.5
astrocytoma SNB-75	3.0	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	33.4
glioma U251	0.5	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	51.1	Breast ca.* (pl.ef) T47D	0.0
Heart	11.1	Breast ca. BT-549	0.3
Skeletal muscle (fetal)	82.9	Breast ca. MDA-N	0.0
Skeletal muscle	3.1	Ovary	21.3
Bone marrow	3.3	Ovarian ca. OVCAR-3	0.0
Thymus	3.6	Ovarian ca. OVCAR-4	0.0
Spleen	45.7	Ovarian ca. OVCAR-5	3.8
Lymph node	22.1	Ovarian ca. OVCAR-8	6.2
Colorectal	5.0	Ovarian ca. IGROV-1	0.4
Stomach	33.2	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	43.8	Uterus	31.2
Colon ca. SW480	0.0	Placenta	8.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	13.8
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.5
Colon ca. HCT-116	0.0	Testis	6.3
Colon ca. CaCo-2	0.2	Melanoma Hs688(A).T	49.0
Colon ca. tissue(ODO3866)	52.5	Melanoma* (met) Hs688(B).T	25.3
Colon ca. HCC-2998	2.6	Melanoma UACC-62	0.0

Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	9.9	Melanoma LOX IMVI	0.0
Trachea	18.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	7.7	Adipose	34.2

Table AH, Panel 4D

Tissue Name	Rel. Exp.(%) Ag918, Run 152842729	Tissue Name	Rel. Exp.(%) Ag918, Run 152842729
Secondary Th1 act	0.0	HUVEC IL-1beta	24.5
Secondary Th2 act	0.0	HUVEC IFN gamma	58.2
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	10.2
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	19.2
Secondary Th2 rest	0.0	HUVEC IL-11	44.8
Secondary Tr1 rest	0.0	Lung Microvascular EC none	55.1
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	43.5
Primary Th2 act	0.2	Microvascular Dermal EC none	37.1
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	26.6
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.3
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	2.5	Coronary artery SMC rest	7.7
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	7.7
CD8 lymphocyte act	0.0	Astrocytes rest	2.8
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	5.3
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-	0.0	CCD1106	0.0

CD95 CH11		(Keratinocytes) none	
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	6.7
LAK cells IL-2+IL-12	0.0	Lupus kidney	2.6
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.2
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.4
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.5
Two Way MLR 5 day	0.0	HPAEC none	85.9
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL- 1 beta	66.0
PBMC rest	0.0	Lung fibroblast none	15.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	3.1
PBMC PHA-L	0.0	Lung fibroblast IL-4	20.9
Ramos (B cell) none	0.0	Lung fibroblast IL-9	21.3
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	15.4
B lymphocytes PWM	0.9	Lung fibroblast IFN gamma	26.6
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	8.4
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	8.1
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	5.6
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	4.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	13.9
Dendritic cells anti- CD40	0.0	IBD Colitis 2	0.8
Monocytes rest	0.0	IBD Crohn's	5.7
Monocytes LPS	0.0	Colon	18.2
Macrophages rest	0.0	Lung	97.3
Macrophages LPS	0.0	Thymus	11.5
HUVEC none	47.6	Kidney	2.7
HUVEC starved	100.0		

CNS_neurodegeneration_v1.0 Summary: Ag918 This panel confirms the expression of the CG124728-02 gene at low levels in the brains of an independent group of individuals. No differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. See
5 Panel 1.3D for a discussion of this gene in treatment of central nervous system disorders.

Ag6492 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

General_screening_panel_v1.5 Summary: Ag6167 Low levels of expression of the CG124728-02 gene is restricted to colon cancer tissue (CT=34.9). Therefore,
10 expression of this gene may be used to distinguish this sample from other samples used in this panel and also as marker to detect the presence of colon cancer. In addition, therapeutic modulation of this gene may be useful in the treatment of colon cancer.

General_screening_panel_v1.6 Summary: Ag6492 Highest expression of the CG124728-02 gene is detected in colon cancer tissue (CT=32.8). In addition, low levels of
15 expression of this gene is also detected in two melanoma cell lines. Therefore, expression of this gene may be used to distinguish these samples from other samples used in this panel and also as marker to detect the presence of colon cancer and melanoma. Furthermore, therapeutic modulation of this gene may be useful in the treatment of colon cancer and melanoma.

20 In addition, low levels of expression of this gene is also found in brain (substantia nigra) and spinal cord. Therefore, therapeutic modulation of this gene may be useful in the treatment of neurological disorders.

Significant expression of this gene is also seen in some of the normal tissue samples derived from spleen, lymph node, bladder, kidney and fetal lung. Therefore,
25 therapeutic modulation of this gene may be useful in the treatment of diseases that affect these tissues.

Interestingly, expression of this gene is higher in fetal (CT=34) as compared to adult lung (CT=38). Therefore, expression of this gene may be useful in distinguishing between fetal and adult lung. In addition, the relative overexpression of this gene in fetal
30 lung suggests that the protein product may enhance lung growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung related diseases.

Panel 1.3D Summary: Ag918 Highest expression of the CG124728-02 gene is seen in brain (hippocampus) (CT=29.8). In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord.

- 5 Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal
10 muscle, heart, liver and the gastrointestinal tract. This gene encodes a splice variant of the complement C1q tumor necrosis factor-related protein 5 precursor, a member of the C1q family. This family includes proteins such as complement subunit C1q, adiponectin, gliacolin, C1q-related protein, cerebellin, CORS26 etc., all of which are secreted. These proteins have been implicated in tissue differentiation, immune regulation, energy
15 homeostasis, synaptic function and in diseases such as obesity, diabetes and neurodegeneration. Adiponectin, a member of C1q family and protein closely related to complement C1q tumor necrosis factor-related protein 5 precursor, is induced over 100-fold in adipocyte differentiation (Scherer *et al.*, 1995, J Biol Chem 270(45):26746-9 PMID: 7592907) and is involved in adipocyte signaling (Hu *et al.*, 1996, J Biol Chem
20 271(18):10697-703 PMID: 8631877). Recently, adiponectin has been shown to reverse insulin resistance in mouse models of lipoatrophy and obesity (Yamauchi *et al.*, 2001, Nat Med 7(8):941-6 PMID: 11479627). Therefore this protein, and proteins related to it, are potential antigens for development protein therapeutics for use in the treatment of obesity and type II diabetes.

- 25 Low to moderate levels of expression of this gene is also seen in ovarian cancer and melanoma cell lines, as well as in samples derived from colon cancer (ODO3866). Therefore, therapeutic modulation of this gene may be used for the treatment of these cancers.

Panel 4.1D Summary: Ag6167/Ag6492 Expression of this gene is
30 low/undetectable (CTs > 35) across all of the samples on this panel.

Panel 4D Summary: Ag918 Highest expression of the CG124728-02 gene is detected in starved HUVEC cells and lung (CT=29.2). In addition, moderate levels of expression of this gene is also detected stimulated and resting HUVEC, lung microvascular EC, microvascular dermal EC, HPAEC, lung and dermal fibroblasts,

coronary artery SMC, astrocytes and in normal tissues including colon, thymus and kidney. Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or antibodies, might be beneficial in the treatment of autoimmune and inflammatory diseases that involve these cell and tissue types, such as lupus erythematosus, asthma, emphysema, Crohn's disease, ulcerative colitis, rheumatoid arthritis, osteoarthritis, and psoriasis.

Low levels of expression of this gene is also seen in samples derived from liver cirrhosis and lupus erythematosus. Therefore, therapeutic modulation of this gene may be useful in the treatment of liver cirrhosis and lupus erythematosus.

Panel CNS_1.1 Summary: Ag6167 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

B. CG127616-01 and CG127616-02: Erythropoietin precursor

Expression of gene CG127616-01 and CG127616-02 was assessed using the primer-probe sets Ag4746 and Ag6361, described in Tables BA and BB. Results of the RTQ-PCR runs are shown in Tables BC, BD and BE. CG127616-01 represents full-length physical clones.

Table BA. Probe Name Ag4746

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gaatatcacgaaggaagcca-3'	20	155	88
Probe	TET-5'-cctcagctgctccactccgaaca-3'-TAMRA	23	193	89
Reverse	5'-cggaagtggtcagcagtgat-3'	20	216	90

Table BB. Probe Name Ag6361

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-acaatcactgctgacactttcc-3'	22	213	91
Probe	TET-5'-ttccgagtctactccaatttctctccg-3'-TAMRA	26	243	92
Reverse	5'-cctgtgtacagcttcagctttc-3'	22	271	93

Table BC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag4746, Run 214145483	Tissue Name	Rel. Exp.(%) Ag4746, Run 214145483
Adipose	0.0	Renal ca. TK-10	27.4
Melanoma* Hs688(A).T	0.0	Bladder	29.1

Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	1.2
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.5
Melanoma* SK- MEL-5	0.0	Colon ca. SW480	0.2
Squamous cell carcinoma SCC-4	0.6	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.8	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.6	Colon ca. HCT-116	2.4
Prostate Pool	1.0	Colon ca. CaCo-2	33.2
Placenta	1.4	Colon cancer tissue	0.0
Uterus Pool	2.8	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	1.3	Colon ca. Colo-205	0.0
Ovarian ca. SK- OV-3	0.2	Colon ca. SW-48	0.4
Ovarian ca. OVCAR-4	0.0	Colon Pool	12.3
Ovarian ca. OVCAR-5	0.3	Small Intestine Pool	0.2
Ovarian ca. IGROV-1	0.0	Stomach Pool	5.8
Ovarian ca. OVCAR-8	1.5	Bone Marrow Pool	1.9
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.7
Breast ca. MDA- MB-231	3.7	Lymph Node Pool	14.5
Breast ca. BT 549	3.1	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.4	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.2
Breast Pool	4.9	Thymus Pool	1.3
Trachea	0.0	CNS cancer (glio/astro) U87-MG	1.6
Lung	1.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.2	CNS cancer (neuro;met) SK-N-AS	0.2
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF- 539	0.0
Lung ca. LX-1	1.1	CNS cancer (astro)	0.2

