

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
23 February 2006 (23.02.2006)

PCT

(10) International Publication Number
WO 2006/020326 A2

- (51) International Patent Classification: Not classified
- (21) International Application Number: PCT/US2005/025762
- (22) International Filing Date: 20 July 2005 (20.07.2005)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/601,306 13 August 2004 (13.08.2004) US
Not furnished 19 July 2005 (19.07.2005) US
- (71) Applicant (for all designated States except US): **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, New Haven, CT 06511 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **BERGHS, Constance** [—/US]; 459 Orange Street, New Haven, CT 06511 (US). **CATTERTON, Elina** [FI/US]; 584 Boston Post Road, Milford, CT 06460 (US). **CHAUDHURI, Amitabha** [IN/US]; 99 Harbor Avenue, Madison, CT 06443 (US). **ELLERMAN, Karen, E.** [US/US]; 87 Montoya Drive, Brandford, CT 06405 (US). **ORT, Tatiana, A.** [—/US]; 85 Viscount Drive, Apt. 61-A, Milford, CT 06460 (US).
- (74) Agent: **YAHWAK, George, M.**; CuraGen Corporation, 555 Long Wharf Drive, New Haven, CT 06511 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (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.*

(54) Title: METHODS OF IDENTIFYING COMPOUNDS THAT MODULATE PROTEIN ACTIVITY

(57) Abstract: Methods of identifying compounds that modulate target polypeptide activity, where a test compound is combined with a target polypeptide and a substrate of the target polypeptide and where a determination is made as to whether the test compound modulates activity of the target polypeptide. The test compounds could be small molecule drugs used for treatment of obesity, diabetes, insulin resistance, and for enhancement of insulin secretion. Target polypeptides and their corresponding nucleic acids as well as their variants are also disclosed.



WO 2006/020326 A2

5

METHODS OF IDENTIFYING COMPOUNDS THAT MODULATE PROTEIN ACTIVITY

RELATED APPLICATIONS

10 This application claims priority to non-provisional US patent application filed July 19, 2005 (Attorney Docket Number Cura 972; Serial Number unknown), which claims priority to provisional US patent application U.S.S.N. 60/601,306, filed August 13, 2004, each of which is incorporated herein by reference in its entirety.

15 FIELD OF THE INVENTION

 The present invention relates to novel polypeptides that are targets of small molecule drugs and that have 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
20 encompass screening, diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions.

BACKGROUND

 Obesity and diabetes are major public health concerns in the developed and developing world. It is estimated that over half of the adult US population is overweight. This includes those with
25 a body mass index (BMI) greater than the upper limit of normal (25) where the BMI is defined as the weight (Kg) / [height (m)]². A common consequence of being overweight is hyperlipidemia and the development of insulin resistance. This is followed by the development of hyperglycemia, a hallmark of Type II diabetes. Left untreated, the hyperglycemia leads to microvascular disease and end organ damage that includes retinopathy, renal disease, cardiac disease, peripheral neuropathy and
30 peripheral vascular compromise. Currently, over 16 million adults in the US are affected by Type II diabetes and the condition has now become rampant among school-age children as a consequence of the epidemic of obesity in that age group.

 Diabetes mellitus is a disorder in which blood levels of glucose (a simple sugar) are abnormally high because the body doesn't release or respond to insulin adequately. Blood sugar
35 (glucose) levels vary throughout the day, rising after a meal and returning to normal within 2 hours. Blood sugar levels are normally between 70 and 110 milligrams per deciliter (mg/dL) of blood in the morning after an overnight fast. They are usually lower than 120 to 140 mg/dL 2 hours after eating foods or drinking liquids containing sugar or other carbohydrates.

 Insulin, a hormone released from the pancreas, is the primary substance responsible for
40 maintaining appropriate blood sugar levels. Insulin allows glucose to be transported into cells so that they can produce energy or store glucose-derived energy until it's needed. The rise in blood sugar levels after eating or drinking stimulates the pancreas to produce insulin, preventing a greater rise in blood sugar levels and causing them to fall gradually. Because muscles use glucose for energy, blood sugar levels can also fall during physical activity.

5 Diabetes results when the body doesn't produce enough insulin to maintain normal blood sugar levels or when cells don't respond appropriately to insulin. In type II diabetes mellitus, the pancreas continues to manufacture insulin, sometimes even at higher than normal levels. However, the body develops resistance to its effects, resulting in a relative insulin deficiency.

10 The main goal of diabetes treatment is to keep blood sugar levels within the normal range as much as possible. Completely normal levels are difficult to maintain, but the more closely they can be kept within the normal range, the less likely that temporary or long-term complications will develop.

Therefore, a therapeutic that decreases insulin resistance and/or enhances insulin secretion would be beneficial in treatment of obesity and/or diabetes. Additionally, such a therapeutic would be beneficial in treatment of insulin resistance, a condition that often leads to the development of
15 diabetes.

In order to treat diseases, pathologies and other abnormal states or conditions in which a mammalian organism has been diagnosed as being, or as being at risk for becoming, other than in a normal state or condition, it is important to identify new therapeutic agents.

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

Signaling proteins may be classified as endocrine effectors, paracrine effectors or autocrine effectors. Endocrine effectors are signaling molecules secreted by a given organ into the circulatory system, which are then transported to a distant target organ or tissue. The target cells include the receptors for the endocrine effector, and when the endocrine effector binds, a signaling cascade is induced. Paracrine effectors involve secreting cells and receptor cells in close proximity to each other, for example two different classes of cells in the same tissue or organ. One class of cells secretes the paracrine effector, which then reaches the second class of cells, for example by diffusion through the extracellular fluid. The second class of cells contains the receptors for the paracrine effector; binding of the effector results in induction of the signaling cascade that elicits the
30 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.
35

40 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
45 proteins. In certain classes of pathologies the dysregulation is manifested as diminished or

5 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 an increased or up-regulated levels of the protein effector of interest.

Small molecule targets have been implicated in various disease states or pathologies. These targets may be proteins, and particularly enzymatic proteins, which are acted upon by small molecule drugs for the purpose of altering target function and achieving a desired result. Cellular, animal and clinical studies can be performed to elucidate the genetic contribution to the etiology and pathogenesis of conditions in which small molecule targets are implicated in a variety of physiologic, pharmacologic or native states. These studies utilize the core technologies at CuraGen Corporation to look at differential gene expression, protein-protein interactions, large-scale sequencing of expressed genes and the association of genetic variations such as, but not limited to, single nucleotide polymorphisms (SNPs) or splice variants in and between biological samples from experimental and control groups. The goal of such studies is to identify potential avenues for therapeutic intervention in order to prevent, treat the consequences or cure the conditions.

In order to treat diseases, pathologies and other abnormal states or conditions in which a mammalian organism has been diagnosed as being, or as being at risk for becoming, other than in a normal state or condition, it is important to identify new therapeutic agents. Such a procedure includes at least the steps of identifying a target component within an affected tissue or organ, and identifying a candidate therapeutic agent that modulates the functional attributes of the target. The target component may be any biological macromolecule implicated in the disease or pathology. Commonly the target is a polypeptide or protein with specific functional attributes. Other classes of macromolecule may be a nucleic acid, a polysaccharide, a lipid such as a complex lipid or a glycolipid; in addition a target may be a sub-cellular structure or extra-cellular structure that is comprised of more than one of these classes of macromolecule. Once such a target has been identified, it may be employed in a screening assay in order to identify favorable candidate therapeutic agents from among a large population of substances or compounds.

In many cases the objective of such screening assays is to identify small molecule candidates; this is commonly approached by the use of combinatorial methodologies to develop the population of substances to be tested. The implementation of high throughput screening methodologies is advantageous when working with large, combinatorial libraries of compounds.

5 DEFINITIONS

As used herein, the terms and phrases "nucleic acid-molecule", "probe", "isolated", "oligonucleotide", "complementary", "fragment", "homologous nucleic acid sequence", "homologous amino acid sequence", "polypeptide having a biologically active portion of NOVX", "gene", "recombinant gene", "hybridizes under stringent conditions", "stringent hybridization conditions",
10 "coding region", "noncoding region", "NOVX", "PNAs", "peptide nucleic acids", "isolated", "purified", "derivative", analog, "homolog", "substantially free of chemical precursors or other chemicals", "sequence identity", "chimeric protein", "fusion protein", "operatively linked", "antibody", and "monoclonal antibody" are as defined in United States Patent 6,600,019 in columns 68 to 81, the definitions of which are incorporated *in toto* herein.

15 SUMMARY OF THE INVENTION

The invention includes nucleic acid sequences and the novel polypeptides they encode. The novel nucleic acids and polypeptides are referred to herein as NOVX, or NOV1, NOV2, NOV3, *etc.*, nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX"
20 nucleic acid, which represents the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 16, or polypeptide sequences, which represents the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 16.

In one aspect, the invention provides an isolated polypeptide comprising a mature form of a NOVX amino acid. One example is a variant of a mature form of a NOVX amino acid sequence,
25 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. The amino acid can be, for example, a NOVX amino acid sequence or a variant of a NOVX amino acid sequence, wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed.
30 The invention also includes fragments of any of these. In another aspect, the invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof.

Also included in the invention is a NOVX polypeptide that is a naturally occurring allelic variant of a NOVX sequence. In one embodiment, the allelic variant includes an amino acid
35 sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a NOVX nucleic acid sequence. In another embodiment, the NOVX polypeptide is a variant polypeptide described therein, wherein any amino acid specified in the chosen sequence is changed to provide a conservative substitution. In one embodiment, the invention discloses a method for determining the presence or amount of the NOVX polypeptide in a sample. The method involves the
40 steps of: providing a sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the NOVX polypeptide, thereby determining the presence or amount of the NOVX polypeptide in the sample. In another embodiment, the invention provides a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide in a mammalian

5 subject. This method involves the steps of: measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in the sample of the first step to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease, wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

10 In a further embodiment, the invention includes a method of identifying an agent that modulates a NOVX polypeptide. This method can involve the steps of: introducing the polypeptide to the agent; and determining whether the agent binds to the polypeptide. In various embodiments, the agent is a cellular receptor or a downstream effector.

15 In another aspect, the invention provides 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 NOVX polypeptide. The method involves the steps of: providing a cell expressing the NOVX polypeptide and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and
20 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 aspect, the invention describes a method for screening for a modulator of activity or of latency or predisposition to a pathology associated with the NOVX polypeptide. This method involves the following steps: administering a test compound to a test animal at increased risk
25 for a pathology associated with the NOVX polypeptide, wherein the test animal recombinantly expresses the NOVX polypeptide. This method involves the steps of measuring the activity of the NOVX polypeptide in the test animal after administering the compound of step; and comparing the activity of the protein in the test animal with the activity of the NOVX polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the NOVX polypeptide in the
30 test animal relative to the control animal indicates that the test compound is a modulator of latency of, or predisposition to, a pathology associated with the NOVX polypeptide. In one embodiment, the test animal is a recombinant test animal that expresses a test protein transgene or expresses the transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein the promoter is not the native gene promoter of the transgene. In another aspect, the invention includes a method for modulating the activity of the NOVX polypeptide, the method comprising introducing a cell sample expressing the NOVX polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

35 In order to treat diseases, pathologies and other abnormal states or conditions in which a mammalian organism has been diagnosed as being, or as being at risk for becoming, other than in a normal state or condition, it is important to identify new therapeutic agents. Such a procedure includes at least the steps of identifying a target component within an affected tissue or organ, and identifying a candidate therapeutic agent that modulates the functional attributes of the target. The target component may be any biological macromolecule implicated in the disease or pathology.
40 Commonly the target is a polypeptide or protein with specific functional attributes. Other classes of

5 macromolecule may be a nucleic acid, a polysaccharide, a lipid such as a complex lipid or a glycolipid; in addition a target may be a sub-cellular structure or extra-cellular structure that is comprised of more than one of these classes of macromolecule. Once such a target has been identified, it may be employed in a screening assay in order to identify favorable candidate therapeutic agents from among a large population of substances or compounds.

10 In many cases the purpose of such screening assays is to identify small molecule candidates; this is commonly approached by the use of combinatorial methodologies to develop the population of substances to be tested. The implementation of high throughput screening methodologies is advantageous when working with large, combinatorial libraries of compounds.

15 It is a purpose of this invention to describe cell lines that recombinantly or endogenously express the target biopolymer or an isolated target biopolymer that is intended to serve as the macromolecular component in a screening assay for identifying candidate pharmaceutical agents.

It is another purpose of the present invention to describe screening assays that positively identify candidate pharmaceutical agents from among a combinatorial library of low molecular weight substances or compounds.

20 It is still a further aspect of this invention to employ the candidate pharmaceutical agents in any of a variety of in vitro, ex vivo and in vivo assays in order to identify pharmaceutical agents with advantageous therapeutic applications in the treatment of a disease, pathology, or abnormal state or condition in a mammal.

25 In another aspect, the present invention describes a method of identifying a test compound as a candidate therapeutic agent, for treating a disease, pathology, or an abnormal state or condition using a target polypeptide (NOVX) having a specific association with the disease. This method includes:

- (a) combining a test compound with a target polypeptide and a substrate of the target polypeptide; and
- 30 (b) determining whether the test compound modulates the activity of the target polypeptide.

In one embodiment of this method, the chemical compound is a member of a combinatorial library of compounds; the combining in step (a) is conducted on one or more replicate samples of the biopolymer; and the replicate sample is contacted with at least one member of the combinatorial library. In additional embodiments of this method, the biopolymer is included within a cell and is functionally expressed therein. In still a further embodiment, the binding of the compound modulates the function of the biopolymer, and it is the modulation that provides the identification that the compound is a potential therapeutic agent. In yet further embodiments of this method, the target biopolymer is a polypeptide.

40 As used herein, a "substrate" is any compound capable of binding to or interacting with a target polypeptide, including but not limited to a peptide, a polypeptide, a nucleic acid, a carbohydrate moiety, a lipid, a small molecule (*e.g.*, cyclic AMP, ATP), an agonist, an antagonist, and an inhibitor.

In another aspect of the invention, a method for identifying a pharmaceutical agent for treating a disease, pathology, or an abnormal state or condition is described. The method includes the steps of:

- (1) identifying a candidate therapeutic agent for treating said disease, pathology, or abnormal state or condition by the method described in the preceding paragraphs;
- (2) contacting a biological sample associated with the disease, pathology, or abnormal state or condition with the candidate therapeutic agent;
- (3) determining whether the candidate induces an effect on the biological sample associated with a therapeutic response therein; and
- (4) identifying a candidate exerting such an effect as a pharmaceutical agent.

In significant embodiments of the method, the biological sample includes a cell, a tissue or organ, or is a nonhuman mammal.

Several cellular, animal and clinical studies were performed to elucidate the genetic contribution to the etiology and pathogenesis of these conditions in a variety of physiologic, pharmacologic or native states. These studies utilized the core technologies at CuraGen Corporation to look at differential gene expression, protein-protein interactions, large-scale sequencing of expressed genes and the association of genetic variations such as, but not limited to, single nucleotide polymorphisms (SNPs) or splice variants in and between biological samples from experimental and control groups. The goal of such studies is to identify various therapeutic interventions in order to prevent, treat the consequences or cure the conditions of obesity and/or diabetes.

The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with various diseases or pathologies. The proteins, and related proteins that are similar to them, are encoded by a cDNA and/or by genomic DNA. The proteins, polypeptides and their cognate nucleic acids were identified by the inventors in certain cases. Additionally, the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify therapeutic antibodies and/or therapeutic small molecules which modulate activity of the disclosed NOVX polypeptides.

The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. In a preferred embodiment, the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant. In another embodiment, the nucleic acid encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant. In another embodiment, the nucleic acid molecule differs by a single nucleotide from a NOVX nucleic acid sequence. In one embodiment, the NOVX nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 16, or a complement of the nucleotide sequence. In another aspect, the invention provides a vector or a cell expressing a NOVX nucleotide sequence.

In one embodiment, the invention discloses a method for modulating the activity of a NOVX polypeptide. The method includes the steps of: introducing a cell sample expressing the NOVX 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 includes an isolated NOVX nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising a NOVX amino acid sequence or a variant of a mature form of the NOVX amino acid sequence, wherein any amino

5" 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. In another embodiment, the invention includes an amino acid sequence that is a variant of the NOVX amino acid sequence, 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.