		SNB-75	
Lung ca. NCI-H146	1.1	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.3	CNS cancer (glio) SF-295	1.4
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.1
Lung ca. NCI-H23	18.0	Brain (fetal)	1.1
Lung ca. NCI-H460	0.2	Brain (Hippocampus) Pool	0.4
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	6.1
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	18.8	Brain (whole)	4.8
Liver ca. HepG2	100.0	Spinal Cord Pool	0.0
Kidney Pool	4.9	Adrenal Gland	0.0
Fetal Kidney	1.0	Pituitary gland Pool	0.0
Renal ca. 786-0	1.3	Salivary Gland	0.0
Renal ca. A498	0.5	Thyroid (female)	0.0
Renal ca. ACHN	0.4	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	7.7

Table BD. General_screening_panel_v1.6

Tissue Name	Rel. Exp.(%) Ag6361, Run 277221717	Tissue Name	Rel. Exp.(%) Ag6361, Run 277221717
Adipose	0.4	Renal ca. TK-10	44.8
Melanoma* Hs688(A).T	0.0	Bladder	38.4
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.1
Melanoma* M14	0.6	Gastric ca. KATO III	0.4
Melanoma* LOXIMVI	0.2	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.5	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.6	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.8	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	1.0	Colon ca. HCT-116	2.2

Prostate Pool	1.5	Colon ca. CaCo-2	32.5
Placenta	2.8	Colon cancer tissue	0.4
Uterus Pool	1.1	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.4	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	8.5
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	2.5
Ovarian ca. IGROV-1	0.2	Stomach Pool	1.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	3.6
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.3	Heart Pool	2.2
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	8.6
Breast ca. BT 549	0.4	Fetal Skeletal Muscle	0.3
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	9.1	Thymus Pool	1.3
Trachea	0.4	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.2	CNS cancer (glio/astro) U-118-MG	0.5
Fetal Lung	0.6	CNS cancer (neuro;met) SK-N-AS	0.1
Lung ca. NCI-N417	0.2	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	1.0	CNS cancer (astro) SNB-75	1.1
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.8
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.2
Lung ca. NCI-H526	0.0	Brain (cerebellum)	1.1
Lung ca. NCI-H23	0.0	Brain (fetal)	0.9
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.2
Lung ca. HOP-62	0.2	Cerebral Cortex Pool	0.1

Lung ca. NCI-H522	0.5	Brain (Substantia nigra) Pool	0.2
Liver	0.6	Brain (Thalamus) Pool	0.0
Fetal Liver	24.3	Brain (whole)	3.1
Liver ca. HepG2	100.0	Spinal Cord Pool	0.1
Kidney Pool	4.8	Adrenal Gland	0.0
Fetal Kidney	0.6	Pituitary gland Pool	0.2
Renal ca. 786-0	0.0	Salivary Gland	0.5
Renal ca. A498	0.2	Thyroid (female)	0.0
Renal ca. ACHN	0.6	Pancreatic ca. CAPAN2	0.3
Renal ca. UO-31	0.2	Pancreas Pool	9.9

Table BE. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4746, Run 204171268	Rel. Exp.(%) Ag6361, Run 262456361	Tissue Name	Rel. Exp.(%) Ag4746, Run 204171268	Rel. Exp.(%) Ag6361, Run 262456361
Secondary Th1 act	0.0	2.1	HUVEC IL-1beta	0.0	0.0
Secondary Th2 act	0.0	2.5	HUVEC IFN gamma	0.0	0.0
Secondary Tr1 act	3.8	0.0	HUVEC TNF alpha + IFN gamma	0.0	0.0
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	0.0	0.0
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	0.0	0.0
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC none	0.0	1.2
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0	1.4
Primary Th2 act	0.0	0.0	Microvascular Dermal EC none	0.0	0.0
Primary Tr1 act	0.0	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0	0.0
Primary Th1 rest	2.1	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0	0.0
Primary Th2 rest	0.0	0.0	Small airway	0.0	0.0

			epithelium none		
Primary Tr1 rest	0.0	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0	0.0
CD45RA CD4 lymphocyte act	0.0	0.0	Coronary artery SMC rest	0.0	0.0
CD45RO CD4 lymphocyte act	13.5	1.7	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0
CD8 lymphocyte act	0.0	0.0	Astrocytes rest	13.7	0.0
Secondary CD8 lymphocyte rest	1.0	1.3	Astrocytes TNFalpha + IL-1beta	3.5	7.8
Secondary CD8 lymphocyte act	0.0	0.0	KU-812 (Basophil) rest	0.0	1.8
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	0.0	2.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	0.0	CCD1106 (Keratinocytes) none	0.0	0.0
LAK cells rest	0.0	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0	3.2
LAK cells IL-2	0.0	0.0	Liver cirrhosis	100.0	100.0
LAK cells IL-2+IL-12	0.0	0.0	NCI-H292 none	0.0	0.0
LAK cells IL-2+IFN gamma	0.0	1.3	NCI-H292 IL-4	0.0	0.0
LAK cells IL-2+IL-18	0.0	1.5	NCI-H292 IL-9	6.9	0.0
LAK cells PMA/ionomycin	0.0	0.0	NCI-H292 IL-13	0.0	0.0
NK Cells IL-2 rest	0.0	0.7	NCI-H292 IFN gamma	0.0	0.0
Two Way MLR 3 day	0.0	0.0	HPAEC none	0.0	0.0
Two Way MLR 5 day	0.0	0.0	HPAEC TNF alpha + IL-1 beta	0.0	0.0
Two Way MLR 7 day	0.0	0.0	Lung fibroblast none	0.0	0.0
PBMC rest	0.0	0.0	Lung fibroblast TNF alpha + IL-1	0.0	0.0

			beta		
PBMC PWM	0.0	0.0	Lung fibroblast IL-4	0.0	0.0
PBMC PHA-L	0.0	0.0	Lung fibroblast IL-9	0.0	0.0
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-13	0.0	0.0
Ramos (B cell) ionomycin	0.0	1.0	Lung fibroblast IFN gamma	0.0	0.0
B lymphocytes PWM	0.0	0.0	Dermal fibroblast CCD1070 rest	0.0	0.0
B lymphocytes CD40L and IL-4	3.0	0.0	Dermal fibroblast CCD1070 TNF alpha	1.0	0.0
EOL-1 dbcAMP	0.0	1.9	Dermal fibroblast CCD1070 IL-1 beta	0.0	1.8
EOL-1 dbcAMP PMA/ionomycin	0.0	0.0	Dermal fibroblast IFN gamma	0.0	2.7
Dendritic cells none	0.0	0.0	Dermal fibroblast IL-4	11.1	4.4
Dendritic cells LPS	0.0	0.0	Dermal Fibroblasts rest	0.0	0.0
Dendritic cells anti-CD40	0.0	0.0	Neutrophils TNFa+LPS	0.0	0.0
Monocytes rest	3.2	0.0	Neutrophils rest	0.0	0.0
Monocytes LPS	0.0	0.0	Colon	0.0	0.0
Macrophages rest	0.0	0.0	Lung	0.0	0.0
Macrophages LPS	0.0	0.0	Thymus	3.4	1.3
HUVEC none	0.0	0.0	Kidney	5.6	9.2
HUVEC starved	0.0	0.0			

CNS_neurodegeneration_v1.0 Summary: Ag4746/Ag6361 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

General_screening_panel_v1.4 Summary: Ag4746 Highest expression of this gene is seen in a liver cancer cell line (CT=30.3). Moderate levels of expression are also
5 seen in colon, renal and lung cancer cell lines, with low but significant expression detectable in lymph node and fetal liver. The transcript for this gene encodes a putative variant of erythropoietin (Epo), which is produced in the kidney and liver of normal adults. Human erythropoietin (Epo) is an acidic glycoprotein hormone that mediates the production of red blood cells, promotes erythroid differentiation, initiates hemoglobin
10 synthesis. In addition, Epo has been shown to be a potent growth factor for the

development of red blood cells from hematopoietic stem cells. Thus, the expression in hematopoietic tissues in this panel is consistent with the characterization of this novel protein as a novel variant of Epo. Therefore, modulation of the expression or function of this gene product may will be useful in the treatment of hematopoietic disorders.

5 **General_screening_panel_v1.6 Summary:** Ag6361 Expression on this panel is consistent with expression on Panel 1.4. Highest expression is seen in a liver cancer cell line (CT=30), with moderate levels of expression seen in colon and renal cancer cell lines, with low but significant levels of expression of this gene in kidney, fetal liver, and lymph node. See Panel 1.4 for discussion of this gene in inflammation.

10 **Panel 4.1D Summary:** Ag4746/Ag6361 Two experiments with two probe and primer sets produce results that are in excellent agreement. Significant expression is limited to the cirrhotic liver (CTs=32-33), consistent with expression in hematopoietic tissues seen in Panels 1.4 and 1.6.

C. CG128348-01: Apolipoprotein E precursor-like.

15 Expression of gene CG128348-01 was assessed using the primer-probe set Ag6834, described in Table CA. Results of the RTQ-PCR runs are shown in Tables CB, CC and CD.

Table CA. Probe Name Ag6834

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cgcgacgaggtgaa-3'	15	386	94
Probe	TET-5'-cgcgacacctcgccacct-3'-TAMRA	18	406	95
Reverse	5'-gtatctgctgggcctgct-3'	18	436	96

Table CB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag6834, Run 278022745	Tissue Name	Rel. Exp.(%) Ag6834, Run 278022745
AD 1 Hippo	59.0	Control (Path) 3 Temporal Ctx	15.0
AD 2 Hippo	100.0	Control (Path) 4 Temporal Ctx	20.2
AD 3 Hippo	13.0	AD 1 Occipital Ctx	16.2
AD 4 Hippo	25.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	40.1	AD 3 Occipital Ctx	7.6
AD 6 Hippo	51.1	AD 4 Occipital Ctx	25.5
Control 2 Hippo	40.9	AD 5 Occipital Ctx	21.8
Control 4 Hippo	58.2	AD 6 Occipital Ctx	19.3

Control (Path) 3 Hippo	12.2	Control 1 Occipital Ctx	5.4
AD 1 Temporal Ctx	15.4	Control 2 Occipital Ctx	33.2
AD 2 Temporal Ctx	51.4	Control 3 Occipital Ctx	16.3
AD 3 Temporal Ctx	8.3	Control 4 Occipital Ctx	20.6
AD 4 Temporal Ctx	25.2	Control (Path) 1 Occipital Ctx	41.2
AD 5 Inf Temporal Ctx	35.1	Control (Path) 2 Occipital Ctx	8.1
AD 5 Sup Temporal Ctx	48.6	Control (Path) 3 Occipital Ctx	7.0
AD 6 Inf Temporal Ctx	47.6	Control (Path) 4 Occipital Ctx	9.7
AD 6 Sup Temporal Ctx	31.6	Control 1 Parietal Ctx	15.2
Control 1 Temporal Ctx	17.4	Control 2 Parietal Ctx	42.9
Control 2 Temporal Ctx	45.4	Control 3 Parietal Ctx	33.0
Control 3 Temporal Ctx	25.7	Control (Path) 1 Parietal Ctx	22.4
Control 3 Temporal Ctx	33.9	Control (Path) 2 Parietal Ctx	20.3
Control (Path) 1 Temporal Ctx	29.5	Control (Path) 3 Parietal Ctx	10.4
Control (Path) 2 Temporal Ctx	26.6	Control (Path) 4 Parietal Ctx	17.6

Table CC. General_screening_panel_v1.6

Tissue Name	Rel. Exp.(%) Ag6834, Run 278019622	Tissue Name	Rel. Exp.(%) Ag6834, Run 278019622
Adipose	0.8	Renal ca. TK-10	15.2
Melanoma* Hs688(A).T	0.0	Bladder	5.9
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.1
Melanoma* M14	4.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.1	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.2	Colon ca. SW480	3.4

Squamous cell carcinoma SCC-4	0.2	Colon ca.* (SW480 met) SW620	0.3
Testis Pool	1.5	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.1	Colon ca. HCT-116	0.1
Prostate Pool	0.0	Colon ca. CaCo-2	44.8
Placenta	1.1	Colon cancer tissue	39.0
Uterus Pool	0.5	Colon ca. SW1116	0.1
Ovarian ca. OVCAR-3	7.2	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.1	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.5	Colon Pool	1.2
Ovarian ca. OVCAR-5	100.0	Small Intestine Pool	0.3
Ovarian ca. IGROV-1	0.7	Stomach Pool	0.9
Ovarian ca. OVCAR-8	64.6	Bone Marrow Pool	0.7
Ovary	11.3	Fetal Heart	0.3
Breast ca. MCF-7	0.3	Heart Pool	0.4
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.7
Breast ca. BT 549	0.5	Fetal Skeletal Muscle	0.5
Breast ca. T47D	3.0	Skeletal Muscle Pool	0.1
Breast ca. MDA-N	0.0	Spleen Pool	5.3
Breast Pool	2.4	Thymus Pool	3.1
Trachea	0.7	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.9	CNS cancer (glio/astro) U-118-MG	0.2
Fetal Lung	0.2	CNS cancer (neuro;met) SK-N-AS	1.1
Lung ca. NCI-N417	0.2	CNS cancer (astro) SF-539	0.2
Lung ca. LX-1	0.1	CNS cancer (astro) SNB-75	0.7
Lung ca. NCI-H146	0.1	CNS cancer (glio) SNB-19	0.1
Lung ca. SHP-77	0.2	CNS cancer (glio) SF-295	0.1
Lung ca. A549	0.3	Brain (Amygdala) Pool	20.9

Lung ca. NCI-H526	1.4	Brain (cerebellum)	11.5
Lung ca. NCI-H23	0.7	Brain (fetal)	3.5
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	17.6
Lung ca. HOP-62	3.4	Cerebral Cortex Pool	9.0
Lung ca. NCI-H522	9.9	Brain (Substantia nigra) Pool	26.4
Liver	31.2	Brain (Thalamus) Pool	12.6
Fetal Liver	26.1	Brain (whole)	7.7
Liver ca. HepG2	52.9	Spinal Cord Pool	28.1
Kidney Pool	0.9	Adrenal Gland	52.5
Fetal Kidney	2.9	Pituitary gland Pool	0.6
Renal ca. 786-0	0.0	Salivary Gland	0.4
Renal ca. A498	0.0	Thyroid (female)	8.1
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.1
Renal ca. UO-31	0.7	Pancreas Pool	1.6

Table CD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag6834, Run 278022645	Tissue Name	Rel. Exp.(%) Ag6834, Run 278022645
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.7
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.1
Secondary Th2 rest	0.0	HUVEC IL-11	0.3
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.1	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.2	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	3.9
Primary Th2 rest	0.0	Small airway epithelium none	1.4
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.1
CD45RA CD4	0.0	Coronary artery SMC	0.3

lymphocyte act		rest	
CD45RO CD4 lymphocyte act	0.1	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.3
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	1.9
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	5.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	3.5
LAK cells rest	30.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.7
LAK cells IL-2	0.7	Liver cirrhosis	16.8
LAK cells IL-2+IL-12	1.0	NCI-H292 none	0.6
LAK cells IL-2+IFN gamma	1.2	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.3	NCI-H292 IL-9	1.1
LAK cells PMA/ionomycin	35.1	NCI-H292 IL-13	0.6
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	1.4	HPAEC none	0.0
Two Way MLR 5 day	2.1	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	1.6	Lung fibroblast none	1.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.5
PBMC PWM	0.0	Lung fibroblast IL-4	0.7
PBMC PHA-L	0.3	Lung fibroblast IL-9	0.6
Ramos (B cell) none	0.0	Lung fibroblast IL-13	1.4
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	1.4
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	1.1
B lymphocytes CD40L and IL-4	0.4	Dermal fibroblast CCD1070 TNF alpha	0.6
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.2
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	22.7	Dermal fibroblast IL-4	0.8
Dendritic cells LPS	30.6	Dermal Fibroblasts rest	0.0

Dendritic cells anti-CD40	31.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	6.0	Colon	1.8
Macrophages rest	92.7	Lung	11.4
Macrophages LPS	100.0	Thymus	1.4
HUVEC none	0.4	Kidney	94.6
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag6834 This profile confirms the expression of this gene at moderate levels in the brain. See Panel 1.4 for discussion of this gene in the central nervous system.

General_screening_panel_v1.6 Summary: Ag6834 Highest expression of this gene is seen in an ovarian cancer cell line (CT=28.3). Moderate levels of expression are also seen in the brain, adrenal, liver, ovary, and spleen. This expression is consistent with the characterization of the protein encoded by this gene as a putative apolipoprotein E molecule. Apolipoprotein E, a plasma protein, is a protein component of very-low-density lipoprotein (VLDL) and chylomicrons that is involved in the transport of cholesterol and other hydrophobic molecules.

Certain forms of apolipoprotein E have been linked to disorders of cholesterol metabolism and coronary heart disease. Interestingly, ApoE4 is a major risk factor for Alzheimer's disease and may play a similar role in other types of dementia. Masliah suggested that apoE may play a role in maintaining the stability of the synapto-dendritic apparatus (Cedazo-Minguez A, J Cell Mol Med 2001 Jul-Sep;5(3):254-66; Masliah E, Prog Neurobiol 1996 Dec;50(5-6):493-503). Thus, based on the homology of this protein to apolipoprotein E, and its localization to the brain and vascular tissues, modulation of the expression or function of this gene may be useful in the treatment of neurological and/or cardiovascular disorders.

Panel 4.1D Summary: Ag6834 Highest expression of this gene is seen in macrophages (CT=29), with moderate levels of expression also detected in dendritic cells, LAK cells, and normal kidney. Since macrophages are one of the cell types that synthesize apolipoprotein E, prominent expression in macrophages on this panel is in agreement with the identification of the protein encoded by this gene as a novel apolipoprotein E. Boisvert *et al.* demonstrated the contribution of macrophage-derived apoE to hepatic clearance of serum cholesterol by performing bone marrow transplantation on hypercholesterolemic apoE-deficient 'knockout' mice and showing dramatic reduction in serum cholesterol

levels the extent of atherosclerosis in the bone marrow-treated mice. (J Clin Invest 1995 Aug;96(2):1118-24). Thus, expression in this panel is consistent with expression in Panel 1.6 and suggests that therapeutic modulation of the expression or function of this gene product may be useful in the treatment of cardiovascular disease.