10 In one embodiment, the invention discloses a NOVX nucleic acid fragment encoding at least a portion of a NOVX polypeptide 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. In another embodiment, the invention includes the complement of any of the NOVX nucleic acid molecules or a naturally occurring allelic nucleic acid variant. In another embodiment, the invention discloses a NOVX nucleic acid molecule that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant. In another embodiment, the invention discloses a NOVX nucleic acid, wherein the nucleic acid molecule differs by a single nucleotide from a NOVX nucleic acid sequence.

15 In another aspect, the invention includes a NOVX nucleic acid, wherein one or more nucleotides in the NOVX nucleotide sequence is changed to a different nucleotide provided that no more than 15% of the nucleotides are so changed. In one embodiment, the invention discloses a nucleic acid fragment of the NOVX nucleotide sequence and a nucleic acid fragment wherein one or more nucleotides in the NOVX 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 are so changed. In another embodiment, the invention includes a nucleic acid molecule wherein the nucleic acid molecule hybridizes under stringent conditions to a NOVX nucleotide sequence or a complement of the NOVX nucleotide sequence. In one embodiment, the invention includes a nucleic acid molecule, wherein the sequence is changed such that no more than 15% of the nucleotides in the coding sequence differ from the NOVX nucleotide sequence or a fragment thereof.

20 In a further aspect, the invention includes a method for determining the presence or amount of the NOVX nucleic acid in a sample. The method involves the steps of: 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 NOVX nucleic acid molecule, thereby determining the presence or amount of the NOVX nucleic acid molecule in the sample. In one embodiment, the presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

25 In another aspect, the invention discloses a method for determining the presence of or predisposition to a disease associated with altered levels of the NOVX nucleic acid molecule of in a first mammalian subject. The method involves the steps of: measuring the amount of NOVX 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 NOVX 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.

30 35 40 45

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. 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 become readily apparent from the following detailed description and claims.

BRIEF DESCRIPTION OF THE FIGURES

Figure D1 illustrates the novel protein-protein interactions of PPAR gamma (PPARG) detected by the PathCalling® Technology. The large circular shapes represent polypeptides. The lines connecting the circular shapes indicate interactions between the polypeptides as detected by PathCalling®. PPARG interacted with PRKCN, NROB1, NROB2, RORC, and NRIP1.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences, their encoded polypeptides, antibodies, and other related compounds. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." 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 (amino acid)	Homology
NOV1a	CG181171-01	1	2	Aquaporin 7 (Aquaporin-7 like) (Aquaporin adipose) (AQPap) - Homo sapiens
NOV1b	CG181171-02	3	4	Aquaporin 7 (Aquaporin-7 like) (Aquaporin adipose) (AQPap) - Homo sapiens
NOV2a	CG127635-01	5	6	Bile acid receptor (Farnesoid X-activated receptor) (Farnesol receptor HRR-1) (Retinoid X receptor-interacting protein 14) (RXR-interacting protein 14) - Homo sapiens
NOV3a	CG197218-01	9	10	P2Y purinoceptor 1 (ATP receptor) (P2Y1) (Purinergic receptor) - Homo sapiens
NOV4a	CG197755-01	13	14	Protein kinase C, nu type (EC 2.7.1.-) (nPKC-nu) (Protein kinase EPK2) - Homo sapiens
NOV4b	CG197755-02	15	16	Protein kinase C, nu type (EC 2.7.1.-) (nPKC-nu) (Protein kinase EPK2) - Homo sapiens
NOV4c	CG197755-03	17	18	Protein kinase C, nu type (EC 2.7.1.-) (nPKC-nu) (Protein kinase EPK2) - Homo sapiens
NOV4d	CG197755-04	19	20	Protein kinase C, nu type (EC 2.7.1.-) (nPKC-nu) (Protein kinase EPK2) - Homo sapiens

NOV5a	CG184446-01	21	22	Delta-6 fatty acid desaturase (Fatty acid desaturase 2) (Hypothetical protein FLJ90458) - Homo sapiens
NOV5b	CG184446-02	23	24	Delta-6 fatty acid desaturase (Fatty acid desaturase 2) (Hypothetical protein FLJ90458) - Homo sapiens
NOV6a	CG56918-02	25	26	Monocarboxylate transporter 13 - Homo sapiens
NOV6b	CG56918-01	27	28	Monocarboxylate transporter 13 - Homo sapiens
NOV6c	CG56918-03	29	30	Monocarboxylate transporter 13 - Homo sapiens
NOV6d	CG56918-04	31	32	Monocarboxylate transporter 13 - Homo sapiens

5

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

Pathologies, diseases, disorders and condition and the like that are associated with NOVX sequences include, but are not limited to, *e.g.*, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, metabolic disturbances associated with obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic disorders, 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, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers, as well as conditions such as transplantation and fertility.

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

5 that are characteristic of, other members of such protein families. Details of the sequence relatedness and domain analysis for each NOVX are presented in Examples for identification of human sequence in individual sections for each NOVX polypeptide.

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
10 according to the invention may be used as targets for the identification of small molecules that modulate or inhibit diseases associated with the protein families listed in Table 1.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are Examples showing expression profiles in individual sections for each NOVX polypeptide. Accordingly, the NOVX nucleic acids, polypeptides,
15 antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the detection of a variety of diseases with differential expression in normal vs. diseased tissues, *e.g.*, detection of a variety of cancers. SNP analysis for each NOVX, if applicable, is presented in SNP Examples in individual sections for each NOVX polypeptide.

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

NOVX clones

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

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

The NOVX nucleic acids and proteins of the invention are useful in potential diagnostic and
35 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
40 (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) a biological defense weapon.

In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1
45 and 16; (b) a variant of a mature form of the amino acid sequence selected from the group consisting

5 of SEQ ID NO: 2n, wherein n is an integer between 1 and 16, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 16; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer
10 between 1 and 16 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
15 selected from the group consisting of: (a) a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 16; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 16 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
20 mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 16; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 16, 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
25 nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 16 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and (f) the complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 16; (b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1,
35 wherein n is an integer between 1 and 16 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 16; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:
40 2n-1, wherein n is an integer between 1 and 16 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX
45 polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid

5 "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.

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

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

35 A nucleic acid molecule of the invention can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:2*n*-1 as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques well known in the art.

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

In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NO:2*n*-1. Oligonucleotides may be chemically synthesized and may also be used as probes.

10 In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 16, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of a NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence of SEQ ID NO:2*n*-1 is one that is sufficiently complementary to the nucleotide sequence of SEQ ID NO:2*n*-1 that it can hydrogen bond with few or no mismatches to the nucleotide sequence shown in SEQ ID
15 NO:2*n*-1 thereby forming a stable duplex.

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

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

In the present invention, homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions) in SEQ ID NO:2*n*-1 as well as a polypeptide possessing NOVX biological activity.

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

The nucleotide sequences determined from the cloning of the human NOVX genes allows for
45 the generation of probes and primers designed for use in identifying and/or cloning NOVX

5 homologues in other cell types, *e.g.*, from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1; or an anti-sense strand nucleotide sequence of SEQ ID
10 NO:2*n*-1; or of a naturally occurring mutant of SEQ ID NO:2*n*-1.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe has a detectable label attached, *e.g.*, the label can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for
15 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 nucleic acid fragment encoding a "biologically-active portion of NOVX" according to the present invention can be prepared by isolating a portion of SEQ ID NO:2*n*-1 that encodes a
20 polypeptide having a NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide
25 sequences of SEQ ID NO:2*n*-1 due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 16. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence of SEQ ID NO:2*n*.

In addition to the human NOVX nucleotide sequences of SEQ ID NO:2*n*-1, it will be
30 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. 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
35 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 a human SEQ ID NO:2*n*-1 are intended to be within the
40 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.

5 " Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region.

10 The conditions for stringent conditions, as used in the present invention, 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, 15 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 a sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 16, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA 20 molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In another embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1 or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Reinhardt's solution, 25 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.

In still another embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NO:2*n*-1 or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are 35 well known in the art (*e.g.*, as employed for cross-species hybridizations). See, *e.g.*, *PNAS USA* 78: 6789 (1981).

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into 40 the nucleotide sequences of SEQ ID NO:2*n*-1 thereby leading to changes in the amino acid sequences of the encoded NOVX protein, without altering the functional ability of that NOVX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:2*n*. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering 45 their biological activity, whereas an "essential" amino acid residue is required for such biological

5 activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are not particularly 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
10 differ in amino acid sequence from SEQ ID NO:2*n*-1 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 of SEQ ID NO:2*n*. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NO:2*n*; more preferably at least about 70% homologous to
15 SEQ ID NO:2*n*; still more preferably at least about 80% homologous to SEQ ID NO:2*n*; even more preferably at least about 90% homologous to SEQ ID NO:2*n*; and most preferably at least about 95% homologous to SEQ ID NO:2*n*.

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

Mutations can be introduced any one of SEQ ID NO:2*n*-1 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more non-essential amino acid residues. A "conservative amino
25 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
30 (*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 non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a
35 NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of a nucleic acid of SEQ ID NO:2*n*-1 the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

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

5 " STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

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

Interfering RNA

15 In another aspect of the invention, NOVX gene expression can be attenuated by RNA interference. One approach well-known in the art is short interfering RNA (siRNA) mediated gene silencing where expression products of a NOVX gene are targeted by specific double stranded NOVX derived siRNA nucleotide sequences that are complementary to at least a 19-25 nt long segment of the NOVX gene transcript, including the 5' untranslated (UT) region, the ORF, or the 3' UT region [see PCT applications WO00/44895, WO99/32619, WO01/75164, WO01/92513, WO 01/29058, WO01/89304, WO02/16620, and WO02/29858]. Targeted genes can be a NOVX gene, or an upstream or downstream modulator of the NOVX gene. Nonlimiting examples of upstream or downstream modulators of a NOVX gene include, e.g., a transcription factor that binds the NOVX gene promoter, a kinase or phosphatase that interacts with a NOVX polypeptide, and polypeptides involved in a NOVX regulatory pathway.

25 According to the methods of the present invention, NOVX gene expression is silenced using short interfering RNA. A NOVX polynucleotide according to the invention includes a siRNA polynucleotide. Such a NOVX siRNA can be obtained using a NOVX polynucleotide sequence, for example, by processing the NOVX ribopolynucleotide sequence in a cell-free system, such as but not limited to a Drosophila extract, or by transcription of recombinant double stranded NOVX RNA or by chemical synthesis of nucleotide sequences homologous to a NOVX sequence [see Genes & Dev. 13:3191 (1999)]. When synthesized, a typical 0.2 micromolar-scale RNA synthesis provides about 1 milligram of siRNA, which is sufficient for 1000 transfection experiments using a 24-well tissue culture plate format.

35 The most efficient silencing is generally observed with siRNA duplexes composed of a 21-nt sense strand and a 21-nt antisense strand, paired in a manner to have a 2-nt 3' overhang. The sequence of the 2-nt 3' overhang makes an additional small contribution to the specificity of siRNA target recognition. The contribution to specificity is localized to the unpaired nucleotide adjacent to the first paired bases. In one embodiment, the nucleotides in the 3' overhang are ribonucleotides. In an alternative embodiment, the nucleotides in the 3' overhang are deoxyribonucleotides. Using 2'-deoxyribonucleotides in the 3' overhangs is as efficient as using ribonucleotides, but deoxyribonucleotides are often cheaper to synthesize and are most likely more nuclease resistant.

40 A contemplated recombinant expression vector of the invention comprises a NOVX DNA molecule cloned into an expression vector comprising operatively-linked regulatory sequences flanking the

5 NOVX sequence in a manner that allows for expression (by transcription of the DNA molecule) of both strands. An RNA molecule that is antisense to NOVX mRNA is transcribed by a first promoter (e.g., a promoter sequence 3' of the cloned DNA) and an RNA molecule that is the sense strand for the NOVX mRNA is transcribed by a second promoter (e.g., a promoter sequence 5' of the cloned DNA). The sense and antisense strands may hybridize in vivo to generate siRNA constructs for
10 silencing of the NOVX gene. Alternatively, two constructs can be utilized to create the sense and anti-sense strands of a siRNA construct. Finally, cloned DNA can encode a construct having secondary structure, wherein a single transcript has both the sense and complementary antisense sequences from the target gene or genes. In an example of this embodiment, a hairpin RNAi product is homologous to all or a portion of the target gene. In another example, a hairpin RNAi product is a
15 siRNA. The regulatory sequences flanking the NOVX sequence may be identical or may be different, such that their expression may be modulated independently, or in a temporal or spatial manner. In a specific embodiment, siRNAs are transcribed intracellularly by cloning the NOVX gene templates into a vector containing, e.g., a RNA pol III transcription unit from the smaller nuclear RNA (snRNA) U6 or the human RNase P RNA H1. One example of a vector system is the GeneSuppressor™ RNA
20 Interference kit (commercially available from Imgenex). The U6 and H1 promoters are members of the type III class of Pol III promoters. The +1 nucleotide of the U6-like promoters is always guanosine, whereas the +1 for H1 promoters is adenosine. The termination signal for these promoters is defined by five consecutive thymidines. The transcript is typically cleaved after the second uridine. Cleavage at this position generates a 3' UU overhang in the expressed siRNA, which
25 is similar to the 3' overhangs of synthetic siRNAs. Any sequence less than 400 nucleotides in length can be transcribed by these promoter, therefore they are ideally suited for the expression of around 21-nucleotide siRNAs in, e.g., an approximately 50-nucleotide RNA stem-loop transcript.

A siRNA vector appears to have an advantage over synthetic siRNAs where long term knock-down of expression is desired. Cells transfected with a siRNA expression vector would
30 experience steady, long-term mRNA inhibition. In contrast, cells transfected with exogenous synthetic siRNAs typically recover from mRNA suppression within seven days or ten rounds of cell division. The long-term gene silencing ability of siRNA expression vectors may provide for applications in gene therapy.

In general, siRNAs are chopped from longer dsRNA by an ATP-dependent ribonuclease
35 called DICER. DICER is a member of the RNase III family of double-stranded RNA-specific endonucleases. The siRNAs assemble with cellular proteins into an endonuclease complex. *In vitro* studies in *Drosophila* suggest that the siRNAs/protein complex (siRNP) is then transferred to a second enzyme complex, called an RNA-induced silencing complex (RISC), which contains an endoribonuclease that is distinct from DICER. RISC uses the sequence encoded by the antisense
40 siRNA strand to find and destroy mRNAs of complementary sequence. The siRNA thus acts as a guide, restricting the ribonuclease to cleave only mRNAs complementary to one of the two siRNA strands.

A NOVX mRNA region to be targeted by siRNA is generally selected from a desired NOVX sequence beginning 50 to 100 nt downstream of the start codon. Alternatively, 5' or 3' UTRs and
45 regions nearby the start codon can be used but are generally avoided, as these may be richer in

5 regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. An initial BLAST homology search for the selected siRNA sequence is done against an available nucleotide sequence library to ensure that only one gene is targeted. Specificity of target recognition by siRNA duplexes indicate that a single point mutation located in the paired region of an siRNA duplex is sufficient to abolish target mRNA degradation [see EMBO J. 20(23):6877 (2001)]. Hence, consideration should be taken to accommodate SNPs, polymorphisms, allelic variants or species-specific variations when targeting a desired gene.

10 In one embodiment, a complete NOVX siRNA experiment includes the proper negative control. A negative control siRNA generally has the same nucleotide composition as the NOVX siRNA but lack significant sequence homology to the genome. Typically, one would scramble the nucleotide sequence of the NOVX siRNA and do a homology search to make sure it lacks homology to any other gene.

Two independent NOVX siRNA duplexes can be used to knock-down a target NOVX gene. This helps to control for specificity of the silencing effect. In addition, expression of two independent genes can be simultaneously knocked down by using equal concentrations of different NOVX siRNA duplexes, *e.g.*, a NOVX siRNA and an siRNA for a regulator of a NOVX gene or polypeptide. Availability of siRNA-associating proteins is believed to be more limiting than target mRNA accessibility.

20 A targeted NOVX region is typically a sequence of two adenines (AA) and two thymidines (TT) divided by a spacer region of nineteen (N19) residues (*e.g.*, AA(N19)TT). A desirable spacer region has a G/C-content of approximately 30% to 70%, and more preferably of about 50%. If the sequence AA(N19)TT is not present in the target sequence, an alternative target region would be AA(N21). The sequence of the NOVX sense siRNA corresponds to (N19)TT or N21, respectively. In the latter case, conversion of the 3' end of the sense siRNA to TT can be performed if such a sequence does not naturally occur in the NOVX polynucleotide. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3' overhangs. Symmetric 3' overhangs may help to ensure that the siRNPs are formed with approximately equal ratios of sense and antisense target RNA-cleaving siRNPs [see Genes & Dev. 15:188 (2001)]. The modification of the overhang of the sense sequence of the siRNA duplex is not expected to affect targeted mRNA recognition, as the antisense siRNA strand guides target recognition.