5 D. CG128348-02: Apolipoprotein E Precursor.

Expression of gene CG128348-02 was assessed using the primer-probe set Ag6835, described in Table DA. Results of the RTQ-PCR runs are shown in Tables DB, DC and DD.

Table DA. Probe Name Ag6835

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cctggacgaggtgaagga-3'	18	503	97
Probe	TET-5'-cagcttggcgcgcacctcc-3'-TAMRA	19	530	98
Reverse	5'-cctgcaggcgtatctgct-3'	18	562	99

10 Table DB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag6835, Run 278022746	Tissue Name	Rel. Exp.(%) Ag6835, Run 278022746
AD 1 Hippo	51.4	Control (Path) 3 Temporal Ctx	11.9
AD 2 Hippo	100.0	Control (Path) 4 Temporal Ctx	17.6
AD 3 Hippo	11.0	AD 1 Occipital Ctx	14.2
AD 4 Hippo	31.2	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	33.2	AD 3 Occipital Ctx	10.7
AD 6 Hippo	54.0	AD 4 Occipital Ctx	24.8
Control 2 Hippo	42.3	AD 5 Occipital Ctx	14.3
Control 4 Hippo	55.5	AD 6 Occipital Ctx	18.6
Control (Path) 3 Hippo	9.0	Control 1 Occipital Ctx	5.0
AD 1 Temporal Ctx	16.7	Control 2 Occipital Ctx	28.5
AD 2 Temporal Ctx	49.0	Control 3 Occipital Ctx	14.3
AD 3 Temporal Ctx	11.4	Control 4 Occipital Ctx	20.0

AD 4 Temporal Ctx	25.7	Control (Path) 1 Occipital Ctx	41.8
AD 5 Inf Temporal Ctx	27.2	Control (Path) 2 Occipital Ctx	6.7
AD 5 Sup Temporal Ctx	38.7	Control (Path) 3 Occipital Ctx	5.3
AD 6 Inf Temporal Ctx	35.4	Control (Path) 4 Occipital Ctx	6.5
AD 6 Sup Temporal Ctx	28.5	Control 1 Parietal Ctx	12.6
Control 1 Temporal Ctx	19.9	Control 2 Parietal Ctx	39.0
Control 2 Temporal Ctx	46.3	Control 3 Parietal Ctx	24.5
Control 3 Temporal Ctx	21.6	Control (Path) 1 Parietal Ctx	27.4
Control 4 Temporal Ctx	34.9	Control (Path) 2 Parietal Ctx	17.9
Control (Path) 1 Temporal Ctx	24.1	Control (Path) 3 Parietal Ctx	8.7
Control (Path) 2 Temporal Ctx	21.3	Control (Path) 4 Parietal Ctx	15.6

Table DC. General_screening_panel_v1.6

Tissue Name	Rel. Exp.(%) Ag6835, Run 278019624	Tissue Name	Rel. Exp.(%) Ag6835, Run 278019624
Adipose	1.4	Renal ca. TK-10	17.1
Melanoma* Hs688(A).T	0.0	Bladder	5.8
Melanoma* Hs688(B).T	0.1	Gastric ca. (liver met.) NCI-N87	0.2
Melanoma* M14	1.8	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.1
Melanoma* SK- MEL-5	0.3	Colon ca. SW480	2.9
Squamous cell carcinoma SCC-4	0.1	Colon ca.* (SW480 met) SW620	0.2
Testis Pool	1.5	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.2	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	45.7
Placenta	1.6	Colon cancer tissue	49.0
Uterus Pool	0.4	Colon ca. SW1116	0.0

Ovarian ca. OVCAR-3	6.8	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.1	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.6	Colon Pool	1.3
Ovarian ca. OVCAR-5	100.0	Small Intestine Pool	0.7
Ovarian ca. IGROV-1	0.3	Stomach Pool	0.5
Ovarian ca. OVCAR-8	73.2	Bone Marrow Pool	0.7
Ovary	14.6	Fetal Heart	0.2
Breast ca. MCF-7	0.3	Heart Pool	0.4
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.8
Breast ca. BT 549	0.8	Fetal Skeletal Muscle	0.6
Breast ca. T47D	4.1	Skeletal Muscle Pool	0.3
Breast ca. MDA-N	0.0	Spleen Pool	4.8
Breast Pool	1.8	Thymus Pool	3.0
Trachea	0.4	CNS cancer (glio/astro) U87-MG	0.1
Lung	2.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.2	CNS cancer (neuro;met) SK-N-AS	0.5
Lung ca. NCI-N417	0.1	CNS cancer (astro) SF-539	0.2
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.5
Lung ca. NCI-H146	0.2	CNS cancer (glio) SNB-19	0.4
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	1.1	Brain (Amygdala) Pool	20.7
Lung ca. NCI-H526	0.6	Brain (cerebellum)	15.8
Lung ca. NCI-H23	0.8	Brain (fetal)	4.5
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	23.8
Lung ca. HOP-62	3.3	Cerebral Cortex Pool	10.5
Lung ca. NCI-H522	13.2	Brain (Substantia nigra) Pool	34.6
Liver	27.7	Brain (Thalamus) Pool	13.7

Fetal Liver	43.2	Brain (whole)	15.4
Liver ca. HepG2	59.0	Spinal Cord Pool	26.2
Kidney Pool	1.1	Adrenal Gland	51.8
Fetal Kidney	1.1	Pituitary gland Pool	0.8
Renal ca. 786-0	0.0	Salivary Gland	0.6
Renal ca. A498	0.1	Thyroid (female)	7.7
Renal ca. ACHN	0.2	Pancreatic ca. CAPAN2	0.2
Renal ca. UO-31	0.6	Pancreas Pool	1.6

Table DD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag6835, Run 278022647	Tissue Name	Rel. Exp.(%) Ag6835, Run 278022647
Secondary Th1 act	0.1	HUVEC IL-1beta	0.2
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	1.8
Primary Th2 rest	0.0	Small airway epithelium none	1.5
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.3
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.1
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	1.8

CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	3.2
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	1.2
LAK cells rest	17.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.3
LAK cells IL-2	0.1	Liver cirrhosis	13.6
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.3
LAK cells IL-2+IFN gamma	0.2	NCI-H292 IL-4	0.4
LAK cells IL-2+ IL-18	0.1	NCI-H292 IL-9	0.5
LAK cells PMA/ionomycin	32.8	NCI-H292 IL-13	0.2
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.5
Two Way MLR 3 day	0.7	HPAEC none	0.0
Two Way MLR 5 day	1.2	HPAEC TNF alpha + IL- 1 beta	0.0
Two Way MLR 7 day	0.9	Lung fibroblast none	0.1
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.3
PBMC PWM	0.0	Lung fibroblast IL-4	0.7
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.9
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.6
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	1.8
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.4
B lymphocytes CD40L and IL-4	0.3	Dermal fibroblast CCD1070 TNF alpha	0.3
EOL-1 dbcAMP	0.1	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	13.9	Dermal fibroblast IL-4	0.1
Dendritic cells LPS	16.6	Dermal Fibroblasts rest	0.1
Dendritic cells anti- CD40	19.3	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	5.6	Colon	0.2
Macrophages rest	46.7	Lung	5.6
Macrophages LPS	100.0	Thymus	0.2
HUVEC none	0.0	Kidney	65.5

HUVEC starved	0.0		
---------------	-----	--	--

CNS_neurodegeneration_v1.0 Summary: Ag6835 This profile confirms the expression of this gene at moderate levels in the brain. This expression is consistent with expression seen in CG128348-01. See that gene for discussion of this gene in CNS disorders.

- 5 **General_screening_panel_v1.6 Summary:** Ag6835 Highest expression of this gene is seen in an ovarian cancer cell line (CT=26). Moderate levels of expression are also seen in two colon cancer cell lines, brain, adrenal, liver, ovary, and spleen. This expression is consistent with expression seen in CG128348-01. See that gene for discussion of this gene in disease.
- 10 **Panel 4.1D Summary:** Ag6835 Highest expression of this gene is seen in macrophages (CT=27), with prominent levels of expression also detected in dendritic cells, LAK cells, and normal kidney. This expression is consistent with expression seen in CG128348-01. See that gene for discussion of this gene in disease.

E. CG128348-03: Apolipoprotein E Precursor

- 15 Expression of gene CG128348-03 was assessed using the primer-probe set Ag6836, described in Table EA.

Table EA. Probe Name Ag6836

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gccgatgacctggagga-3'	17	525	100
Probe	TET-5'-cccagcagatacgctgcaggc-3'-TAMRA	22	547	101
Reverse	5'-aggcgggcctggaag-3'	15	575	102

CNS_neurodegeneration_v1.0 Summary: Ag6836 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

- 20 **General_screening_panel_v1.6 Summary:** Ag6836 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

Panel 4.1D Summary: Ag6836 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

F. CG129136-01: 4-1BB LIGAND.

Expression of gene CG129136-01 was assessed using the primer-probe set Ag4938, described in Table FA. Results of the RTQ-PCR runs are shown in Tables FB and FC.