25 Alternatively, if the NOVX target mRNA does not contain a suitable AA(N21) sequence, one may search for the sequence NA(N21). Further, the sequence of the sense strand and antisense strand may still be synthesized as 5' (N19)TT, as it is believed that the sequence of the 3'-most nucleotide of the antisense siRNA does not contribute to specificity. Unlike antisense or ribozyme technology, the secondary structure of the target mRNA does not appear to have a strong effect on silencing [see J. Cell Science 114:4557 (2001)].

40 Transfection of NOVX siRNA duplexes can be achieved using standard nucleic acid transfection methods, for example, OLIGOFECTAMINE Reagent (available from Invitrogen). An assay for NOVX gene silencing is generally performed approximately 2 days after transfection. No

5 NOVX gene silencing has been observed in the absence of transfection reagent, allowing for a comparative analysis of the wild-type and silenced NOVX phenotypes. In a specific embodiment, for one well of a 24-well plate, approximately 0.84 μ g of the siRNA duplex is generally sufficient. Cells are typically seeded the previous day, and are transfected at about 50% confluence. The choice of cell culture media and conditions are routine to those of skill in the art, and will vary with the choice of
10 cell type. The efficiency of transfection may depend on the cell type, but also on the passage number and the confluency of the cells. The time and the manner of formation of siRNA-liposome complexes (e.g., inversion versus vortexing) are also critical. Low transfection efficiencies are the most frequent cause of unsuccessful NOVX silencing. The efficiency of transfection needs to be carefully examined for each new cell line to be used. Preferred cell are derived from a mammal,
15 more preferably from a rodent such as a rat or mouse, and most preferably from a human. Where used for therapeutic treatment, the cells are preferentially autologous, although non-autologous cell sources are also contemplated as within the scope of the present invention.

For a control experiment, transfection of 0.84 μ g single-stranded sense NOVX siRNA will have no effect on NOVX silencing, and 0.84 μ g antisense siRNA has a weak silencing effect when
20 compared to 0.84 μ g of duplex siRNAs. Control experiments again allow for a comparative analysis of the wild-type and silenced NOVX phenotypes. To control for transfection efficiency, targeting of common proteins is typically performed, for example targeting of lamin A/C or transfection of a CMV-driven EGFP-expression plasmid (e.g., available from Clontech). In the above example, a determination of the fraction of lamin A/C knockdown in cells is determined the next day by such
25 techniques as immunofluorescence, Western blot, Northern blot or other similar assays for protein expression or gene expression. Lamin A/C monoclonal antibodies may be obtained from Santa Cruz Biotechnology.

Depending on the abundance and the half life (or turnover) of the targeted NOVX polynucleotide in a cell, a knock-down phenotype may become apparent after 1 to 3 days, or even
30 later. In cases where no NOVX knock-down phenotype is observed, depletion of the NOVX polynucleotide may be observed by immunofluorescence or Western blotting. If the NOVX polynucleotide is still abundant after 3 days, cells need to be split and transferred to a fresh 24-well plate for re-transfection. If no knock-down of the targeted protein is observed, it may be desirable to analyze whether the target mRNA (NOVX or a NOVX upstream or downstream gene) was effectively
35 destroyed by the transfected siRNA duplex. Two days after transfection, total RNA is prepared, reverse transcribed using a target-specific primer, and PCR-amplified with a primer pair covering at least one exon-exon junction in order to control for amplification of pre-mRNAs. RT/PCR of a non-targeted mRNA is also needed as control. Effective depletion of the mRNA yet undetectable reduction of target protein may indicate that a large reservoir of stable NOVX protein may exist in the
40 cell. Multiple transfections in sufficiently long intervals may be necessary until the target protein is finally depleted to a point where a phenotype may become apparent. If multiple transfection steps are required, cells are split 2 to 3 days after transfection. The cells may be transfected immediately after splitting.

An inventive therapeutic method of the invention contemplates administering a NOVX siRNA
45 construct as therapy to compensate for increased or aberrant NOVX expression or activity. The

5 NOVX ribopolynucleotide is obtained and processed into siRNA fragments, or a NOVX siRNA is synthesized, as described above. The NOVX siRNA is administered to cells or tissues using known nucleic acid transfection techniques. A NOVX siRNA specific for a NOVX gene will decrease or knockdown NOVX transcription products, which will lead to reduced NOVX polypeptide production, resulting in reduced NOVX polypeptide activity in the cells or tissues.

10 The present invention also encompasses a method of treating a disease or condition associated with the presence of a NOVX protein in an individual comprising administering to the individual an RNAi construct that targets the mRNA of the protein (the mRNA that encodes the protein) for degradation. A specific RNAi construct includes a siRNA or a double stranded gene transcript that is processed into siRNAs. Upon treatment, the target protein is not produced or is not
15 produced to the extent it would be in the absence of the treatment.

Where the NOVX gene function is not correlated with a known phenotype, a control sample of cells or tissues from healthy individuals provides a reference standard for determining NOVX expression levels. Expression levels are detected using the assays described, *e.g.*, RT-PCR, Northern blotting, Western blotting, or ELISA. A subject sample of cells or tissues is taken from a
20 mammal, preferably a human subject, suffering from a disease state. The NOVX ribopolynucleotide is used to produce siRNA constructs that are specific for the NOVX gene product. These cells or tissues are treated by administering NOVX siRNA's to the cells or tissues by methods described for the transfection of nucleic acids into a cell or tissue, and a change in NOVX polypeptide or polynucleotide expression is observed in the subject sample relative to the control sample, using the
25 assays described. This NOVX gene knockdown approach provides a rapid method for determination of a NOVX minus (NOVX⁻) phenotype in the treated subject sample. The NOVX⁻ phenotype observed in the treated subject sample thus serves as a marker for monitoring the course of a disease state during treatment.

In specific embodiments, a NOVX siRNA is used in therapy. Methods for the generation and
30 use of a NOVX siRNA are known to those skilled in the art.

Production of RNAs

Sense RNA (ssRNA) and antisense RNA (asRNA) of NOVX are produced using known methods such as transcription in RNA expression vectors. In the initial experiments, the sense and antisense RNA are about 500 bases in length each. The produced ssRNA and asRNA (0.5 μ M) in
35 10 mM Tris-HCl (pH 7.5) with 20 mM NaCl were heated to 95° C for 1 min then cooled and annealed at room temperature for 12 to 16 h. The RNAs are precipitated and resuspended in lysis buffer (below). To monitor annealing, RNAs are electrophoresed in a 2% agarose gel in TBE buffer and stained with ethidium bromide. See, *e.g.*, Sambrook et al., *supra*.

Lysate Preparation

40 Untreated rabbit reticulocyte lysate (Ambion) are assembled according to the manufacturer's directions. dsRNA is incubated in the lysate at 30° C for 10 min prior to the addition of mRNAs. Then NOVX mRNAs are added and the incubation continued for an additional 60 min. The molar ratio of double stranded RNA and mRNA is about 200:1. The NOVX mRNA is radiolabeled and its stability is monitored by gel electrophoresis.

5 In a parallel experiment made with the same conditions, the double stranded RNA is internally radiolabeled with a ^{32}P -ATP. Reactions are stopped by the addition of 2X-proteinase-K buffer and deproteinized [see Genes Dev., 13:3191 (1999)]. Products are analyzed by electrophoresis in 15% or 18% polyacrylamide sequencing gels using appropriate RNA standards. By monitoring the gels for radioactivity, the natural production of 10 to 25 nt RNAs from the double
10 stranded RNA can be determined.

The band of double stranded RNA, about 21-23 bps, is eluted. The efficacy of these 21-23 mers for suppressing NOVX transcription is assayed in vitro using the same rabbit reticulocyte assay described above using 50 nanomolar of double stranded 21-23 mer for each assay. The sequence of these 21-23 mers is then determined using standard nucleic acid sequencing techniques.

15 RNA Preparation

21 nt RNAs, based on the sequence determined above, are chemically synthesized using Expedite RNA phosphoramidites and thymidine phosphoramidite. Synthetic oligonucleotides are deprotected and gel-purified (see Genes & Dev. 15:188 (2001)], followed by Sep-Pak C18 cartridge (Waters, Milford, A) purification (Biochemistry, 32:11658 (1993)).

20 These RNAs (20 μM) single strands are incubated in annealing buffer (100 mM potassium acetate, 30 mM HEPES-KOH at pH 7.4, 2 mM magnesium acetate) for 1 min at 90° C followed by 1 h at 37° C.

Cell Culture

A cell culture known in the art to regularly express NOVX is propagated using standard
25 conditions. 24 hours before transfection, at approx. 80% confluency, the cells are trypsinized and diluted 1:5 with fresh medium without antibiotics (1-3 X 10⁵ cells/ml) and transferred to 24-well plates (500 μl /well). Transfection is performed using a commercially available lipofection kit and NOVX expression is monitored using standard techniques with positive and negative control. A positive control is cells that naturally express NOVX while a negative control is cells that do not express
30 NOVX. Base-paired 21 and 22 nt siRNAs with overhanging 3' ends mediate efficient sequence-specific mRNA degradation in lysates and in cell culture. Different concentrations of siRNAs are used. An efficient concentration for suppression in vitro in mammalian culture is between 25 nM to 100 nM final concentration. This indicates that siRNAs are effective at concentrations that are several orders of magnitude below the concentrations applied in conventional antisense or
35 ribozyme gene targeting experiments.

The above method provides a way both for the deduction of NOVX siRNA sequence and the use of such siRNA for in vitro suppression. In vivo suppression may be performed using the same siRNA using well known in-vivo transfection or gene therapy transfection techniques.

Antisense Nucleic Acids

40 Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2n-1, 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

5 an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO:2*n*, or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO:2*n*-1 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. 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.

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense
15 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
20 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 can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to
25 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-carboxymethylaminomethyl-2-thiouridine, 5-(carboxyhydroxymethyl) uracil,
30 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 5-methoxyuracil, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, 2-thiouracil, 4-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 2-methylthio-N6-
35 isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation.

40 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. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the
45 major groove of the double helix. An example of a route of administration of antisense nucleic acid

5 molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

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

Ribozymes and PNA Moieties

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

25 In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX cDNA disclosed herein (i.e., SEQ ID NO:2n-1, wherein n is an integer between 1 and 16). 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. Patents 4,987,071 and 35 5,116,742. NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules.

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 *Anticancer Drug Des.* 6:569 (1991); *Ann. N.Y. Acad. Sci.* 660:27 (1992); and *Bioassays* 14: 807 (1992)].

40 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

5 molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids [see *Bioorg Med Chem* 4: 5 (1996)].

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
10 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, or as probes or primers for DNA sequence and hybridization.

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
15 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
20 selected in terms of base stacking, number of bonds between the nucleotide bases, and orientation. The synthesis of PNA-DNA chimeras can be performed as described in *Nucl Acids Res* 24:3357 (1996). 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
25 [see *Nucl Acid Res* 17:5973 (1989)]. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment [see *Bioorg. Med. Chem. Lett.* 5:1119 (1975)]

In other embodiments, the oligonucleotide may include other appended groups such as
30 peptides or agents facilitating transport across the cell membrane [see *PNAS U.S.A.* 86:6553 (1989); *PNAS* 84: 648 (1987)] or the blood-brain barrier (PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents [see *BioTechniques* 6:958 (1988)] or intercalating agents [see *Pharm. Res.* 5:539 (1988)]. To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, a hybridization triggered
35 cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in any one of SEQ ID NO:2n while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment
40 thereof.

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

5 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
10 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.

15 Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence of SEQ ID NO:2*n* 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.
20 A biologically-active portion of a NOVX protein can be a polypeptide that 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.

25 In an embodiment, the NOVX protein has an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 16. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO:2*n* and retains the functional activity of the protein of SEQ ID NO:2*n*, 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
30 acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO:2*n* and retains the functional activity of the NOVX proteins of SEQ ID NO:2*n*.

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
35 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
40 nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. [see, Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453]. Using GCG GAP software with the following settings for nucleic acid

5 sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence of SEQ ID NO:2*n*-1.

10 "Substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

15 The invention also provides NOVX chimeric or fusion proteins. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a NOVX protein of SEQ ID NO:2*n*, 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
20 protein. In one embodiment, a NOVX fusion protein comprises at least one biologically-active portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at least two biologically-active portions of a NOVX protein. In yet another embodiment, a NOVX fusion protein comprises at least three biologically-active portions of a NOVX protein.

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

In another embodiment, the fusion protein is a NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.
30 In yet another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal
35 transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject,
40 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

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

15 **NOVX Agonists and Antagonists**

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the
20 naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset
25 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 or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment,
30 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
35 set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of
40 potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art.

Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of

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

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products
 15 having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which
 20 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 *PNAS USA* 89:7811 (1992); *Protein Engineering* 6:327 (1993)].

Anti-NOVX Antibodies

25 Included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins.
 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
 30 the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence of SEQ ID NO:2n, wherein n is an integer between 1 and 16, 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
 35 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
 40 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. See, e.g., Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and

5 Doolittle 1982, *J. Mol. Biol.* 157: 105-142. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

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

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

20 Various well known and standard 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.

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 of the sequence of a human immunoglobulin, and contain 30 minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the methods described in Nature, 321:522 (1986); Nature, 332:323-327 (1988); or 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 35 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 40 immunoglobulin constant region (Fc), typically that of a human immunoglobulin..

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

5 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. 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. PNAS USA 80: 10 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 (J. Mol. Biol., 227:381 (1991); J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals. For 15 example, 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; and 5,661,016.

20 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 (PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's 25 genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, i.e., the Xenomouse™ [see PCT publications WO 96/33735 and WO 30 96/34096]. This animal produces B cells that secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to 35 obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous 40 heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

5 PCT/US2005/025762
A method for producing an antibody of interest, such as a human antibody (U.S. Patent No. 5,916,771) 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.

10 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 (PCT publication WO 99/53049).

F_{ab} Fragments and Single Chain Antibodies

15 According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries [see Science 246:1275 (1989)] 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
20 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.

25 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.
30 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 (Nature, 305:537 (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
35 which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography. Similar procedures are disclosed in WO 93/08829, and EMBO J., 10:3655 (1991).

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

5 separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Methods in Enzymology, 121:210 (1986).

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

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g., F(ab')₂ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Science [229:81 (1985)] describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')₂ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. J. Exp. Med. 175:217(1992) describe the production of a fully humanized bispecific antibody F(ab')₂ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described in PNAS USA (90:6444 (1993)) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker that 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

5 making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported [see J. Immunol. 152:5368 (1994)].

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

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

Heteroconjugate Antibodies

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

Effector Function Engineering

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

Immunoconjugates

40 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

5 bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

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

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

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

Immunoliposomes

30 The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art [see U.S. Pat. Nos. 4,485,045 and 4,544,545]. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in 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 J. National Cancer Inst., 81(19):1484 (1989)].

40 Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme linked immunosorbent assay (ELISA) and other immunologically mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of

5 hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Antibodies directed against a NOVX protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of a NOVX protein (*e.g.*, for use in
10 measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies specific to a NOVX protein, or derivative, fragment, analog or homolog thereof, that contain the antibody derived antigen binding domain, are utilized as pharmacologically active compounds (referred to hereinafter as "Therapeutics").

15 An antibody specific for a NOVX protein of the invention (*e.g.*, a monoclonal antibody or a polyclonal antibody) can be used to isolate a NOVX polypeptide by standard techniques, such as immunoaffinity, chromatography or immunoprecipitation. An antibody to a NOVX polypeptide can facilitate the purification of a natural NOVX antigen from cells, or of a recombinantly produced NOVX antigen expressed in host cells. Moreover, such an anti-NOVX antibody can be used to detect the
20 antigenic NOVX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic NOVX protein. Antibodies directed against a NOVX protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples
25 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,
30 fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibody Therapeutics

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

5 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 that may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

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

Pharmaceutical Compositions of Antibodies

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

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

5 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.