Table FA. Probe Name Ag4938

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gcccaaggctggagtctactatg-3'	22	437	103
Probe	TET-5'-tcacttgcgctgcacctgca-3'-TAMRA	20	361	104
Reverse	5'-cctggcctcagtggaagat-3'	20	201	105

5 Table FB. General_screening_panel_v1.5

Tissue Name	Rel. Exp.(%) Ag4938, Run 228850853	Tissue Name	Rel. Exp.(%) Ag4938, Run 228850853
Adipose	0.0	Renal ca. TK-10	5.7
Melanoma* Hs688(A).T	2.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.6
Melanoma* M14	4.0	Gastric ca. KATO III	1.8
Melanoma* LOXIMVI	1.4	Colon ca. SW-948	58.2
Melanoma* SK- MEL-5	0.9	Colon ca. SW480	2.3
Squamous cell carcinoma SCC-4	0.8	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	12.3
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	8.8
Uterus Pool	0.0	Colon ca. SW1116	3.8
Ovarian ca. OVCAR-3	3.5	Colon ca. Colo-205	4.5
Ovarian ca. SK- OV-3	27.0	Colon ca. SW-48	0.5
Ovarian ca. OVCAR-4	2.9	Colon Pool	0.0
Ovarian ca. OVCAR-5	12.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	73.7	Stomach Pool	0.0
Ovarian ca. OVCAR-8	38.7	Bone Marrow Pool	0.0

Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	2.2	Lymph Node Pool	0.0
Breast ca. BT 549	5.9	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	7.9	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.7	CNS cancer (glio/astro) U87-MG	27.5
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.5
Fetal Lung	0.9	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.4
Lung ca. LX-1	0.6	CNS cancer (astro) SNB-75	9.9
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	100.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	1.3
Lung ca. A549	5.5	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	2.1	Brain (cerebellum)	0.6
Lung ca. NCI-H23	0.4	Brain (fetal)	0.8
Lung ca. NCI-H460	13.7	Brain (Hippocampus) Pool	0.9
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	4.7
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	17.6	Salivary Gland	0.0
Renal ca. A498	3.4	Thyroid (female)	0.0
Renal ca. ACHN	2.4	Pancreatic ca. CAPAN2	1.4
Renal ca. UO-31	9.1	Pancreas Pool	0.0

Table FC. Panel 4.1D

Tissue Name	Rel. Exp.(%)	Tissue Name	Rel. Exp.(%)
-------------	--------------	-------------	--------------

	Ag4938, Run 225245203		Ag4938, Run 225245203
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	2.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	3.8	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	1.8	Lung Microvascular EC none	0.0
Primary Th1 act	1.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	1.3	Microvascular Dermal EC none	0.0
Primary Tr1 act	3.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	4.7
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	1.2
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	1.7
Secondary CD8 lymphocyte rest	0.7	Astrocytes TNFalpha + IL-1beta	4.9
Secondary CD8 lymphocyte act	0.9	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	1.4

LAK cells PMA/ionomycin	2.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.6
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.6	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	4.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	2.9	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	1.1
Dendritic cells LPS	1.2	Dermal Fibroblasts rest	1.4
Dendritic cells anti- CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.7
Monocytes LPS	4.1	Colon	1.1
Macrophages rest	0.0	Lung	3.0
Macrophages LPS	3.9	Thymus	6.0
HUVEC none	0.0	Kidney	100.0
HUVEC starved	0.0		

General_screening_panel_v1.5 Summary: Ag4938 Highest expression of this

gene is seen in a brain cancer cell line (CT=31.7). Overall, expression of this gene appears to be associated exclusively with cancer cell lines, with moderate levels of expression also seen in cell lines from colon, ovarian, lung, and renal cancers. Thus, expression of this

5 gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of colon, ovarian, lung, and renal cancers. This gene is a variant of CG129136-02. See the

discussion of expression for that gene for further discussion of this gene in the treatment of immunological disease and cancer.

Panel 4.1D Summary: Ag4938 This gene is only expressed at detectable levels in the kidney (CT=32.7). Thus, expression of this gene could be used to differentiate the kidney derived sample from other samples on this panel and as a marker of kidney tissue. In addition, therapeutic targeting of the expression or function of this gene may modulate kidney function and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

G. CG129136-03: 4-1BB LIGAND.

Expression of gene CG129136-03 was assessed using the primer-probe sets Ag6625 and Ag6950, described in Tables GA and GB. Results of the RTQ-PCR runs are shown in Tables GC and GD.

Table GA. Probe Name Ag6625

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gaagcgtcagaggcgtattc-3'	20	78	106
Probe	TET-5'-atggaatacgcctctgacgctgtcat-3'-TAMRA	26	51	107
Reverse	5'-ctgacgctgtcaggaatacg-3'	20	17	108

Table GB. Probe Name Ag6950

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gccggcagcgagca-3'	14	193	109
Probe	TET-5'-agcagcagccccgcgaccagggcc-3'-TAMRA	24	161	110
Reverse	5'-tcgcgcctgccgcgta-3'	16	137	111

Table GC. General screening panel_v1.6

Tissue Name	Rel. Exp.(%) Ag6950, Run 278388889	Tissue Name	Rel. Exp.(%) Ag6950, Run 278388889
Adipose	0.0	Renal ca. TK-10	15.8
Melanoma* Hs688(A).T	1.7	Bladder	1.1
Melanoma* Hs688(B).T	3.2	Gastric ca. (liver met.) NCI-N87	3.2
Melanoma* M14	4.4	Gastric ca. KATO III	6.0
Melanoma* LOXIMVI	7.1	Colon ca. SW-948	100.0

Melanoma* SK-MEL-5	8.1	Colon ca. SW480	4.2
Squamous cell carcinoma SCC-4	5.0	Colon ca.* (SW480 met) SW620	0.9
Testis Pool	0.4	Colon ca. HT29	0.3
Prostate ca.* (bone met) PC-3	0.9	Colon ca. HCT-116	13.9
Prostate Pool	0.0	Colon ca. CaCo-2	0.6
Placenta	0.0	Colon cancer tissue	23.7
Uterus Pool	0.3	Colon ca. SW1116	6.7
Ovarian ca. OVCAR-3	0.4	Colon ca. Colo-205	12.2
Ovarian ca. SK-OV-3	44.1	Colon ca. SW-48	2.3
Ovarian ca. OVCAR-4	7.5	Colon Pool	0.8
Ovarian ca. OVCAR-5	32.3	Small Intestine Pool	0.6
Ovarian ca. IGROV-1	54.0	Stomach Pool	2.9
Ovarian ca. OVCAR-8	41.5	Bone Marrow Pool	0.5
Ovary	0.0	Fetal Heart	0.3
Breast ca. MCF-7	0.5	Heart Pool	0.3
Breast ca. MDA-MB-231	2.6	Lymph Node Pool	0.3
Breast ca. BT 549	17.2	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.7	Skeletal Muscle Pool	0.2
Breast ca. MDA-N	6.1	Spleen Pool	0.3
Breast Pool	0.6	Thymus Pool	0.4
Trachea	1.7	CNS cancer (glio/astro) U87-MG	23.2
Lung	0.4	CNS cancer (glio/astro) U-118-MG	0.7
Fetal Lung	10.9	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.2	CNS cancer (astro) SF-539	1.9
Lung ca. LX-1	3.9	CNS cancer (astro) SNB-75	9.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	55.1
Lung ca. SHP-77	3.1	CNS cancer (glio) SF-295	8.7

Lung ca. A549	5.2	Brain (Amygdala) Pool	1.3
Lung ca. NCI-H526	2.2	Brain (cerebellum)	1.0
Lung ca. NCI-H23	1.5	Brain (fetal)	1.1
Lung ca. NCI-H460	14.6	Brain (Hippocampus) Pool	1.1
Lung ca. HOP-62	3.0	Cerebral Cortex Pool	1.4
Lung ca. NCI-H522	1.7	Brain (Substantia nigra) Pool	2.8
Liver	0.2	Brain (Thalamus) Pool	1.4
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	1.0	Spinal Cord Pool	0.8
Kidney Pool	0.1	Adrenal Gland	0.3
Fetal Kidney	0.2	Pituitary gland Pool	0.2
Renal ca. 786-0	28.3	Salivary Gland	0.2
Renal ca. A498	12.8	Thyroid (female)	0.1
Renal ca. ACHN	9.7	Pancreatic ca. CAPAN2	9.0
Renal ca. UO-31	12.9	Pancreas Pool	1.8

Table GD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag6950, Run 283838220	Tissue Name	Rel. Exp.(%) Ag6950, Run 283838220
Secondary Th1 act	3.1	HUVEC IL-1beta	0.0
Secondary Th2 act	7.2	HUVEC IFN gamma	0.0
Secondary Tr1 act	3.3	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	1.8	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	2.7	HUVEC IL-11	0.0
Secondary Tr1 rest	7.6	Lung Microvascular EC none	0.0
Primary Th1 act	5.7	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	11.8	Microvascular Dermal EC none	0.0
Primary Tr1 act	20.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.7	Bronchial epithelium TNFalpha + IL1beta	6.4
Primary Th2 rest	0.0	Small airway epithelium none	3.8
Primary Tr1 rest	0.0	Small airway epithelium	3.6