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

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

ELISA Assay

 An agent for detecting an analyte protein may be a polyclonal or monoclonal antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. An intact
25 antibody, or a fragment thereof (*e.g.*, F_{ab} or F_{(ab)2}) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and
30 end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be
35 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
40 hybridizations. Furthermore, *in vivo* techniques for detection of an analyte protein include introducing into a subject a labeled anti-analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

5 NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular
10 double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell
15 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, useful expression vectors in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most
20 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
25 expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the
30 vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements. 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
35 sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, *etc.* The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including
40 fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, *etc.*).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells

5 or mammalian cells. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (*Gene* 67:31 (1998)), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

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

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

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

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 [see *Mol. Cell. Biol.* 3:2156 (1983)], and the pVL series [see *Virology* 170:31 (1989)].

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 (*Nature* 329:840 (1987)) and pMT2PC (*EMBO J.* 6:187 (1987)). 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.

45 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

5 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; *Genes Dev.* 1:268 (1987)), lymphoid-specific promoters (*Adv. Immunol.* 43:235 (1988)), in particular promoters of T cell receptors (*EMBO J.* 8:729 (1989)) and immunoglobulins (*Cell* 33: 729 (1983); *Cell* 33: 741 (1988)), neuron-specific promoters (e.g., the neurofilament promoter; *PNAS USA* 86: 5473 (1989)), pancreas-specific promoters (*Science* 230: 912 (1985)), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (*Science* 249:374 (1990)) and the α -fetoprotein promoter (*Genes Dev.* 3:537 (1989)).

15 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, 20 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.

25 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 30 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 35 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 40 co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

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 45 of interest. Various selectable markers include those that confer resistance to drugs, such as G418,

5 hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

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

Transgenic NOVX Animals

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

35 A transgenic animal in accordance with the invention can be created by introducing a NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences, *i.e.*, any one of SEQ ID NO:2*n*-1 may be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the
40 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

5 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. 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
10 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
15 alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of any one of SEQ ID NO:2*n*-1, but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO:2*n*-1 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
20 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
25 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,
30 several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector [see Cell 51:503 (1987)] 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 Cell, 69:915 (1992)].

35 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
40 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 *Curr. Opin. Biotechnol.* 2: 823 (1991) and 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
45 selected systems that allow for regulated expression of the transgene. One example of such a

5 system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, see, *PNAS USA* 89: 6232 (1992). Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* [see *Science* 251:1351 (1991)]-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
10 can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in *Nature* 385: 810 (1997). In brief, a cell (e.g., a somatic cell)
15 from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal
20 will be a clone of the animal from which the cell is isolated.

Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such
25 compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's
30 Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional
35 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
40 administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as

5 acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions
10 (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage
15 and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the
20 use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an
25 agent that 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
30 contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be
35 enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials
40 can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose
45 or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

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

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

15 The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery. In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such
20 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
25 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
30 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.

35 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 or by stereotactic injection. 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
40 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.

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

5 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
10 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
15 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

20 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.

25 In one embodiment, the invention provides assays for screening candidate or test compounds that 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;
30 synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. See, *e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

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

40 Examples of methods for the synthesis of molecular libraries can be found in the art, for example in *PNAS U.S.A.* 90:6909 (1993); *PNAS U.S.A.* 91:11422 (1994); *J. Med. Chem.* 37:2678 (1994); *Science* 261:1303 (1993); *Angew. Chem. Int. Ed. Engl.* 33:2059 (1994); and *J. Med. Chem.* 37:1233 (1994). Libraries of compounds may also be presented in solution (*Biotechniques* 13: 412 (1992)), or on beads (*Nature* 354:82 (1991)), on chips (*Nature* 364:555 (1993)), bacteria (U.S. Patent

- 5 No. 5,223,409), spores (U.S. Patent 5,233,409), plasmids (*PNAS USA* 89:1865 (1992) or on phage (U.S. Patent No. 5,233,409.).

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

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining
30 the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule
35 in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention. In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.*, a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and
40 into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

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

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

10 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.

20 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*.

25 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.

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

5 dimethylamminiol-1-propane sulfonate (CHAPS), or
3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

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

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

5 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.

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

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

In yet another aspect of the invention a method for identifying compounds that modulate target polypeptide (NOVX) activity is disclosed wherein the method comprises: (a) combining a test compound with a target polypeptide and a substrate of the target polypeptide; and (b) determining
30 whether the test compound modulates the activity of the target polypeptide; wherein the target polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:2n, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain of SEQ ID NO:2n.

35 The method further comprises a step of identifying the test compound that modulates the target polypeptide activity by modulating the target polypeptide activity as modulator of the target polypeptide. Such modulator could be an inhibitor, an activator, an antagonist, or an agonist of NOVX target polypeptide.

The method also further comprises a step of identifying the test compound that modulates
40 the target polypeptide activity as an enhancer of insulin secretion, or as a therapeutic for treatment of insulin resistance, obesity and/or diabetes.

In the above described method, the target polypeptide (NOVX) could be an isolated polypeptide.

45 The target polypeptide may be produced by a process comprising culturing a recombinant host cell, the recombinant host cell comprising a nucleic acid encoding the target polypeptide, under

5 ~~FIG. 1~~ conditions promoting expression of the target polypeptide. In such a method, the nucleic acid comprises a nucleotide sequence selected from the group consisting of: (a) SEQ ID NO:2n-1; (b) nucleotides encoding an amino acid sequence of the at least one domain of SEQ ID NO:2n; and (c) a nucleotide sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2n, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain of SEQ ID NO:2n.

10 Alternatively, the target polypeptide could be produced by expression of a recombinant vector comprising a nucleic acid, the nucleic acid encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2n, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain of SEQ ID NO:2n. Here, the test compound could be combined with the target polypeptide in a mammalian cell grown in culture. Also, the test compound could be combined with the target polypeptide in vitro. In this method, the nucleic acid comprises a nucleotide sequence selected from the group consisting of: (a) SEQ ID NO:2n-1; (b) nucleotides encoding an amino acid sequence of the at least one domain of SEQ ID NO:2n; and (c) a nucleotide sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2n, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain of SEQ ID NO:2n.

25 In yet another embodiment, the target polypeptide is produced by expression of an endogenous nucleic acid, the endogenous nucleic acid encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2n, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain of SEQ ID NO:2n. Here as well, the test compound could be combined with the target polypeptide in a mammalian cell grown in culture. Also, the test compound could be combined with the target polypeptide in vitro. In this method, the nucleic acid comprises a nucleotide sequence selected from the group consisting of: (a) SEQ ID NO:2n-1; (b) nucleotides encoding an amino acid sequence of the at least one domain of SEQ ID NO:2n; and (c) a nucleotide sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2n, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain of SEQ ID NO:2n.

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

40 **Detection Assays**

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an

5 individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample.

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
10 chromosome mapping. Accordingly, portions or fragments of the NOVX sequences of SEQ ID NO:2n-1 or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

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

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

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

35 Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that
40 the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. For a review of this technique, see, Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites

5 and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, 10 *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected 15 with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based 20 on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more 25 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 30 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 35 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 40 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms.

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

- 5 differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If coding sequences, such as those of SEQ ID NO:2*n*-1 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

- 10 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
15 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
20 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
25 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 (*i.e.*, "pharmacogenomics"). Pharmacogenomics allows for the selection of agents
30 (*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.

Diagnostic Assays

- 35 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
40 sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1 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.

5 An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

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

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

40 Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic

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

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide,
15 nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and
20 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
25 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
35 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,
40 buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (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) (*Science* 241:1077 (1988); and *PNAS USA* 91: 360 (1994)), the latter of which can be particularly useful for detecting point mutations
45 in the NOVX-gene (*Nucl. Acids Res.* 23: 675 (1995)).

5 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
10 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 (*PNAS USA* 87:1874 (1990)), transcriptional amplification system (*PNAS USA* 86:1173 (1989)); Q β Replicase
15 (*BioTechnology* 6: 197 (1988)), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

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

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes [see *Human Mutation* 7:244 (1996); *Nat. Med.* 2:753 (1996) . For example, genetic mutations in NOVX can be identified in two-dimensional arrays containing
30 light-generated DNA. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations
35 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 [e.g. those described in *PNAS*
40 *USA* 74:560 (1997) or *PNAS USA* 74: 5463 (1977)]. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays [see *Biotechniques* 19:448 (1995)], including sequencing by mass spectrometry [see *Adv. Chromatography* 36:127 (1996); or *-Appl. Biochem. Biotechnol.* 38:147 (1993)]. .

Other methods for detecting mutations in the NOVX gene include methods in which
45 protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA

5 "heteroduplexes [see *Science* 230:1242 (1985)]. 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.

10 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 *PNAS USA* 85:4397 (1988); *Methods Enzymol.* 217: 286 (1992)]. 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 (*Carcinogenesis* 15: 1657 (1994)). 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

25 detected from electrophoresis protocols or the like [see U.S. Patent No. 5,459,039].

In other embodiments, alterations in electrophoretic mobility may 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., *PNAS USA*: 86: 2766 (1989), *Mutat. Res.* 285: 125 (1993)]. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility [see *Trends Genet.* 7:5 (1991)].

30

35

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant may be assayed using denaturing gradient gel electrophoresis (DGGE) [see *Nature* 313:495 (1985)]. 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 may be used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA [see *Biophys. Chem.* 265:12753 (1987)].

40

5 " 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 a 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., *Nature* 324:163 (1986); and *PNAS USA* 86: 6230 (1989)]. Such allele specific
10 oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

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

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

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

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (e.g., NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders. The disorders include but are not
35 limited to, e.g., those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table 1.

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
40 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

5 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 (*e.g.*, *Clin. Exp.*
10 *Pharmacol. Physiol.*, 23: 983 (1996) or *Clin. Chem.*, 43:254 (1997)). 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,
15 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
20 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
25 (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
30 the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

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

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied

5 "not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

15 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.

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

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions

5 listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table 1.

Diseases and Disorders

10 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.*,
15 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
20 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
25 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).
30 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

35 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.
40 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

5 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
10 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
15 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
20 methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a NOVX protein or nucleic acid
25 molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situations* in which NOVX is abnormally 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
30 such a situation is where the subject has a gestational disease (e.g., preclampsia).

Determination of the Biological Effect of the Therapeutic

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

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

5 **Prophylactic and Therapeutic Uses of the Compositions of the Invention**

The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders. As, for example, those associated with homologs of a NOVX protein, such as those summarized in Table 1.

10 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 various diseases, disorders, conditions and the like.

15 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.

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

NOVX Polypeptides

The following sections describe in detail the NOVX polypeptides of the invention and methods of screening for modulators of NOVX polypeptides:

A. NOV1 -- Human Aquaporin Adipose

25 The mammalian aquaporin family of water-transporting channels has 11 members (Agre P, J Physiol. 2002 Jul 1;542(Pt 1):3-16. Review.). Three of the aquaporins also transport glycerol in addition to water; the aquaglyceroporins are aquaporins 3, 7 and 9. Aquaporin 7 (AQP7) is also called aquaporin adipose because it is predominantly found in adipose tissue, while the
30 aquaglyceroporins 7 and 9 are not highly expressed in that tissue (Kishida K, J Biol Chem. 2000 Jul 7;275(27):20896-902.). AQP7 is an integral membrane protein that transports glycerol out of the adipocyte (fat cell). In the mouse, AQP7 levels are increased by fasting, and suppressed by refeeding, in parallel with plasma levels of free fatty acids and glycerol (Kishida et al.). Suppression by feeding, and an increase by fasting, suggest that AQP7 is the physiological channel for adipocyte
35 glycerol efflux in humans. In the fasting state, adipocytes release free fatty acids and glycerol, the end products of triglyceride hydrolysis. The breakdown of adipocyte triglyceride stores during fasting provides the body with free fatty acids as a source of energy, and glycerol, which serves as a substrate for glucose production (gluconeogenesis) in the liver.

40 Changes in AQP7 levels in man have not yet been reported. AQP7 levels in adipose tissue have been observed to be upregulated in db/db mice (Kuriyama H, Diabetes. 2002 Oct;51(10):2915-21). The db/db strain of mouse is a model for obesity and Type 2 diabetes. However, this model is not reflective of the majority of obesity and Type 2 diabetes in man. Db/db mice are obese and diabetic because they have a mutation in the leptin receptor which leads to hyperphagia (increased eating), obesity and dysregulated glucose and lipid metabolism (Chen H, Cell. 1996 Feb 9;84(3):491-
45 5.) No more than 3% of humans with obesity and/or Type 2 diabetes have leptin receptor mutations.

5

In particular the invention relates to the use of the aquaporin adipose protein as a diagnostic and/or target for small molecule drugs and antibody therapeutics. The inventors have discovered that the CG181171-01 mRNA was observed to be upregulated two-fold in the adipose tissue of a genetic model of obesity (of unknown pathogenesis) in mice. CG181171-01 mRNA was also observed to be upregulated in adipose tissue during the transition from normoglycemia to hyperglycemia in a diet-induced obesity (DIO) model in mice. Data from the DIO model demonstrate that release of glycerol through CG181171-01 contributes to the development of hyperglycemia. Diet-induced obesity is considered to be the main cause of Type 2 diabetes in man. Inhibition of CG181171-01 inhibits hepatic glucose production and is a treatment to prevent or ameliorate fasting hyperglycemia in Type 2 diabetes.

In a particular embodiment of the invention, aquaporin adipose is a target for screening. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify aquaporin adipose antagonist therapeutic antibodies and/or therapeutic small molecules beneficial in the treatment of obesity and/or insulin resistance and diabetes.

Not to be limited by a particular mechanism of action, the inventors nevertheless have discovered that inhibition of aquaporin adipose has beneficial effects for treating obesity and/or diabetes by acting in many metabolic tissues, including adipose. Specifically, in adipose tissue, aquaporin adipose inhibition will lead to a decrease in glycerol release that will modulate hepatic glucose production and hyperglycemia in diabetes. The finding that CG181171-01 causes up-regulation of the aquaporin adipose in adipose tissue of obese mice and in adipose tissue of mice during the transition from normoglycemia to hyperglycemia, indicates a critical role for aquaporin adipose in peripheral metabolism. Therefore, an antagonist of aquaporin adipose is useful for the treatment of obesity and/or insulin resistance and diabetes.

Thus, aquaporin adipose nucleic acids and proteins are useful for screening for an inhibitor/antagonist of aquaporin adipose for the treatment of obesity and or diabetes. These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in diagnostic and/or therapeutic methods.

Furthermore, our results indicate that a modulator of aquaporin adipose activity, such as an inhibitor, activator, antagonist, or agonist of aquaporin adipose may be useful for treatment of such disorders as obesity, diabetes, and insulin resistance, as well as for enhancement of insulin secretion.

Discovery Process

The following sections describe the study design(s) and the techniques used to identify the aquaporin adipose – encoded protein, and any variants thereof, as being suitable as diagnostic markers, targets for antibody therapeutic and targets for a small molecule drugs for treatment of Obesity and Diabetes.

5 Example A1. Genetically Obese Mice vs. Genetically Lean Mice Study - Protocol for Genetically Obese Mice vs. Genetically Lean Mice Study is disclosed in Example Q6.

10 A large number of mouse strains have been identified that differ in body mass and composition. The AKR and NZB strains are obese, the SWR, C57L and C57BL/6 strains are of average weight whereas the SM/J and Cast/Ei strains are lean. Understanding the gene expression differences in the major metabolic tissues from these strains will elucidate the pathophysiological basis for obesity. These specific strains of rat were chosen for differential gene expression analysis because quantitative trait loci (QTL) for body weight and related traits had been reported in published genetic studies. Tissues included whole brain, skeletal muscle, visceral adipose, and liver.

15 Results

A fragment of the mouse aquaporin adipose gene was initially found to be up-regulated by 2.2 fold in adipose from genetically obese mice (AKR strain) relative to adipose from C57L/J average weight mice using CuraGen Corporation's GeneCalling® method of differential gene expression (described in Example Q7). A differentially expressed mouse gene fragment migrating, at approximately 143 nucleotides in length was definitively identified as a component of the mouse aquaporin adipose cDNA. The method of competitive PCR was used for confirmation of the gene assessment. The electropherographic peaks corresponding to the gene fragment of the mouse aquaporin adipose were ablated when a gene-specific primer (shown in Table A1) competes with primers in the linker-adaptors during the PCR amplification. The peaks at 143 nt in length were ablated in the sample from both the genetically obese AKR mice and the average weight C57L/J strain.