		TNFalpha + IL-1beta	
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	1.9
CD45RO CD4 lymphocyte act	5.1	Coronary artery SMC TNFalpha + IL-1beta	3.8
CD8 lymphocyte act	1.7	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	14.4
Secondary CD8 lymphocyte act	10.1	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	2.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	6.8	CCD1106 (Keratinocytes) none	1.5
LAK cells rest	1.4	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.5	NCI-H292 none	0.6
LAK cells IL-2+IFN gamma	2.4	NCI-H292 IL-4	1.0
LAK cells IL-2+ IL-18	2.1	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	15.6	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	1.5	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	1.6	HPAEC none	3.2
Two Way MLR 5 day	0.8	HPAEC TNF alpha + IL-1 beta	2.2
Two Way MLR 7 day	4.8	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	3.6	Lung fibroblast IL-4	0.0
PBMC PHA-L	3.4	Lung fibroblast IL-9	0.0
Ramos (B cell) none	1.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.7	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.6	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	6.3	Dermal fibroblast CCD1070 TNF alpha	1.9
EOL-1 dbcAMP	15.1	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	3.5	Dermal fibroblast IFN gamma	5.9

Dendritic cells none	0.0	Dermal fibroblast IL-4	6.7
Dendritic cells LPS	1.6	Dermal Fibroblasts rest	2.4
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	100.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	10.5	Thymus	1.9
HUVEC none	2.8	Kidney	0.0
HUVEC starved	0.0		

General_screening_panel_v1.6 Summary: Ag6625 Highest expression of this gene is seen in a colon cancer cell line (CT=28). Overall, expression of this gene appears to be more highly associated with cancer cell lines, with moderate levels of expression also seen in cell lines from brain, breast, ovarian, lung, melanoma and renal cancers. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of these cancers. This gene has homology to 4-1BBL, a cytokine that binds to TNFRSF9 and induces the proliferation of activated peripheral blood T cells and may also play a role in activation-induced cell death (AICD). Therefore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of colon, ovarian, lung, and renal cancers.

Panel 4.1D Summary: Ag6950 Highest expression of this gene is seen in LPS treated monocytes. (CT=31.7). Low but significant levels of expression are also seen in TNF- α and IL-1 β treated astrocytes, PMA/ionomycin treated LAK cells, eosinophils and LPS activated macrophages. This gene encodes a protein with homology to 4-1BB-L, a member of a family of receptors found on the surfaces of cells of the immune system that have been shown to enhance primary T-cell proliferation (Alderson, Eur J Immunol 1994 Sep;24(9):2219-27). Therefore, modulation of the expression of the putative cytokine encoded by this transcript may prevent the recruitment of monocytes and the initiation of the inflammatory process, and reduce the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, or rheumatoid arthritis.

H. CG133483-01: Win4 like protein

Expression of gene CG133483-01 was assessed using the primer-probe set Ag4836, described in Table HA.

Table HA. Probe Name Ag4836

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gctggaactgctccacact-3'	19	301	112
Probe	TET-5'-actccttgcccgtcttcggcaag-3'-TAMRA	23	322	113
Reverse	5'-gacatccactgtctcctccat-3'	21	373	114

General_screening_panel_v1.5 Summary: Ag4836 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

Panel 4.1D Summary: Ag4836 Expression of this gene is low/undetectable in all 5 samples on this panel (CTs>35).

I. CG135316-01: FGF14 splice variant.

Expression of gene CG135316-01 was assessed using the primer-probe sets Ag4899 and Ag5109, described in Tables IA and IB. Results of the RTQ-PCR runs are shown in Tables IC, ID and IE.

10 **Table IA. Probe Name Ag4899**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gacccaagtaaaagcacaag-3'	21	669	115
Probe	TET-5'-tgcgctctaaatccatttcagatatactccg-3'-TAMRA	30	690	116
Reverse	5'-tggcgttaagtttggttcatta-3'	22	729	117

Table IB. Probe Name Ag5109

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-atttcagatatactccgtcctgtt-3'	24	703	118
Probe	TET-5'-aatgaaccaaacttaacgccatcccc-3'-TAMRA	26	730	119
Reverse	5'-gggaacgcagccaga-3'	15	759	120

Table IC. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag5109, Run 233609875	Tissue Name	Rel. Exp.(%) Ag5109, Run 233609875
AD 1 Hippo	7.4	Control (Path) 3 Temporal Ctx	1.7
AD 2 Hippo	12.2	Control (Path) 4 Temporal Ctx	24.5
AD 3 Hippo	3.6	AD 1 Occipital Ctx	8.9
AD 4 Hippo	2.9	AD 2 Occipital Ctx (Missing)	0.0

AD 5 Hippo	92.0	AD 3 Occipital Ctx	3.5
AD 6 Hippo	30.8	AD 4 Occipital Ctx	7.2
Control 2 Hippo	14.6	AD 5 Occipital Ctx	28.7
Control 4 Hippo	3.6	AD 6 Occipital Ctx	14.5
Control (Path) 3 Hippo	2.0	Control 1 Occipital Ctx	1.3
AD 1 Temporal Ctx	7.9	Control 2 Occipital Ctx	52.5
AD 2 Temporal Ctx	13.2	Control 3 Occipital Ctx	12.2
AD 3 Temporal Ctx	2.7	Control 4 Occipital Ctx	1.3
AD 4 Temporal Ctx	13.8	Control (Path) 1 Occipital Ctx	64.2
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	8.0
AD 5 Sup Temporal Ctx	21.5	Control (Path) 3 Occipital Ctx	1.6
AD 6 Inf Temporal Ctx	29.9	Control (Path) 4 Occipital Ctx	14.5
AD 6 Sup Temporal Ctx	25.3	Control 1 Parietal Ctx	1.7
Control 1 Temporal Ctx	1.2	Control 2 Parietal Ctx	24.3
Control 2 Temporal Ctx	30.8	Control 3 Parietal Ctx	13.1
Control 3 Temporal Ctx	10.2	Control (Path) 1 Parietal Ctx	70.2
Control 3 Temporal Ctx	3.0	Control (Path) 2 Parietal Ctx	13.6
Control (Path) 1 Temporal Ctx	35.8	Control (Path) 3 Parietal Ctx	1.1
Control (Path) 2 Temporal Ctx	25.2	Control (Path) 4 Parietal Ctx	32.3

Table ID. General_screening_panel_v1.5

Tissue Name	Rel. Exp.(%) Ag5109, Run 228969350	Tissue Name	Rel. Exp.(%) Ag5109, Run 228969350
Adipose	0.4	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	42.9	Bladder	2.1
Melanoma* Hs688(B).T	17.1	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.3	Gastric ca. KATO III	0.0

Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	37.9	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	4.7	Colon ca.* (SW480 met) SW620	1.3
Testis Pool	1.3	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	5.6	Colon ca. HCT-116	0.0
Prostate Pool	0.6	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	1.1
Uterus Pool	0.6	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK- OV-3	0.1	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	1.2
Ovarian ca. OVCAR-5	2.3	Small Intestine Pool	0.7
Ovarian ca. IGROV-1	0.7	Stomach Pool	1.0
Ovarian ca. OVCAR-8	0.4	Bone Marrow Pool	2.6
Ovary	0.6	Fetal Heart	1.4
Breast ca. MCF-7	0.0	Heart Pool	1.7
Breast ca. MDA- MB-231	0.7	Lymph Node Pool	1.6
Breast ca. BT 549	0.3	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.7
Breast ca. MDA-N	6.7	Spleen Pool	0.2
Breast Pool	1.4	Thymus Pool	0.4
Trachea	5.3	CNS cancer (glio/astro) U87-MG	10.4
Lung	0.3	CNS cancer (glio/astro) U-118-MG	10.4
Fetal Lung	2.5	CNS cancer (neuro;met) SK-N-AS	39.2
Lung ca. NCI-N417	2.9	CNS cancer (astro) SF- 539	0.8
Lung ca. LX-1	0.1	CNS cancer (astro) SNB-75	1.7
Lung ca. NCI-H146	25.9	CNS cancer (glio) SNB-19	0.4

Lung ca. SHP-77	47.3	CNS cancer (glio) SF-295	12.2
Lung ca. A549	0.0	Brain (Amygdala) Pool	11.1
Lung ca. NCI-H526	0.0	Brain (cerebellum)	100.0
Lung ca. NCI-H23	0.1	Brain (fetal)	20.3
Lung ca. NCI-H460	11.2	Brain (Hippocampus) Pool	18.4
Lung ca. HOP-62	1.0	Cerebral Cortex Pool	20.2
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	10.5
Liver	0.0	Brain (Thalamus) Pool	24.8
Fetal Liver	0.2	Brain (whole)	18.8
Liver ca. HepG2	0.0	Spinal Cord Pool	3.8
Kidney Pool	1.7	Adrenal Gland	1.2
Fetal Kidney	1.0	Pituitary gland Pool	4.7
Renal ca. 786-0	0.1	Salivary Gland	0.1
Renal ca. A498	0.1	Thyroid (female)	0.1
Renal ca. ACHN	0.1	Pancreatic ca. CAPAN2	0.7
Renal ca. UO-31	0.3	Pancreas Pool	0.9

Table IE. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5109, Run 229739339	Tissue Name	Rel. Exp.(%) Ag5109, Run 229739339
Secondary Th1 act	0.0	HUVEC IL-1beta	21.0
Secondary Th2 act	3.5	HUVEC IFN gamma	15.2
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	5.7
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	6.4
Secondary Th2 rest	0.0	HUVEC IL-11	10.7
Secondary Tr1 rest	0.0	Lung Microvascular EC none	100.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	15.6
Primary Th2 act	3.5	Microvascular Dermal EC none	74.7
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	18.3
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium	2.7

		none	
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	3.1
CD45RA CD4 lymphocyte act	4.7	Coronary artery SMC rest	31.9
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	9.2
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	2.8	KU-812 (Basophil) rest	29.9
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	68.3
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	10.6
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	2.5
LAK cells PMA/ionomycin	9.2	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	3.2
Two Way MLR 3 day	0.0	HPAEC none	37.1
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL- 1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	77.9
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	10.9
PBMC PWM	0.0	Lung fibroblast IL-4	30.8
PBMC PHA-L	0.0	Lung fibroblast IL-9	92.7
Ramos (B cell) none	0.0	Lung fibroblast IL-13	28.3
Ramos (B cell) ionomycin	6.2	Lung fibroblast IFN gamma	82.9
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	39.2
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	10.8
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	3.3

EOL-1 dbcAMP PMA/ionomycin	3.8	Dermal fibroblast IFN gamma	12.5
Dendritic cells none	0.0	Dermal fibroblast IL-4	38.2
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	75.8
Dendritic cells anti- CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	7.6	Lung	3.1
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	17.7	Kidney	10.8
HUVEC starved	25.0		

CNS_neurodegeneration_v1.0 Summary: Ag5109 This panel confirms the expression of the CG135316-01 gene at low levels in the brains of an independent group of individuals. See Panel 1.5 for a discussion of this gene in treatment of central nervous system disorders.