Table A1. Competitive PCR primer for the mouse aquaporin adipose: Nucleotides 466-1311 of the 1311 nucleotide-long cDNA sequence of GENBANK-ID: AB010100 are shown. Gene-specific primers used for competitive PCR are shown below in bold (fragment from 947 to 1089 in **bold**, band size: 191). The gene-specific primers at the 5' and 3' ends of the fragment are underlined.

466 AGGAGAAAAC TCTGGCAGCT ATCTCGGTGT CAACTTGGGT TTTGGATTCTG
GAGTGACCAT

526 GGGAGTCCAT GTAGCAGGCG GCATCTCTGG GGCCCACATG AATGCCGCAG
TGACTTTCAC

586 CAATTGTGCA CTAGGCCGAA TGACCTGGAA GAAGTTCCT GTATATGTGC
TGGGTCAGTT

646 CCTGGGCTCC TTCTCAGCTG CAGCTACCAC CTACTTAATT TTCTATGGTG
CCATTAACCA

706 CTTTGCAGGC GGAGACCTGT TGGTGACAGG TTCCAAGGCC ACTGCAAACA
TTTTTGCCAC

766 CTATCTTCCT GAATACATGA CACTGTGGCG GGGCTTCCTG GATGAGGCAT
TCGTGACTGG

826 GATGCTGCAG CTGTGTCTCT TCGCCATCAC CGACAAGAAG AACAGTCCAG
CACTTCAAGG

886 GACCGAGCCC CTCGTGATAG GCATCCTTGT TACCGTCCTT GGGGTGTGCG
TAGGCATGAA

946 CTCCGGATAT GCAATCAACC CATCCCGTGA CCTGCCTCCC CGGTTGTTCA
CTTTCATTGC

1006 TGGCTGGGGC AAACAAGTGT TCAGAGCCGG AAACAACTGG TGGTGGGTGC
CGGTGGTGGC

1066 ACCACTTCTG GGCGCCTACC TAGGTGGTAT TGTATACCTG GGTTTAATTC
ACCCAGCAT

1126 ACCACAGGAT CCTCAGAGAC TGGAGAATTT TACAGCAAGA GACCAGAAGG
TAACTGCATC

1186 ATACAAGAAT GCAGCTTCTG CAAACATAAG TGGCTCTGTG CCTCTAGAGC
ACTTCTAAGT

1246 AGAGCTTCTC TTAAACCACA ACCGTACTGC AATAAAGCAC ACCACCTATC
AAAAAAAAAA

1306 AAAAAA (SEQ ID NO: 33)

5

Example A2. Mouse Dietary – Induced Obesity Study (BP24.02)

A protocol for Mouse Dietary-Induced Obesity study is disclosed in Example Q1.

10

The predominant cause for obesity in clinical populations is excess caloric intake. This so-called diet-induced obesity (DIO) is mimicked in animal models by feeding high fat diets of greater than 40% fat content. The DIO study was established to identify the gene expression changes contributing to the development and progression of diet-induced obesity. In addition, the study design seeks to

15 identify the factors that lead to the ability of certain individuals to resist the effects of a high fat diet and thereby prevent obesity. The sample groups for the study were selected from C57BL/6J mice and had body weights +1 S.D. (sd1), + 4 S.D. (sd4) and + 7 S.D. of the chow-fed controls (below). In addition, the biochemical profile of the + 7 S.D. mice revealed a further stratification of these animals into mice that retained a normal glycemic profile in spite of obesity (ngsd7) and mice that

20 demonstrated hyperglycemia (hgds7). Tissues examined included hypothalamus, brainstem, liver, retroperitoneal white adipose tissue (WAT), epididymal WAT, brown adipose tissue (BAT), gastrocnemius muscle (fast twitch skeletal muscle) and soleus muscle (slow twitch skeletal muscle). The differential gene expression profiles for these tissues should reveal genes and pathways that can be used as therapeutic targets for obesity.

25

Results

A fragment of the mouse aquaporin adipose gene was initially found to be up-regulated by 1.6 fold in the adipose (epididymal fat pads) of hyperglycemic obese mice relative to euglycemic obese mice using CuraGen's GeneCalling® method of differential gene expression (described in Example Q7). A differentially expressed mouse gene fragment migrating, at approximately 191

30 nucleotides in length was definitively identified as a component of the mouse aquaporin adipose

5' cDNA. The method of competitive PCR was used for confirmation of the gene assessment. The electropherographic peaks corresponding to the gene fragment of the mouse aquaporin adipose were ablated when a gene-specific primer (shown in Table A2) competes with primers in the linker-adaptors during the PCR amplification. The peaks at 191 nt in length were ablated in the sample from both the hyperglycemic obese mice and the euglycemic mice. The mouse aquaporin adipose gene was also found to be up-regulated by 1.7 fold in the adipose (efp) of obese, hyperglycemic (hgsd7) mice relative to normal weight, chow-fed controls and up-regulated by 1.8 fold in adipose (efp and retroperitoneal fat pads) of sd4 mice (mice with body weights 4 standard deviations above the chow-fed normal controls) relative to ngsd7 mice (normoglycemic mice with weights 7 standard deviations above the control mice). Taken together, these data show that inhibition of aquaporin adipose is useful for the inhibition of glycerol release and the reduction of the development of hyperglycemia in type 2 diabetes. In addition, the finding that aquaporin adipose is up-regulated in diabetic adipose supports the hypothesis that inhibition of aquaporin adipose is beneficial for the treatment of diabetes.

Table A2. Competitive PCR primer for the mouse aquaporin adipose: Nucleotides 466-1311 of the 1311 nucleotide-long cDNA sequence of GENBANK-ID: AB010100 are shown. Gene-specific primers used for competitive PCR are shown below in bold (fragment from 947 to 1137 in **bold**, band size: 191). The gene-specific primers at the 5' and 3' ends of the fragment are underlined.

466 AGGAGAAAAC TCTGGCAGCT ATCTCGGTGT CAACTTGGGT TTTGGATTCTG
GAGTGACCAT

526 GGGAGTCCAT GTAGCAGGCG GCATCTCTGG GGCCACATG AATGCCGCAG
TGACTTTCAC

586 CAATTGTGCA CTAGGCCGAA TGACCTGGAA GAAGTTCCT GTATATGTGC
TGGGTCAGTT

646 CCTGGGCTCC TTCTCAGCTG CAGCTACCAC CTACTTAATT TTCTATGGTG
CCATTAACCA

706 CTTTGCAGGC GGAGACCTGT TGGTGACAGG TTCCAAGGCC ACTGCAAACA
TTTTTGCCAC (SEQ ID NO: 34)

766 CTATCTTCCT GAATACATGA CACTGTGGCG GGGCTTCCTG GATGAGGCAT
TCGTGACTGG

826 GATGCTGCAG CTGTGTCTCT TCGCCATCAC CGACAAGAAG AACAGTCCAG
CACTTCAAGG

886 GACCGAGCCC CTCGTGATAG GCATCCTTGT TACCGTCCTT GGGGTGTGCG
TAGGCATGAA

946 CTCCGGATAT GCAATCAACC CATCCCGTGA CCTGCCTCCC CGGTTGTTCA
CTTTCATTGC

1006 TGGCTGGGGC AAACAAGTGT TCAGAGCCGG AAACAAGTGG TGGTGGGTGC
CGGTGGTGGC

1066 ACCACTTCTG GGCGCCTACC TAGGTGGTAT TGTATACCTG GGTTTAATTC
ACCCAGCAT

1126 ACCACAGGAT CCTCAGAGAC TGGAGAATTT TACAGCAAGA GACCAGAAGG
TAACTGCATC

1186 ATACAAGAAT GCAGCTTCTG CAAACATAAG TGGCTCTGTG CCTCTAGAGC
ACTTCTAAGT

1246 AGAGCTTCTC TTTAACCACA ACCGTACTGC AATAAAGCAC ACCACCTATC
AAAAAAAAAA

1306 AAAAAA (SEQ ID NO: 33)

Example A3. Human Aquaporin-Adipose Sequence Identification

The sequence of Human aquaporin-adipose was derived by laboratory screening of cDNA library. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were 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 protocol for identification of human sequence(s) is disclosed in Example Q8.

Table A3 shows an alignment (ClustalW) of the protein sequences of the human (CG181171-01) and mouse homologs of the aquaporin adipose. Table A4 shows protein sequence of mouse homolog of the aquaporin adipose.

Table A3. An alignment (ClustalW) of the protein sequences of the human (CG181171-01) (SEQ ID NO: 2) and mouse homologs of the aquaporin adipose (SEQ ID NO: 35).			
NP_031499_Aquaporin_Mus	1	MAPRSVLETIQSVLQKNMVREFLAEFLLSTYVMMVFGLGSAHVMVL	45
CG181171-01-PROT	1	MVQASGHRSTRGSKMVSWSVIAKIQEILQKRMVREFLAEFMSTYVMMVFGLGSAHVMVL	60
NP_031499_Aquaporin_Mus	46	GENSGSYLGVNLGFGFGVTMGVHVAGGISGAHMNAAVTFTNCALGRMTWKKFPVYVLGQF	105
CG181171-01-PROT	61	NKKYGSYLGVNLGFGFGVTMGVHVAGRISGAHMNAAVTFANALGRVYWKKFPVYVLGQF	120
NP_031499_Aquaporin_Mus	106	LGSFSAAATTYLIFYGAINHFAGDGLLVTCGSKATANIFATYLPPEYMTLWRGFLDEAFVTG	165
CG181171-01-PROT	121	LGSFLAAATTYSLIFYTAILHFSGQLMVTGPFVATAGIFATYLPDHMTLWRGFLNEAWLTG	180
NP_031499_Aquaporin_Mus	166	MLQLCLFAITDKKNSPALQGTEPLVIGILVTVLGVSLGMNSGYAINPSRDLPPRLFTFFIA	225
CG181171-01-PROT	181	MLQLCLFAITDQENNPALPGTEALVIGILVTVLGVSLGMNTGYAINPSRDLPPRLFTFFIA	240
NP_031499_Aquaporin_Mus	226	GWGKQVFRAGNNWWWVPVWAPLLGAYLGGIVYVLGLIHPSIPQDPQRLNFTARDQ	281
CG181171-01-PROT	241	GWGKQVFSNGENWWWVPVWAPLLGAYLGGIYLVFLGSTIPREPLKLEDVAYEDHGITY	300
NP_031499_Aquaporin_Mus	282	VTASYKNAASANISGSSVPLEHF	303
CG181171-01-PROT	301	LPKMGSHPTISPLTPVSVSPANRSSVHPAPFLHESMALEHF	342

Table A4. Protein sequence of mouse homolog of the aquaporin adipose.	
>NP_031499_Aquaporin_Mus	
MAPRSVLETIQSVLQKNMVREFLAEFLLSTYVMMVFGLGSAHVMVLGENSGSYLGVNLGFGFGVTMGVHVA GGISGAHMNAAVTFTNCALGRMTWKKFPVYVLGQFLGSFSAAATTYLIFYGAINHFAGDGLLVTCGSKATA NIFATYLPPEYMTLWRGFLDEAFVTGMLQLCLFAITDKKNSPALQGTEPLVIGILVTVLGVSLGMNSGYAI NPSRDLPPRLFTFIAGWGKQVFRAGNNWWWVPVWAPLLGAYLGGIVYVLGLIHPSIPQDPQRLNFTARDQ KVTASYKNAASANISGSSVPLEHF (SEQ ID NO: 34)	

5 The laboratory cloning was performed using one or more of the methods summarized in Example Q8.

The NOV1 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table A5.

Table A5. NOV1 Sequence Analysis				
NOV1a,	CG181171-01	SEQ ID NO: 1	1256 bp	
DNA Sequence		ORF Start: ATG at 173	ORF Stop: TAA at 1199	
GGCTCTGGACTGGGGACACAGGGATAGCTGAGCCCCAGCTGGGGGTGGAAGCTGAGCCAGGGACAGTC ACGGAGGAACAAGATCAAGATGCGCTGTAAGTGAAGCCCCCAAGGCGGAGGCTGAGAATCAGAGAC ATTTTCAGCAGACATCTACAAATCTGAAAGACAAAACATGGTTCAAGCATCCGGGCACAGGCGGTCCAC CCGTGGCTCCAAAATGGTCTCCTGGTCCGTGATAGCAAAGATCCAGGAAATACTGCAGAGGAAGATGG TGCGAGAGTTCTTGGCCGAGTTTCATGAGCACATATGTCATGATGGTATTCGGCCTTGGTTCCGTGGCC CATATGGTTCTAAATAAAAAATATGGGAGCTACCTTGGTGTCAACTTGGGTTTGGCTTCGGAGTCAC CATGGGAGTGCACGTGGCAGGCCGCATCTCTGGAGCCCACATGAACGCAGCTGTGACCTTTTGCTAACT GTGCGCTGGGCCGCGTGCCCTGGAGGAAGTTTCCGGTCTATGTGCTGGGGCAGTTCTTGGGCTCCTTC CTGGCGGTGCCACCATCTACAGTCTCTTCTACACGGCCATTTCTCCACTTTTCGGGTGGACAGCTGAT GGTGACCGGTCCCCTGCTACAGCTGGCATTTTGGCCACCTACCTTCTGATCACATGACATTGTGGC GGGGCTTCTGAATGAGGCGTGGCTGACCGGGATGCTCCAGCTGTGTCTCTTCGCCATCACGGACCAG GAGAACAACCCAGCACTGCCAGGAACAGAGGCGCTGGTGTATAGGCATCCTCGTGGTCATCATCGGGGT GTCCCTTGGCATGAACACAGGATATGCCATCAACCCGTCCTGGGACCTGCCCCCGCATCTTACCT TCATTGCTGGTTGGGGCAACAGGTCTTCAGCAATGGGGAGAACTGGTGGTGGGTGCCAGTGGTGGCA CCACTTCTGGGTGCCATCTAGGTGGCATCATCTACCTGGTCTTCATTGGCTCCACCATCCCACGGGA GCCCCTGAAATTGGAGGATTCTGTGGCGTATGAAGACCACGGGATAACCGTATTGCCCAAGATGGGAT CTCATGAACCCACGATCTCTCCCTCACCCCGTCTCTGTGAGCCCTGCCAACAGATCTTCAGTCCAC CCTGCCCCACCCTTACATGAATCCATGGCCCTAGAGCACTTCTAAGCAGAGATTATTTGTGATCCCAT CCATTCCCCAATAAAGCAAGGCTTGTCCGACA				
NOV1a,	CG181171-01	SEQ ID NO: 2	342 aa	MW at 37231.2kD
Protein Sequence				
MVQASGHRSTRGSKMVSWSVIAKIQEILQRKMVREFLAEFMSTYVMMVFGLGSVAHMLNKKYGSYL GVNLFGFGVMTMGVHVAGRISGAHMNAAVTFANCALGRVPWRKFPVYVLGQFLGSFLAAATISLFTY AILHFSGGQLMVTGPVATAGIFATYLPDHMTLWRGFLNEAWLTGMLQLCLFAITDQENNPALPGTEAL VIGILVVIIGVSLGMNTGYAINPSRDLPPRIFTFIAGWGKQVFSNGENWWWVPVAPLLGAYLGIIY LVFIGSTIPREPLKLEDSVAYEDHGITVLPKMGSHIPTISPLTPVSVSPANRSSVHPAPPLHESMALE HF				
NOV1b,	CG181171-02	SEQ ID NO: 3	1048 bp	
DNA Sequence		ORF Start: ATG at 14	ORF Stop: end of sequence	
CACCGGATCCACCATGGTTCAAGCATCCGGGCACAGGCGGTCCACCCGTGGCTCCAAAATGGTCTCCT GGTCCGTGATAGCAAAGATCCAGGAAATACTGCAGAGGAAGATGGTGGCAGAGTTCTTGGCCGAGTTC ATGAGCACATATGTCATGATGGTATTCGGCCTTGGTTCCGTGGCCCATATGGTTCTAAATAAAAAATA TGGGAGCTACCTTGGTGTCAACTTGGGTTTGGCTTCGGAGTCAACATGGGAGTGCACGTGGCAGGCC GCATCTCTGGAGCCCATGAACGCAGCTGTGACCTTTGCTAACTGTGCGCTGGGCCGCGTGCCCTGG AGGAAGTTTCCGGTCTATGTGCTGGGGCAGTTCTTGGGCTCCTTCTGGCGGCTGCCACCATCTACAG TCTCTTCTACACGGCCATTCTCCACTTTTCGGGTGGACAGCTGATGGTGACCGGTCCCGTCGCTACAG CTGGCATTTTTGCCACCTACCTTCTGATCACATGACATTGTGGCGGGGCTTCTGAATGAGGCGTGG CTGACCGGGATGCTCCAGCTGTGTCTCTTCGCCATCACGACCAGGAGAACAACCCAGCACTGCCAGG AACAGAGGCGCTGGTGTAGGCATCCTCGTGGTTCATCATCGGGGTGTCCCTTGGCATGAACACAGGAT ATGCCATCAACCCGTCCCGGGACCTGCCCCCGCATCTTACCTTCATTGCTGGTTGGGGCAAACAG GTCTTCAGCAATGGGGAGAACTGGTGGTGGGTGCCAGTGGTGGCACCCTTCTGGGTGCCATCTAGG TGGCATCATCTACCTGGTCTTCATTGGCTCCACCATCCACGGGAGCCCTGAAATTGGAGGATTCTG TGGCGTATGAAGACCAGGATAACCGTATTGCCCAAGATGGGATCTCATGAACCCAGCATCTCTCCC CTCACCCCGTCTCTGTGAGCCCTGCCAACAGATCTTCAGTCCACCCTGCCCCACCCTTACATGAATC CATGGCCCTAGAGCACTTC				
NOV1b,	CG181171-02	SEQ ID NO: 4	345 aa	MW at 37530.6kD
Protein Sequence				
MVQASGHRSTRGSKMVSWSVIAKIQEILQRKMVREFLAEFMSTYVMMVFGLGSVAHMLNKKYGSYL GVNLFGFGVMTMGVHVAGRISGAHMNAAVTFANCALGRVPWRKFPVYVLGQFLGSFLAAATISLFTY AILHFSGGQLMVTGPVATAGIFATYLPDHMTLWRGFLNEAWLTGMLQLCLFAITDQENNPALPGTEAL VIGILVVIIGVSLGMNTGYAINPSRDLPPRIFTFIAGWGKQVFSNGENWWWVPVAPLLGAYLGIIY LVFIGSTIPREPLKLEDSVAYEDHGITVLPKMGSHIPTISPLTPVSVSPANRSSVHPAPPLHESMALE HF				

5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table A6.