- 5 Ag4899 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

- General_screening_panel_v1.5 Summary:** Ag5109 Highest expression of the CG135316-01 gene is detected in brain (cerebellum) (CT=24.9). this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

- 15 High to moderate levels of expression of this gene is also seen in number of cancer cell lines derived from melanoma, pancreatic, brain, lung, breast, ovarian, and prostate cancers. Therefore, therapeutic modulation of the protein encoded by this gene may be useful in the treatment of these cancers.

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

Expression of this gene is higher in adult (CT=32) as compared to fetal skeletal muscle (CT=40). Therefore, expression of this gene may be used to distinguish between fetal and adult skeletal muscle.

5 In addition, this gene is expressed at much higher levels in fetal (CT=30-33) when compared to adult lung and liver respectively (CT=33-38). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung and liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance growth or development of lung and liver in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the
10 protein encoded by this gene could be useful in treatment of lung and liver related diseases.

Ag4899 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel

Panel 4.1D Summary: Ag5109 Highest expression of the CG135316-01 gene is
15 detected in lung microvascular EC cells (CT=33.4). Low levels of expression of this gene is also seen in microvascular dermal EC, and HPAEC. Interestingly, expression of this gene is downregulated in cytokine stimulated endothelial cells. In addition, low levels of expression of this gene is also seen in basophils, lung fibroblasts and dermal fibroblast cells. Therefore, therapeutic modulation of the FGF encoded by this gene or use of this
20 gene as protein therapeutic may be useful in the treatment of autoimmune and inflammatory diseases that involve endothelial cells, such as lupus erythematosus, asthma, emphysema, Crohn's disease, ulcerative colitis, rheumatoid arthritis, osteoarthritis, and psoriasis.

Ag4899 Expression of this gene is low/undetectable (CTs > 35) across all of the
25 samples on this panel.

Example D: Identification of Single Nucleotide Polymorphisms in NOVX nucleic acid sequences

Variant sequences are also included in this application. A variant sequence can
30 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. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a

nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be
5 silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation,
10 intensity of expression, and stability of transcribed message.

SeqCalling assemblies produced by the exon linking process were selected and extended using the following criteria. Genomic clones having regions with 98% identity to all or part of the initial or extended sequence were identified by BLASTN searches using the relevant sequence to query human genomic databases. The genomic clones that
15 resulted were selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies. These sequences were analyzed for putative coding regions as well as for similarity to the known DNA and protein sequences. Programs used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs.

20 Some additional genomic regions may have also been identified because selected SeqCalling assemblies map to those regions. Such SeqCalling sequences may have overlapped with regions defined by homology or exon prediction. They may also be included because the location of the fragment was in the vicinity of genomic regions identified by similarity or exon prediction that had been included in the original predicted
25 sequence. The sequence so identified was manually assembled and then may have been extended using one or more additional sequences taken from CuraGen Corporation's human SeqCalling database. SeqCalling fragments suitable for inclusion were identified by the CuraToolsTM program SeqExtend or by identifying SeqCalling fragments mapping to the appropriate regions of the genomic clones analyzed.

30 The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments or from discrepancies between predicted exon junctions, EST locations and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze

SeqCalling assemblies and genomic clones was reiterated to derive the full length sequence (Alderborn *et al.*, Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. Genome Research. 10 (8) 1249-1265, 2000).

5 Variants are reported individually but any combination of all or a select subset of variants are also included as contemplated NOVX embodiments of the invention.

NOV4a SNP Data

Three polymorphic variants of NOV4a have been identified and are shown in Table 16A.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13379729	192	G	A	60	Gly	Arg
13379730	269	A	G	85	Ala	Ala
13379731	564	A	C	184	Ser	Arg

10 NOV5a SNP Data

One polymorphic variant of NOV5a has been identified and is shown in Table 16B.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13378449	351	T	C	57	Ile	Thr

NOV6a SNP Data

Four polymorphic variants of NOV6a have been identified and are shown in Table 16C.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13379732	139	G	A	41	Gly	Asp
13379733	167	C	T	50	Arg	Arg
13379736	550	T	G	178	Val	Gly
13379737	556	C	T	180	Ala	Val

15

NOV7a SNP Data

Two polymorphic variants of NOV7a have been identified and are shown in Table 16D.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13374892	50	C	G	17	Pro	Ala
13379683	591	G	A	197	Gly	Glu

NOV14a SNP Data

Eight polymorphic variants of NOV14a have been identified and is shown in Table 16E.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13374704	26	C	G	9	Leu	Val
13379710	139	G	A	46	Leu	Leu
13374705	143	A	T	48	Thr	Ser
13379725	334	C	T	111	Pro	Pro
13379724	421	T	G	140	Pro	Pro
13379723	605	C	G	202	His	Asp
13379704	616	C	T	205	Leu	Leu
13379703	688	G	C	229	Val	Val

5 Example E.

Each of the clones listed below is related to a clone or family of clones listed in Example A. The relationship is identifiable as the clone listed below will have the same NOVX number as the clones to which it is related. For example, NOV2B below is related to NOV2a of Example A.

10 NOV2B

The NOV2b clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 17A.

Table 17A. NOV2b Sequence Analysis		
	SEQ ID NO: 121	173 bp
NOV2b, CG109754-01 DNA Sequence	aaatcctccgggctcttaggaaatttcactccgcttctgccagtggtctctttggtcgg cagggtgttcttctcctgcgtctccgttttcttcagcttggccctattgaagctggcgat tccccccacgtctggtttgtctgccattttcttaaaacaatcggtaccatccg	
	SEQ ID NO: 122	43 aa
NOV2b, CG109754-01	MADKPDVGGIASFNRAKLKKTETQEKNLPTKETTGQKRSEIS	

Protein Sequence	
------------------	--

NOV3b

The NOV3b clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 18A.

Table 18A. NOV3b Sequence Analysis		
	SEQ ID NO: 123	330 bp
NOV3b, CG114834-01 DNA Sequence	atgtcagatgcagctgtagacaccagctctgaaatcattgccaggacttaaaggagaag aaggaagtgtgaaagaggcggaaaatggaagagacgcccctgctaacgggaatgctaata gaggaaaatggggagcaggaggctgacaaggaggtagatgaagaaggggaagaaagtggg gaggaagaggaggaggaaaaagaaggtgatggtgaggaagaggatggagatgaagaggaa gctgagtctgctacaggcaagcgggcagctgaagatgatgaggatgatgatgtcgatacc aagaagcagaagaccgacaaggatgactaa	
	SEQ ID NO: 124	109 aa
NOV3b, CG114834-01 Protein Sequence	MSDAAVDTSSSEIIAKDLKEKKEVVKEAENGRDAPANGNANEENGEQADKEVDEEGEESG EEEEEEKEGDGEEEDGDEEEAESATGKRAAEDDEDDVDTKKQKTDKDD	

5

NOV10b

The NOV10b clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 19A.

Table 19A. NOV10b Sequence Analysis		
	SEQ ID NO: 125	646 bp
NOV10b, CG53473-01 DNA Sequence	agcgcgcccgaacgaagccgcccggggcacagcatggcccgcggcgggagggcgctcg gatgttcggcagcctcctgcacttcgccctgctcgtgcggcgctcgtcccgtcagctg ggatctcccgagccccgcagccgagccagcaagatccgagtgactcgcgaggcaagct ctgggccatcgggtcacttcattgggcaagaagagtctggagccttcagcccatccccatt ggggacagctccccacacctccctgaggaccagcgactgcagctgagtcagtcgtctgct cggaatcctcctgctaagaaggctctgggcgtgagcctcagccgccccgcaccccaaat ccagtacaggaggctgctggtacaaatactgcagaaatgacaccaataatggggcagaca caacagcgtggcttagattgtgcccaccaggggaaggtgctgaatgggaccctgttgatg gccccatctggatgtaaatcctgagctcaaatctctgttactccattactgtgatttctg gctgggtcaccagaaatatcgctgatgcagacacagattatgttcctgctgtatttcctg cttcctgttgattggtgaataaaaccttgctctatacatataaaa	
	SEQ ID NO: 126	112 aa
NOV10b, CG53473-01 Protein Sequence	MFGSLLHFALLAAGVVPLSWDLPEPRSRASKIRVHSRGLWAIGHFMGKKSLEPSSPSPL GTAPHTSLRDQRLQLSHDLLGILLKKALGVSLSRPAPQIQYRRLVLVQILQK	

NOV11b

The NOV11b clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 20A.

Table 20A. NOV11b Sequence Analysis		
	SEQ ID NO: 127	147 bp
NOV11b, CG54725-01 DNA Sequence	agaaaatggcacacaaaactagacctggaagaaattgccagcttgataaggccaagctga aggccacagagatgcagaagaacactctgatgaccaagagaccacagagcaggagaagt ggagtgaatttcctgagagcctcgag	
	SEQ ID NO: 128	43 aa
NOV11b, CG54725-01 Protein Sequence	MAHKLDLEEIASLKD KALKATEMQNTLMTKETTEQEKWSEIS	

5

NOV12d

The NOV12d clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 21A.

Table 21A. NOV12d Sequence Analysis		
	SEQ ID NO: 129	348 bp
NOV12d, CG56983-01 DNA Sequence	ctatccctatgggtgtcgggtgtgcaggccgtggcctgctgtggccatagcacttctggctc tgctggctgtgcctggggcgctgggtcgacacctgccccatcaaaccgaggtcctggcg aagacgagtccttgaggagctgagccactattatgcttccctgtgccactacctcaacg tggtcaccagacagttaatttcagagagaaacctaccagacaccattgtgtccaaggaag tatttttcacaagcacaaggaagacctgtgaggacacagaaggaaggttgccatctgc aagccaaggagagaagcctctgaaaaaaccaacctgctggcaccttg	
	SEQ ID NO: 130	104 aa
NOV12d, CG56983-01 Protein Sequence	MVSVCRPWPAVAIALALLVCLGALVDTCPKPEAPGEDESLEELSHYYASLCHYLNVT RQLISERNLPDTIVSKEVFFTSTKERPVRTQKEGCHLQAKERSL	

NOV15b

10

The NOV15b clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 22A.

Table 22A. NOV15b Sequence Analysis		
	SEQ ID NO: 131	515 bp
NOV15b, CG89614-01 DNA	gctgacaccatgctgccgcctgcttctcggcctactggccttctcctccgcgtgcta cttcagaactgcccgagggcggaagagggccatgtccgacctggagctgagacagtg cctccctgcgccccggggcgaaggccgctgcttcgggccagcatttgctgcgcgga cgagctgggctgcttcgtgggcacggctgaggcgctgcgctgccaggaggagaactacct	

Sequence	gccgtcgccctgccagtcgggccagaaggcgtgcgggagcgggggcccgtgcgccgccttcggcggtttgctgcaacgacgagagctgcgtgaccgagtcgagtgccgcgagggctttcacgcccgcgcccgcgcccagcgaccggagcaacgccacgcaactggacaggccggccggggccttgctgctgcggctggtgcagctggccggggcgcccagccctttgagcccgcgcagccgacgcctactgagccccgcgctcgccccaccggc	
	SEQ ID NO: 132	160 aa
NOV15b, CG89614-01 Protein Sequence	MLPACFLGLLAFSSACYFQNCPRGGKRAMSDLELRQCLPCGPGGKGRCFGPSICCADELGCFVGTAEALRCQEENYLPSPCQSGQKACGSGGRCAAFGVCCNDESCVTESECREGFHRRARASDRSNATQLDRPAGALLRLVQLAGAPEPFEPAPDAY	

OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has
5 been done by way of example for purposes of illustration only, and is not intended to be
limiting with respect to the scope of the appended claims, which follow. In particular, it is
contemplated by the inventors that various substitutions, alterations, and modifications
may be made to the invention without departing from the spirit and scope of the invention
as defined by the claims. The choice of nucleic acid starting material, clone of interest, or
10 library type is believed to be a matter of routine for a person of ordinary skill in the art
with knowledge of the embodiments described herein. Other aspects, advantages, and
modifications considered to be within the scope of the following claims. The claims
presented are representative of the inventions disclosed herein. Other, unclaimed
inventions are also contemplated. Applicants reserve the right to pursue such inventions
15 in later claims.

We claim:

1. An isolated polypeptide comprising the mature form of an amino acid sequenced selected from the group consisting of SEQ ID NOs: 122, 124, 126, 128, 130, 132 and SEQ ID NO:2n, wherein n is an integer between 1 and 39.
2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 122, 124, 126, 128, 130, 132 and SEQ ID NO:2n, wherein n is an integer between 1 and 39.
3. An isolated polypeptide comprising an amino acid sequence which is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 122, 124, 126, 128, 130, 132 and SEQ ID NO:2n, wherein n is an integer between 1 and 39.
4. An isolated polypeptide, wherein the polypeptide comprises an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of SEQ ID NOs: 122, 124, 126, 128, 130, 132 and SEQ ID NO:2n, wherein n is an integer between 1 and 39.
5. The polypeptide of claim 1 wherein said polypeptide is naturally occurring.
6. A composition comprising the polypeptide of claim 1 and a carrier.
7. A kit comprising, in one or more containers, the composition of claim 6.
8. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein the therapeutic comprises the polypeptide of claim 1.
9. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:
 - (a) providing said sample;
 - (b) introducing said sample to an antibody that binds immunospecifically to the polypeptide; and
 - (c) determining the presence or amount of antibody bound to said polypeptide, thereby determining the presence or amount of polypeptide in said sample.
10. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the polypeptide of claim 1 in a first mammalian subject, the method comprising:
 - a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and
 - b) comparing the expression of said polypeptide in the sample of step (a) to the expression of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,wherein an alteration in the level of expression of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

11. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:
- (a) introducing said polypeptide to said agent; and
 - (b) determining whether said agent binds to said polypeptide.
12. The method of claim 11 wherein the agent is a cellular receptor or a downstream effector.
13. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:
- (a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;
 - (b) contacting the cell with a composition comprising a candidate substance; and
 - (c) determining whether the substance alters the property or function ascribable to the polypeptide;
- whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition in the absence of the substance, the substance is identified as a potential therapeutic agent.
14. A method for screening for a modulator of activity of or of latency or predisposition to a pathology associated with the polypeptide of claim 1, said method comprising:
- (a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein said test animal recombinantly expresses the polypeptide of claim 1;
 - (b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a); and
 - (c) comparing the activity of said polypeptide in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator activity of or latency or predisposition to, a pathology associated with the polypeptide of claim 1.
15. The method of claim 14, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.
16. A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of claim 1 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.
17. A method of treating or preventing a pathology associated with the polypeptide of claim 1, the method comprising administering the polypeptide of claim 1 to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

18. The method of claim 17, wherein the subject is a human.
19. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NOs: 122, 124, 126, 128, 130, 132 and SEQ ID NO:2n, wherein n is an integer between 1 and 39 or a biologically active fragment thereof.
20. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 121, 123, 125, 127, 129, 131 and SEQ ID NO:2n-1, wherein n is an integer between 1 and 39.
21. The nucleic acid molecule of claim 20, wherein the nucleic acid molecule is naturally occurring.
22. A nucleic acid molecule, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 121, 123, 125, 127, 129, 131 and SEQ ID NO: 2n-1, wherein n is an integer between 1 and 39.
23. An isolated nucleic acid molecule encoding the mature form of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NOs: 122, 124, 126, 128, 130, 132 and SEQ ID NO:2n, wherein n is an integer between 1 and 39.
24. An isolated nucleic acid molecule comprising a nucleic acid selected from the group consisting of SEQ ID NOs: 121, 123, 125, 127, 129, 131 and SEQ ID NOs:2n-1, wherein n is an integer between 1 and 39.
25. The nucleic acid molecule of claim 20, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 121, 123, 125, 127, 129, 131 and SEQ ID NO: 2n-1, wherein n is an integer between 1 and 39, or a complement of said nucleotide sequence.
26. A vector comprising the nucleic acid molecule of claim 20.
27. The vector of claim 26, further comprising a promoter operably linked to said nucleic acid molecule.
28. A cell comprising the vector of claim 26.
29. An antibody that immunospecifically binds to the polypeptide of claim 1.
30. The antibody of claim 29, wherein the antibody is a monoclonal antibody.
31. The antibody of claim 29, wherein the antibody is a humanized antibody.

32. A method for determining the presence or amount of the nucleic acid molecule of claim 20 in a sample, the method comprising:

- (a) providing said sample;
- (b) introducing said sample to a probe that binds to said nucleic acid molecule; and
- (c) determining the presence or amount of said probe bound to said nucleic acid molecule,

thereby determining the presence or amount of the nucleic acid molecule in said sample.

33. The method of claim 32 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

34. The method of claim 33 wherein the cell or tissue type is cancerous.

35. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the nucleic acid molecule of claim 20 in a first mammalian subject, the method comprising:

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

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

36. A method of producing the polypeptide of claim 1, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 121, 123, 125, 127, 129, 131 and SEQ ID NO:2n-1, wherein n is an integer between 1 and 39.

37. The method of claim 36 wherein the cell is a bacterial cell.

38. The method of claim 36 wherein the cell is an insect cell.

39. The method of claim 36 wherein the cell is a yeast cell.

40. The method of claim 36 wherein the cell is a mammalian cell.

41. A method of producing the polypeptide of claim 2, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 121, 123, 125, 127, 129, 131 and SEQ ID NO:2n-1, wherein n is an integer between 1 and 39.

42. The method of claim 41 wherein the cell is a bacterial cell.

43. The method of claim 41 wherein the cell is an insect cell.

44. The method of claim 41 wherein the cell is a yeast cell.
45. The method of claim 41 wherein the cell is a mammalian cell.