PFam analysis predicts that the NOV1a protein contains the domains shown in the Table A6.

Table A6. Domain Analysis of NOV1a			
Pfam Domain	NOV1a Match Region	Identities/ Similarities for the Matched Region	Expect Value
MIP	27..272	82/268 (31%) 187/268 (70%)	3.8e-59

10

Example A4. Human Aquaporin-Adipose Gene Variants and SNPs

The protocol for obtainment of gene variants and SNPs is disclosed in Example Q11. The variants of the human Aquaporin-Adipose were obtained from direct cloning and/or public databases.

15 In addition to the human version of the Aquaporin-Adipose identified as being differentially expressed in the experimental study, other variants have been identified by direct sequencing of cDNAs derived from many different human tissues and from sequences in public databases. No splice-form variants have been identified at CuraGen whereas several amino acid-changing and non-amino acid-changing cSNPs were identified and are shown in Table A7 below, where UCP=uncharged polar,

20 NP=non polar, A=acidic, and B=basic. Those cSNPs with ID "cgsp" refer to CuraGen proprietary SNPs, whereas those labeled "hsnp" are from public databases. The preferred variant of all those identified, to be used for screening purposes, is CG181171-01.

Table A7. CG181171-01 SNPs

25

SNP ID	NUCLEOTIDE POSITION	ALLELES	AMINO ACID POSITION	ORIG AA	NEW AA	AA CHANGE TYPE
hsnp:rs2381004	273	A/G	34	Val	Ala	Conservative
hsnp:rs1885734	284	C/G	38	Leu	Val	Conservative
hsnp:rs1810674	290	C/T	40	Glu	Lys	A to B
hsnp:rs4008659	347	C/G	59	Val	Leu	Conservative
hsnp:rs4008658	360	G/T	63	Lys	Thr	B to UCP
cgsp:13381473	605	C/G	145	Gln	Glu	UCP to A
hsnp:rs2809564	636	A/G	155	Ala	Val	Conservative
hsnp:rs2809563	651	G/T	160	Thr	Asn	UCP to UCP
hsnp:rs4248652	917	A/G	249	Asn	Asp	UCP to A
hsnp:rs3967459	930	A/C	253	Trp	Leu	Conservative
cgsp:c99.989	1101	G/A	310	Thr	Met	UCP to NP

cgsp:c99.988	1119	G/A	316	Pro	Leu	Conservative
cgsp:c99.797	1121	C/T	317	Val	Ile	Conservative
hsnp:rs4248647	1127	A/G	319	Val	Met	Conservative
hsnp:rs4056299	1134	A/G	321	Pro	Leu	Conservative
hsnp:rs4056298	1157	A/G	329	Pro	Ser	NP to UCP
hsnp:rs2759528	1187	C/G	339	Leu	Val	Conservative

5

Example A5. Expression Profile of the Human Aquaporin Adipose Gene (CG181171-01)

- 10 The protocol for quantitative expression analysis is disclosed in Example Q9.

Expression of gene CG181171-01 was assessed using the primer-probe set Ag8222, described in Table A8. Results of the RTQ-PCR runs are shown in Table A9.

Table A8. Probe Name Ag8222

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aggaaatactgcagaggaagatg-3'	23	249	35
Probe	TET-5'-acatatgtgctcatgaactcggccag-3'- TAMRA	26	284	36
Reverse	5'-aaccaaggccgaataccat-3'	19	314	37

- 15 Table A9. General_screening_panel_v1.7

Column A - Rel. Exp.(%) Ag8222, Run 323386686			
Tissue Name	A	Tissue Name	A
Adipose	100.0	Gastric ca. (liver met.) NCI-N87	0.0
HUVEC	0.0	Stomach	0.1
Melanoma* Hs688(A).T	0.0	Colon ca. SW-948	0.2
Melanoma* Hs688(B).T	0.0	Colon ca. SW480	0.0
Melanoma (met) SK-MEL-5	0.0	Colon ca. (SW480 met) SW620	0.9
Testis	2.9	Colon ca. HT29	0.0
Prostate ca. (bone met) PC-3	0.0	Colon ca. HCT-116	0.1
Prostate ca. DU145	0.1	Colon cancer tissue	0.1
Prostate pool	0.1	Colon ca. SW1116	0.0
Uterus pool	0.2	Colon ca. Colo-205	0.1
Ovarian ca. OVCAR-3	0.0	Colon ca. SW-48	0.7
Ovarian ca. (ascites) SK-OV-3	0.0	Colon	1.2
Ovarian ca. OVCAR-4	0.0	Small Intestine	0.0
Ovarian ca. OVCAR-5	0.0	Fetal Heart	2.5
Ovarian ca. IGROV-1	0.0	Heart	2.6
Ovarian ca. OVCAR-8	0.0	Lymph Node pool 1	0.0

Ovary	0.8	Lymph Node pool 2	11.3
Breast ca. MCF-7	0.1	Fetal Skeletal Muscle	2.6
Breast ca. MDA-MB-231	0.0	Skeletal Muscle pool	0.8
Breast ca. BT-549	0.0	Skeletal Muscle	15.4
Breast ca. T47D	0.0	Spleen	0.1
Breast pool	0.0	Thymus	0.2
Trachea	3.5	CNS cancer (glio/astro) SF-268	0.0
Lung	0.7	CNS cancer (glio/astro) T98G	0.0
Fetal Lung	0.2	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	0.1	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. NCI-H23	0.0	Brain (Amygdala)	0.0
Lung ca. NCI-H460	0.0	Brain (Cerebellum)	1.7
Lung ca. HOP-62	0.0	Brain (Fetal)	0.0
Lung ca. NCI-H522	0.0	Brain (Hippocampus)	0.1
Lung ca. DMS-114	0.0	Cerebral Cortex pool	0.0
Liver	4.1	Brain (Substantia nigra)	0.0
Fetal Liver	2.7	Brain (Thalamus)	0.0
Kidney pool	7.1	Brain (Whole)	0.5
Fetal Kidney	2.2	Spinal Cord	0.0
Renal ca. 786-0	0.0	Adrenal Gland	3.0
Renal ca. A498	0.0	Pituitary Gland	0.0
Renal ca. ACHN	0.0	Salivary Gland	0.9
Renal ca. UO-31	0.0	Thyroid	1.4
Renal ca. TK-10	0.0	Pancreatic ca. PANC-1	0.0
Bladder	0.6	Pancreas pool	0.1

5 Results

This gene is primarily expressed in adipose (CT=24.3), consistent with its characterization as aquaporin adipose.

Example A6. Screening Assay Formulation

10

Assays for screening for antibody therapeutics or small molecule drugs targeting human aquaporin adipose can be formulated utilizing the non-exhaustive list of cell lines that express the aquaporin adipose gene from the RTQ-PCR results shown above. To assay the activity of aquaporin adipose the measurement of glycerol can be utilized using a fluorimetric/colorimetric enzyme method (Winartasaputra H, et al. Clin Chem 26:613-617, 1980). Specifically, epinephrine-treated adipocytes or an epinephrine-treated adipocyte cell line, for example, 3T3-L1 cells can be monitored for glycerol

15

5 production. Epinephrine treatment of adipocytes leads to triglyceride hydrolysis and the release of glycerol into the cell culture medium.

B. NOV2 -- HUMAN FARNESOID X-ACTIVATED RECEPTOR - LIKE PROTEINS, NUCLEIC ACIDS ENCODING THE SAME & METHODS OF USE THEREOF

10 Cholesterol is an extremely important biological molecule that has roles in membrane structure as well as being a precursor for the synthesis of the steroid hormones and bile acids. Both dietary cholesterol and that synthesized *de novo* are transported through the circulation in lipoprotein particles. The same is true of cholesteryl esters, the form in which cholesterol is stored in cells.

15 The synthesis and utilization of cholesterol must be tightly regulated in order to prevent over-accumulation and abnormal deposition within the body. Of particular importance clinically is the abnormal deposition of cholesterol and cholesterol-rich lipoproteins in the coronary arteries. Such deposition, eventually leading to atherosclerosis, is the leading contributory factor in diseases of the coronary arteries.

20 Human Farnesoid X-activated receptor (FXR) serves as one of a key regulator in cholesterol homeostasis by governing reverse cholesterol transport from peripheral tissues, bile acid synthesis in liver, and cholesterol absorption in intestine (Chiang JY. Bile Acid regulation of gene expression: roles of nuclear hormone receptors. *Endocr Rev* (2002) 23:443-63 PMID: 12202460).

25 Farnesoid X-activated receptor (FXR) functions as a heterodimer with retinoid X receptor (RXR) and regulates genes associated with lipid, bile acid, and cholesterol metabolism. Bile acids, the physiologic ligands for FXR, are the major products of cholesterol catabolism and are essential for the solubilization and transport of dietary lipids. Activation of RXR/FXR heterodimers counteracts the function of LXR (liver X receptor, nuclear hormone receptor involved in lipid metabolism and transport) by repressing CYP7A1 (cytochrome P450, family 27, subfamily A, polypeptide 1, rate-limiting enzyme of bile acid synthesis) and ABC11 (ATP-binding cassette, sub-family B (MDR/TAP), member 11, reverse-cholesterol transporter), and promoting BSEP (major hepatic canalicular bile acid transport protein), and IBABP (fatty and bile acid binding protein 6) expression.

30 FXR null mice have been distinguished from wild type mice by elevated serum bile acid, cholesterol, and triglycerides, increased hepatic cholesterol and triglycerides, and a proatherogenic serum lipoprotein profile. This finding suggests that activation of FXR may be beneficial for lowering cholesterol level and improving lipid profile in general (Sinal CJ, Tohkin M, Miyata M, Ward JM, Lambert G, Gonzalez FJ. Targeted disruption of the nuclear receptor FXR/BAR impairs bile acid and lipid homeostasis. *Cell* (2000) 102:731-44 PMID: 11030617).

40 Elevated cholesterol and triglycerate levels although often associated with later stage of obesity and diabetes are not causative per se to disease development. We are proposing that antagonizing FXR activity may be beneficial for prevention of obesity based on comparison gene dysregulation in normal and obese animals.

45 The present invention further discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with various diseases or pathologies. The proteins and related proteins that are similar to them, are encoded by a cDNA and/or by genomic DNA. The proteins, polypeptides and their cognate nucleic acids were identified by the inventors in certain cases. In

5 particular, FXR is found to be down-regulated in ileum of rats with severe obesity induced by a high fat diet compared to normal weight rats on the same diet. The data show that obesity is associated with FXR down-regulation, and suggest that activation of FXR reverses the obesity phenotype. Furthermore, the inventors found that FXR, its heterodimer partner RXR and its known responsive genes are all expressed in several cell lines that are therefore suitable for screening compounds that
10 modulate FXR activity. A preferred method of the invention is the use of these cell lines expressing endogenous FXR and related genes for identifying an agonist that would be beneficial in the treatment of obesity. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify such therapeutic antibodies and/or therapeutic small molecules.

15 In particular the invention relates to the use of FXR protein as a diagnostic and/or target for small molecule drugs and antibody therapeutics and to the use of cell lines expressing FXR protein to identify small molecule drugs and antibody therapeutics to modulate FXR activity. The inventors have discovered the down-regulation of the FXR in the ileum of obese rats on a high fat diet compared to expression in the ileum of their lean littermates on the same diet. The finding that
20 down-regulation of FXR is associated with obesity indicates that activation of FXR may reverse the obesity phenotype by triggering the expression of FXR-responsive genes.

Therefore activation with an agonist of FXR is useful for the treatment of obesity. To identify the appropriate screening system for compounds that modulate FXR activity, the inventors have assessed the expression of FXR and its heterodimer partner RXR in different cell lines. FXR and
25 RXR are expressed in the following cell lines: renal carcinoma ACHN, TK-10, A498 and 786-0; in non-small cell lung carcinoma A549 and HOP-62; in hepatoma HepG2 cells. Furthermore, that expression of known FXR-responsive genes BSEP (ABCB11, ATP-binding cassette, sub-family B (MDR/TAP), member 11), I-BABP (FABP6, fatty acid binding protein 6), SHP (NR0B2, nuclear receptor subfamily 0, group B, member 2), PLTP (phospholipid transfer protein), PNMT
30 (phosphatidylethanolamine N-methyltransferase) and CPT-II (carnitine palmitoyltransferase II) have been studied in the same cell lines. CPTII and PLTP are expressed in all of the above mentioned cell lines at high to moderate levels; PNMT is expressed in 786-0; FABP6 is expressed at low but significant levels in A498; NR0B2 is expressed in HepG2 and ABCB11 has no expression in these cell lines at all. The cell lines with high endogenous expression of FXR, RXR and one or more FXR-
35 responsive genes are suitable for assaying FXR activity in HTS screening.

In a particular embodiment of the invention, the FXR protein is the target for screening, and in another embodiment of the invention described above cell lines expressing endogenous FXR are each a target for screening. As such the current invention embodies the use of endogenously expressed protein in various screens to identify FXR agonist therapeutic antibodies and/or
40 therapeutic small molecules beneficial in the treatment of obesity.

Thus, FXR nucleic acids and proteins are useful for screening for an activator/agonist of FXR for the treatment of obesity. These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in diagnostic and/or therapeutic methods.

5 Furthermore, our results indicate that a modulator of FXR activity, such as an activator or agonist of FXR may be useful for treatment of such disorders as obesity. Additionally, modulators of FXR could be useful in treatment of diabetes and insulin resistance and for enhancement of insulin secretion.

Discovery Process

10

The following sections describe the study design(s) used to identify the Farnesoid X-activated receptor-encoded protein and any variants, thereof, as being suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for treatment of Obesity and/or Diabetes.

15

Example B1. Rat Dietary-Induced Obesity Fast-Re-feed Study (BP24.06)

20 This study was designed to examine the chronic gene expression changes in response to dietary-induced obesity (DIO), as well as the acute gene expression changes associated with fasting and re-feeding. The sample groups for the study were selected from male Wistar rats and were either chow-fed, or placed on a high fat (45%) diet. The rats on the high-fat diet were further sub-divided into rats resistant to DIO (<1 standard deviation above the weight of chow-fed control rats) and DIO rats (4 standard deviations above the weight of chow-fed control rats. Changes in gene expression in the three sample groups were examined under normal feeding conditions, after 24 hr fasting, and after 24 hr fasting followed by a 4-hr re-feeding period. The clinical data obtained from each animal included body weight, food intake, glucose levels, insulin levels, free fatty acid levels and blood chemistry. A variety of tissues were harvested, including hypothalamus, brainstem, striatum, epididymal fat pads, subcutaneous fat pad, brown adipose tissue (BAT), gastrocnemius muscle (fast twitch skeletal muscle), soleus muscle (slow twitch skeletal muscle), proximal small intestine, distal small intestine, pituitary, kidneys, adrenal gland, and heart. The differential gene expression profiles for these tissues revealed genes and pathways that can be used as therapeutic targets for obesity.

25

30

Results

35 A fragment of the rat Farnesoid X-activated receptor gene was initially found to be down-regulated 1.8 fold in the ileum of obese rats (sd4) on a high fat diet relative to expression in the ileum of rats resistant to obesity (sd1) under normal feeding conditions using CuraGen's GeneCalling® method of differential gene expression (described in Example Q7). A differentially expressed rat gene fragment migrating, at approximately 87 nucleotides in length was definitively identified as a component of the rat Farnesoid X-activated receptor cDNA. The method of competitive PCR was used for confirmation of the gene assessment. The electropherographic peaks corresponding to the gene fragment of the rat Farnesoid X-activated receptor are ablated when a gene-specific primer (shown in Table B1) competes with primers in the linker-adaptors during the PCR amplification. The peaks at 87 nt in length were ablated in the samples from both the sd4 and sd1 rats. Down-regulation of Farnesoid X-activated receptor observed in obese ileum indicates that activation of FXR in obese subjects would be beneficial for reduce or treat the phenotypes indicative of obesity.

45

Table B1. Competitive PCR primer for the rat Farnesoid X-activated receptor: The sequence of the
--

gene fragment (from 1481 to 1567, band size: 87) and the gene-specific primers used for competitive PCR are indicated on the cDNA sequence of the rat Farnesoid X-activated receptor and are shown below in bold. The gene-specific primers at the 5' and 3' ends of the fragment are underlined (gene length is 2443, only region from 1000 to 2047 shown).

1000 GCAGAAGAAA ATTTTCTCAT ATTAACAGAA ATGGCTACCA GTCACGTACA
GATTCTCGTA

1060 GAATTCACAA AAAGACTTCC AGGGTTTCAG ACACTGGACC ACGAAGACCA
GATTGCTTTG

1120 CTCAAAGGGT CCGCAGTCGA GGCCATGTTT CTCGTTTCAG CGGAGATTTT
CAATAAGAAA

1180 CTTCTGCGG GACACGCAGA CCTGTTGGAA GAAAGAATTC GAAAGAGCGG
CATCTCCGAT

1240 GAGTACATAA CCCCAGTGTT TAGTTTCTAT AAAAGTGTCG GGGAGCTGAA
AATGACCCAG

1300 GAAGAGTACG CTCTGCTCAC AGCAATTGTC ATCCTCTCTC CAGACAGACA
ATACATAAAG

1360 GATAGAGAGG CAGTGGAGAA GCTTCAGGAG CCTCTGCTCG ATGTCCTACA
AAACTCTGC

1420 AAGATCTACC AGCCCGAGAA CCCTCAGCAT TTCGCCTGCC TCCTGGGTGCG
CCTGACAGAA

1480 CT**CCGGACAT TCAACCATCA CCACGCTGAG ATGCTGATGT CTTGGAGGGT**
GAATGACCAC

1540 **AAGTTCACCC CGCTCCTCTG TGAGATCT**GG GATGTGCAGT GAAGGACACG
GGGAGGGGCT

1600 AGCTCCTTGT CCTCCTCAGA GCAGCAACCC GGTATTGGAC TTCCCTTCTT
TTCATTTGTA

1660 CCGGGTCTCA CTCAAGAATC TCAATGAATA TTTATGTGGC AATTATACAA
TTCCACAAC

1720 TGTAATACA GGCTCCATAG AATTGCTTCC CCTACACTGT ATTTACAAG
GCTTTGGGAA

1780 ACCCACTGA CACGCCCTTT TTGCCTCATT AAATCAATTG TTA CTTC AAT TTTGTCAACT

1840 GAGCTAGGGA CCGCCTCGTT TTATCCTCCA TGCGGCAACA TTATATATAT
ATATATTTTA

1900 TCAAATAGCT GTTTTCTCTT CCTTTTCTTT TTTTCTTTT TTCGGAGCTG GGGACCGAAC

1960 CCAGGGCCTT GCGCTTGCTA GGCAAGCGCT CTACCACTGA GCTAAATCCC
CAACCCCTTT

2020 TTTATTTCTT TAACAATGCA AAATATAC (SEQ ID NO: 38)

Example B2. Human FXR Sequence Identification

The sequence of Human FXR (Acc. No. CG127635-01) was derived by laboratory screening of cDNA library. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were 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 protocol for identification of human sequence(s) is disclosed in Example Q8.

Table B2 shows an alignment of the protein sequences of the human (CG127635-01) and rat homologs of the Farnesoid X-activated receptor. Table B3 shows amino acid sequence of rat homologs of the Farnesoid X-activated receptor.

Table B2. An alignment (Clustal W) of the protein sequences of the human (CG127635-01; SEQ ID NO: 6) and rat homologs of the Farnesoid X-activated receptor (SEQ ID NO: 39).

CG127635-01-PROT	1	MGSKMNLIEHSHLPTDFEFSFSENLFGLVTEQVAGPLGQNLDEVEPYSOYSNVQFPQVQPO	60
RNU18374_Rat_FXR	1	----MNLIGPSHLQATDEFALSENLFGLVTEHAAGPLGQNLDESYSFYNNVQFPQVQPO	56
CG127635-01-PROT	61	ISSSSYYSNLGFYFQQPEEWYSPGLYELRRMPAETLVQGETEVAEMPVTKKPRMAAS-AG	119
RNU18374_Rat_FXR	57	ISSSSYYSNLGFYFQQPEDWYSPGLYELRRMPTESVYQGETEVS EMPVTKKPRMAASSAG	116
CG127635-01-PROT	120	RIKGD ELCVV CGDRASGYHYNALTCEGCKGFFRRSITKNAVYKCKNGGNCVMDMYMRRKC	179
RNU18374_Rat_FXR	117	RIKGD ELCVV CGDRASGYHYNALTCEGCKGFFRRSITKNAVYKCKNGGNCVMDMYMRRKC	176
CG127635-01-PROT	180	QECRLRKCKEMGMLAECLLTEIQCKSKRLRKNVKQHADQTVNEDSEGRDLRQVTSSTTKSC	239
RNU18374_Rat_FXR	177	QECRLRKCREMGMLAECLLTEIQCKSKRLRKNVKQHADQTVNEDSEGRDLRQVTSSTTKLC	236
CG127635-01-PROT	240	REKTELTVDQQTLLHFIIMDSYNKQRMPOEITNKILKEEFSAEENFLILTEMATNHVQVLV	299
RNU18374_Rat_FXR	237	REKTELTVDQQTLLDYIMDSYSKQRMPOEITNKILKEEFSAEENFLILTEMATSHVQILV	296
CG127635-01-PROT	300	EFTKRLPGFQTLDDHEDQIALLKGS AVEAMFLRSAEIFNKKLPAGHADLLEERIRKSGISD	359
RNU18374_Rat_FXR	297	EFTKRLPGFQTLDDHEDQIALLKGS AVEAMFLRSAEIFNKKLPAGHADLLEERIRKSGISD	356
CG127635-01-PROT	360	EYITPMFSFYKSVGELKMTQEEYALLTAIVILSPDRQYIKDREAVEKLOEPLLDVLQKLC	419
RNU18374_Rat_FXR	357	EYITPMFSFYKSVGELKMTQEEYALLTAIVILSPDRQYIKDREAVEKLOEPLLDVLQKLC	416
CG127635-01-PROT	420	KIHQPENPQHFACLLGRLTELRFTFNHHHA EMLMSWRVNDHKFTPLLCEIWDVQ	472
RNU18374_Rat_FXR	417	KIVQPENPQHFACLLGRLTELRFTFNHHHA EMLMSWRVNDHKFTPLLCEIWDVQ	469

Table B3. Amino acid sequence of rat homologs of the Farnesoid X-activated receptor.

>RNU18374_Rat_FXR

MNLIGPSHLQATDEFALSENLFGLVTEHAAGPLGQNLDESYSFYNNVQFPQVQPOISSSSYYSNLGFYFQQPEDWYSPGLYELRRMPPTESVYQGETEVSEMPVTKKPRMAASSAGRIKGD ELCVV CGDRASGYHYNALTCEGCKGFFRRSITKNAVYKCKNGGNCVMDMYMRRKCQDCRLRKCREMGMLAECLLTEIQCKSKRLRKNVKQHADQTVNEDSEGRDLRQVTSSTTKLCREKTELTVDQQTLLDYIMDSYSKQRMPOEITNKILKEEFSAEENFLILTEMATSHVQILVEFTKRLPGFQTLDDHEDQIALLKGS AVEAMFLRSAEIFNKKLPAGHADLLEERIRKSGISDEYITPMFSFYKSVGELKMTQEEYALLTAIVILSPDRQYIKDREAVEKLOEPLLDVLQKLCCKIYQPENPQHFACLLGRLTELRFTFNHHHA EMLMSWRVNDHKFTPLLCEIWDVQ (SEQ ID NO: 39)

The laboratory cloning was performed using one or more of the methods summarized in Example Q8. The NOV2 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table B4.

Table B4. NOV2 Sequence Analysis			
NOV2a,	CG127635-01	SEQ ID NO: 5	2233 bp
DNA Sequence		ORF Start: ATG at 363	ORF Stop: TGA at 1779

NOV2a,	CG127635-01	SEQ ID NO: 6	472 aa	MW at 54408.7kD
Protein Sequence				
MGSKMNLIEHSHLPTTDEFSENFSLFGVLTEQVAGPLGQNLVEVEPYSQYSNVQFPQVQVQPISSSSYYS NLGFYPQQPEEWYSPGIYELRRMPAETLYQGTEVAEMPVTKKPRMGASAGRIKGDCLCVVCGDRASG YHYNALTCCEGCKGFFRRSITKNAVYCKKNGGNCVMDMYMRKCKQECRLRKCKEMGMLAECLLITEIQCK SKRLRKNVKQHADQTVNEDSEGRDLRQVTSTTKSCREKTELTPDQQTLLHFIIMDSYNKQMPQEIITNK ILKEEFSABENFLILTEMATNHVQVLFVEFTKKLPGFQTLDHEDQIALLKGSAAVEAMFLRSAEIFNKKL PSGHSDLLEERIRNSGISDEYITPMFSFYKSIGELMKTQEEYALLTAVILSPDRQYIKDREAVEKLQ EPLLVLVOKLCKIHOPENPOHFACLLGLRLTELTRTFNNHHAEMLSWRVNDHKFTPLLCEIWDVQ				

Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-C4	125..200	48/77 (62%) 67/77 (87%)	4.9e-44
Hormone_recep	292..468	68/205 (33%) 149/205 (73%)	1.4e-48

10 The protocol for obtainment of gene variants and SNPs is disclosed in Example Q11. The variants of the human FXR were obtained from direct cloning and/or public databases. In addition to the human version of the FXR identified as being differentially expressed in the experimental study, other

5 variants have been identified by direct sequencing of cDNAs derived from many different human tissues and from sequences in public databases. No splice-form variants have been identified at CuraGen whereas several amino acid-changing and non-amino acid-changing cSNPs were identified and are shown in Table B6 below, where UCP=uncharged polar, NP=non polar, A=acidic, and B=basic. Those cSNPs with ID "cgsp" refer to CuraGen proprietary SNPs, whereas those labeled
 10 "hsnp" are from public databases. In a preferred embodiment, CG127635-01 sequence was used for screening purposes.

Table B6.

SNP ID	SNP MATCH POSITION ON DNA	ALLELES	SNP MATCH POSITION ON PROTEIN	ORIG AA	NEW AA	AA CHANGE TYPE
cgsp:cg111.3254	290	C/-	7	Gln	N/A	
cgsp:13381540	443	A/G	58	Leu	Leu	Silent
cgsp:cg111.3253	450	C/-	61	Gly	N/A	
hsnp:rs1049227	456	C/G	63	Gln	Glu	UCP to A
cgsp:cg111.3252	489	G/-	74	His	N/A	
cgsp:cg111.3251	536	G/-	89	Cys	N/A	
hsnp:rs2650007	642	C/T	125	Leu	Phe	Conservative
cgsp:cg111.3255	977	C/-	236	Val	N/A	

15

Example B4. Expression Profile of the Human Farnesoid X-activated receptor Gene (CG127635-01)

To identify cell lines that endogenously express human Farnesoid X-activated receptor,
 20 expression of gene CG127635-01 was assessed using the primer-probe set Ag7502, described in Table B8. Results of the RTQ-PCR assays are shown in Table B16. The expression of known FXR interactors such as retinoid X receptor alpha (RXRA) and small heterodimer partner nuclear receptor subfamily 0, group B, member 2 (NR0B2/SHP) was also assessed. RXR forms a heterodimer with FXR and is necessary for FXR activity, while SHP represses FXR activity. In addition, the expression
 25 of FXR responsive genes carnitine palmitoyltransferase II (CPTII), phospholipid transfer protein (PLTP), phenylethanolamine N-methyltransferase (PNMT), fatty and bile acid binding protein 6 (IBABP/FABP6), and bile salt export pump (ABCB11/BSEP) was assessed. FXR is responsible for transcription activation of PLTP, PNMT, CPTII in liver/liver-related cell lines and for transcription activation of ABCB11 in GI tract.

30 In addition to the results displayed in Table B16, additional experiments were conducted with FXR and shown in Table B17. Table B17 is a summary of the expression levels detected by RTQ-PCR of FXR and FXR-interacting genes in all the cell lines examined. The values are reported as CT values, which represent the cycle at which a given sample crosses a threshold level of fluorescence. Any sample with a reported CT value below 35 is considered to show significant levels of expression
 35 of that gene. The protocol for quantitative expression analysis is disclosed in Example Q9.

5th Table B7.

Gene	Internal identifier or NCBI Reference Sequence	SEQ ID NO	Primer-probe set	Table
FXR	CG127635-01	5	Ag7502	Table B16
RXRA	NM_002957.3	40	Ag7501	Table B17
SHP	NM_021969.1	41	Ag7837	Table B17
CPTII	NM_000098	42	Ag3588	Table B17
PLTP	NM_006227.2	43	Ag7835	Table B17
PNMT	NM_002686.2	44	Ag7838	Table B17
ABCB11	NM_003742.2	45	Ag7836	Table B17
FABP6	NM_001445.1	46	Ag7834	Table AH

Table B8. FXR, Probe Name Ag7502

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ctctcctcctcctcacctcat-3'	21	18	47
Probe	TET-5'-tgtctccccgacttatcctaagcga-3'- TAMRA	26	39	48
Reverse	5'-attttgctacaaatgctcagaatc-3'	24	70	49

Table B9. RXR, Probe Name Ag7501

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gtacagtgcgtgttggttgaaaa-3'	23	3849	50
Probe	FAM-5'-acctctctactcagtgatgagccgca-3'- TAMRA	26	3888	51
Reverse	5'-atactactacaggtctgggttaaatagaaattct- 3'	33	3914	52

Table B10. SHP, Probe Name Ag7837

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ctccacctcaagtcattc-3'	20	663	53
Probe	fam-5'-cttggggacctcttctttcgccctat- 3'-tamra	26	694	54
Reverse	5'-agcgatgtcaacatctccaa-3'	20	722	55

5 Table B11. CPTII, Probe Name Ag3588

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tgaggagagagcgatatttcttc-3'	21	1670	56
Probe	TET-5'-tggaagtacagaatatcccaacatgtctca-3'-TAMRA	30	1639	57
Reverse	5'-acacagacatgctttgtttgc-3'	21	1615	58

Table B12. PLTP, Probe Name Ag7835

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttcccacattcatagcctgtagt-3'	23	1630	59
Probe	FAM-5'-tcaaattcccgtcttctctgtggcact-3'-TAMRA	26	1669	60
Reverse	5'-tgattatggaattaaattgggtacag-3'	26	1697	61

Table B13. PNMT, Probe Name Ag7838

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cctcatcgacattgggttcag-3'	20	222	62
Probe	fam-5'-cctgcagccacttttgaggacatcac-3'-tamra	25	269	63
Reverse	5'-acctccaggaaatctgtcatg-3'	21	294	64

Table B14. BSEP, Probe Name Ag7836

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-agaccagctgtgtttctaactaaatt-3'	27	4581	65
Probe	fam-5'-tgtcttcaagtgaaccagcttcctc-3'-tamra	28	4608	66
Reverse	5'-ttccctatccttagccttagagat-3'	25	4636	67

Table B15. FABP6, Probe Name Ag7834

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tcctcatccctctgctctct-3'	20	9	68
Probe	fam-5'-ctccatctcgaacttgccggtgaaag-3'-tamra	26	54	69
Reverse	5'-catgaactcatcataattcttctcact-3'	27	80	70

10 Table B16. General_screening_panel_v1.7 for FXR

Tissue Name	Rel. Exp.(%) Ag7502	Tissue Name	Rel. Exp.(%) Ag7502
-------------	------------------------	-------------	------------------------

Adipose	0.1	Gastric ca. (liver met.) NCI-N87	0.0
HUVEC	0.0	Stomach	0.0
Melanoma* Hs688(A).T	0.0	Colon ca. SW-948	0.0
Melanoma* Hs688(B).T	0.0	Colon ca. SW480	0.0
Melanoma (met) SK-MEL-5	0.0	Colon ca. (SW480 met) SW620	0.0
Testis	0.5	Colon ca. HT29	0.0
Prostate ca. (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate ca. DU145	0.0	Colon cancer tissue	0.2
Prostate pool	0.5	Colon ca. SW1116	0.0
Uterus pool	0.0	Colon ca. Colo-205	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. (ascites) SK-OV-3	0.2	Colon	0.2
Ovarian ca. OVCAR-4	0.0	Small Intestine	0.2
Ovarian ca. OVCAR-5	0.2	Fetal Heart	0.2
Ovarian ca. IGROV-1	0.0	Heart	0.0
Ovarian ca. OVCAR-8	0.1	Lymph Node pool 1	0.1
Ovary	18.2	Lymph Node pool 2	1.4
Breast ca. MCF-7	0.0	Fetal Skeletal Muscle	0.0
Breast ca. MDA-MB-231	0.0	Skeletal Muscle pool	0.1
Breast ca. BT-549	0.0	Skeletal Muscle	0.3
Breast ca. T47D	0.0	Spleen	4.7
Breast pool	0.0	Thymus	0.2
Trachea	0.3	CNS cancer (glio/astro) SF-268	0.0
Lung	0.8	CNS cancer (glio/astro) T98G	0.0
Fetal Lung	4.2	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	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. NCI-H23	0.0	Brain (Amygdala)	0.4
Lung ca. NCI-H460	0.0	Brain (Cerebellum)	0.0
Lung ca. HOP-62	100.0	Brain (Fetal)	0.0
Lung ca. NCI-H522	0.0	Brain (Hippocampus)	0.5
Lung ca. DMS-114	0.0	Cerebral Cortex pool	0.1
Liver	40.1	Brain (Substantia nigra)	0.1

Fetal Liver	77.4	Brain (Thalamus)	0.1
Kidney pool	5.6	Brain (Whole)	0.3
Fetal Kidney	1.1	Spinal Cord	0.0
Renal ca. 786-0	0.0	Adrenal Gland	75.8
Renal ca. A498	19.9	Pituitary Gland	5.4
Renal ca. ACHN	6.7	Salivary Gland	0.0
Renal ca. UO-31	1.3	Thyroid	0.1
Renal ca. TK-10	1.0	Pancreatic ca. PANC-1	0.0
Bladder	1.3	Pancreas pool	0.6

5

Expression Data Summary: RTQ-PCR data shows that Farnesoid X-activated receptor is expressed in liver, kidney, adrenal gland, which is in agreement with published data. Expression is also detected in ovary, kidney, and fetal lung. RTQ-PCR does not show expression in small intestine, but expression of this gene in small intestine has been demonstrated both in the literature

10

and the GeneCalling® experiment described in Example B1 above.

Table B17.

Summary of expression of FXR and related genes in different cell lines, reported as CT values

Cell lines	Gene Expression							
	FXR	RXR	CPTII	PLTP	PNMT	FABP6	NR0B2	ABCB11
Colon ca. SW-948	40.0	32.0	34.4	28.3	40.0	34.4	32.1	38.3
Breast ca. MCF-7	40.0	29.2	28.8	30.8	37.7	35.6	35.7	40.0
Breast ca. MDA-MB-231	40.0	29.5	30.3	28.6	34	38.7	N/A	40.0
Breast ca. BT-549	40.0	30.4	29.7	24.9	40.0	40.0	36.1	40.0
Breast ca. T47D	40.0	31.1	N/A	N/A	N/A	N/A	N/A	N/A
Breast ca. MDA-N	40.0	32.5	32.3	N/A	N/A	N/A	N/A	N/A
Ovarian ca. OVCAR-3	40.0	30.4	30.2	26.0	36.1	36.9	34.5	38.7
Ovarian ca. SK-OV-3	40.0	30.0	29.8	36.2	40.0	40.0	36.1	40.0
Ovarian ca. OVCAR-4	40.0	33.0	32.1	25.9	36.3	32.3	32.5	40.0
OVCAR-5_Ovarian carcinoma	34.67	28.7	28.3	28.8	35.5	36.4	N/A	33.4
Ovarian ca. IGROV-1	40.0	31.5	29.8	30.6	37.2	35.9	29.1	40.0

Ovarian ca. OVCAR-8	37.27	31.3	30.3	25.8	33.6	37.7	N/A	40.0
CNS cancer (glio/astro) U87-MG	40.0	30.7	30.6	N/A	40.0	39.7	35.8	40.0
KATO III_Stomach ca,	40.0	29.5	31.2	N/A	33.7	35.9	36.9	34.7
CNS cancer (glio/astro) U-118-MG	37.14	30.1	30.1	N/A	40.0	40.0	34.5	35.7
CNS cancer (neuro;met) SK-N-AS	40.0	31.5	31.2	30.7	37.8	40.0	40.0	40.0
CNS cancer (astro) SF- 539	40.0	30.0	N/A	32.0	35.6	39.0	35.5	37.5
CNS cancer (astro) SNB- 75	37.71	28.9	28.1	25.8	34.9	40.0	38.0	40.0
CNS cancer (glio) SNB- 19	40.0	31.9	29.5	32.6	34.4	37.6	36.6	31.6
NCI- N417_Small cell lung cancer/neuro endocrine	40.0	32.0	35.8	27.5	40.0	40.0	40.0	40.0
CNS cancer (glio) SF-295	40.0	30.2	27.3	33.5	34.1	40.0	36.8	40.0
Colon ca. SW480	40.0	29.4	31.3	32.7	40.0	40.0	40.0	40.0
Colon ca. (SW480 met) SW620	40.0	31.5	31.1	26.9	27.5	34.6	30.4	40.0
Colon ca. HT29	36.81	30.2	32.2	26.2	33.0	35.6	33.1	35.6
Colon ca. HCT-116	40.0	30.1	31.7	29.8	35.4	33.2	33.4	40.0
Colon ca. CaCo-2	34.83	29.4	31.9	N/A	N/A	N/A	N/A	N/A
NCI- N87_Gastric carcinoma	37.19	28.3	27.1	32.6	33.8	40.0	40.0	40.0
Colon cancer tissue	35.04	30.9	31.1	28.5	40.0	39.2	31.7	40.0
Colon ca. SW1116	40.0	32.3	31.7	34.6	31.8	40.0	39.8	38.1
Renal ca.	31.54	29.6	29.2	31.3	28.0	37.2	32.5	40.0

786-0								
Renal ca. A498	30.82	31.7	30.2	30.5	40.0	33.4	35.8	40.0
Colon ca. Colo-205	40.0	31.8	34.0	34.1	33.8	35.8	34.0	37.5
Renal ca. ACHN	30.57	32.8	30.0	25.5	40.0	34.9	33.9	40.0
Renal ca. UO-31	34.45	32.2	30.3	25.6	37.9	37.4	29.9	40.0
Renal ca. TK-10	30.74	29.6	29.2	26.1	38.4	37.1	N/A	37.8
Liver_HepG2	31.13	30.8	31.1	29.5	40	40	26.3	40
LX-1_Small cell lung cancer	38.21	31.2	30.0	32.1	30.7	36.2	32.8	38.6
Lung ca. NCI-H146	40.0	40.0	32.0	N/A	N/A	N/A	N/A	N/A
Lung ca. SHP-77	38.13	33.2	32.3	27.4	31.9	34.9	25.1	40.0
Lung ca. A549	32.21	31.1	30.7	N/A	35	40	38.5	36.6
Lung ca. NCI-H526	40.0	33.7	33.5	N/A	N/A	N/A	N/A	N/A
SW-48 Colon adenocarcino ma	40.0	32.3	32.4	33.2	35.9	34.7	31.7	38
Lung ca. NCI-H23	40.0	31.4	28.4	23.6	32.8	38.3	32.8	40.0
Lung ca. NCI-H460	40.0	30.4	29.1	26.8	36.5	35.3	37.3	35.5
Lung ca. HOP-62	33.28	32.0	29.5	29.6	34.2	36.9	35.2	38.7
Lung ca. NCI-H522	40.0	31.2	29.3	N/A	N/A	N/A	N/A	N/A
Pancreatic ca. CAPAN2	40.0	30.2	29.5	N/A	N/A	N/A	N/A	N/A
Prostate ca. (bone met) PC-3	38.09	31.5	29.6	33.8	40.0	40.0	40.0	40.0
Melanoma Hs688(A).T	40.0	31.1	29.2	39.4	40.0	40.0	40.0	40.0
Melanoma Hs688(B).T	40.0	31.4	29.4	24.6	33.9	33.3	N/A	35.9
Squamous cell carcinoma SCC-4	40.0	31	32.8	N/A	N/A	N/A	N/A	N/A
Melanoma M14	40.0	31.8	29.2	N/A	N/A	N/A	N/A	N/A

Melanoma LOXIMVI	40.0	33.4	34.1	N/A	N/A	40.0	N/A	N/A
Melanoma (met) SK- MEL-5	40.0	32.4	31.0	27.3	40.0	40.0	33.5	40.0

5 **Expression Data Summary:** RTQ-PCR data show that FXR and RXR are both expressed in the following cell lines: in renal carcinoma cell lines ACNH, TK-10, A498, and 786-0; in non-small cell lung carcinoma A549 and HOP-62, and in hepatoma HepG2 cells. Endogenous expression of FXR in these cell lines is confirmed by the observed expression of the FXR responsive genes: CPTII and

10 PLTP are expressed in all of the above mentioned cell lines with high to moderate levels; PNMT is expressed in 786-0; FABP6 is expressed (low level) in A498; NR0B2 is expressed in HepG2.

Based on high endogenous expression of FXR, RXR and one or more FXR-responsive genes, all the cell lines mentioned above are suitable for HTS screening for an agonist for human Farnesoid X-activated receptor.

15 **Example B5. Screening Assay Formulation**

Assays for screening for antibody therapeutics or small molecule drugs targeting endogenously expressed Human Farnesoid X-activated receptor can be formulated utilizing the non-

20 exhaustive list of cell lines that express the Farnesoid X-activated receptor from the RTQ-PCR results shown above.

C. NOV3 -- Human P2Y1 receptor short form - like Proteins, Nucleic Acids Encoding the Same & Methods of Use Thereof

P2 receptors are divided into two main classes based on whether they are ligand-gated ion

25 channels (P2X receptors) or are coupled to G proteins (P2Y receptors). P2Y receptors are a family of G-protein coupled receptors localized to the plasma membrane. To date, at least five distinct P2Y receptors subtypes have been cloned and characterized pharmacologically. The cloned P2Y1 receptor (P2Y1R), and its endogenous (physiologic) counterpart termed P2Y1-like is a receptor for the endogenous ligands ADP, ATP and certain diadenosine phosphates. Sensitivity to ATP seems to

30 vary with cell type; many P2Y1 and P2Y1-like receptors are relatively insensitive to ATP, but are strongly activated by ADP. Characteristically, among all P2Y subtypes, the P2Y1R and its endogenous counterpart are strongly activated by 2MeSATP, ADP, ADP β S, and ADP β F (Ralevic V, Burnstock G., 1998, Receptors for purines and pyrimidines. Pharmacol Rev. 50(3):413-92. Review. PMID: 9755289). Thus, cellular responses to these reagents can be used to infer, but not prove, the

35 presence of P2Y1R at the cell membrane. Agonist stimulation of P2Y1R leads to Gq/11-mediated phospholipase C activation, formation of inositol 1,4,5-triphosphate and mobilization of intracellular calcium from internal stores.

P2Y1R is widely expressed in human tissues including pancreas (Moore DJ, Chambers JK, Wahlin JP, Tan KB, Moore GB, Jenkins O, Emson PC, Murdock PR., 200, Expression pattern of

40 human P2Y receptor subtypes: a quantitative reverse transcription-polymerase chain reaction study. Biochim Biophys Acta. 2001 Oct 31;1521(1-3):107-19. PMID: 11690642). However, the pancreas is

5 comprised both of exocrine cells and endocrine islet cells. Because islet cells constitute only 2% of cells present in the pancreas, pancreatic expression of P2Y1 R does not constitute proof of P2Y1R expression on islet cells. The present invention provides data for human islet cell expression of P2Y1R.

10 A role for P2Y receptors in insulin secretion has been posited based on data from rat islets (Petit P, Hillaire-Buys D, Manteghetti M, Debrus S, Chapal J, Loubatieres-Mariani MM, 1998, Evidence for two different types of P2 receptors stimulating insulin secretion from pancreatic cell. *Br J Pharmacol.* 125:1368-74. PMID: 9863669; Verspohl EJ, Johannwille B, Waheed A, Neye H., 2002, Effect of purinergic agonists and antagonists on insulin secretion from INS-1 cells and rat pancreatic islets. *Can J Physiol Pharmacol.* 80(6):562-8. PMID: 12117305) and human islets (Fernandez-Alvarez J, Hillaire-Buys D, Loubatieres-Mariani MM, Gomis R, Petit P. 2001, P2 receptor agonists stimulate insulin release from human pancreatic islets. *Pancreas.* 22:69-71. PMID: 11138974). Pancreatic islet cells secrete insulin in response to a threshold level of glucose, or to pharmacologic reagents that mimic the actions of glucose. The P2Y1 receptors are strongly activated by the stable ATP analog ADP β S (Ralevic V, Burnstock G., 1998, Receptors for purines and pyrimidines. *Pharmacol Rev.* 20 50(3):413-92. Review. PMID: 9755289). In isolated rat islets, glucose-stimulated insulin secretion is potentiated by ADP β (Petit P, Hillaire-Buys D, Manteghetti M, Debrus S, Chapal J, Loubatieres-Mariani MM, 1998, Evidence for two different types of P2 receptors stimulating insulin secretion from pancreatic B cell. *Br J Pharmacol.* 125:1368-74. PMID: 9863669; Verspohl EJ, Johannwille B, Waheed A, Neye H., 2002, Effect of purinergic agonists and antagonists on insulin secretion from INS-1 cells and rat pancreatic islets. *Can J Physiol Pharmacol.* 80:562-8. PMID: 12117305). In isolated human islets, ADP β also potentiates glucose-stimulated insulin secretion (Fernandez-Alvarez J, Hillaire-Buys D, Loubatieres-Mariani MM, Gomis R, Petit P. 2001, P2 receptor agonists stimulate insulin release from human pancreatic islets. *Pancreas.* 22:69-71. PMID: 11138974). The combined data suggest, based on insulin secretion in response to ADP β S, that islet cells may 25 express P2Y1R at the cell membrane. However, proof of pharmacologic specificity requires confirmation with both selective agonists and antagonists, and there are no such reports in the literature.

The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with various diseases or pathologies. The proteins and related 35 proteins that are similar to them, are encoded by a cDNA and/or by genomic DNA. The proteins, polypeptides and their cognate nucleic acids were identified by the inventors in certain cases. In particular, the P2Y1R protein encoded by CG197218-01 and any variants, thereof, are suitable as diagnostic markers, targets for an antibody therapeutic and targets for small molecule drugs. At least five distinct P2Y receptors subtypes have been cloned and characterized pharmacologically (Ralevic V, Burnstock G., 1998, Receptors for purines and pyrimidines. *Pharmacol Rev.* 40 50:413-92. Review. PMID: 9755289). To date there are no agonists or antagonists that discriminate adequately between families of P2X and P2Y receptors, or between subtypes of receptors within each of these groups (Ralevic V, Burnstock G., 1998, Receptors for purines and pyrimidines. *Pharmacol Rev.* 50:413-92. Review. PMID: 9755289). The lack of specific pharmacologic reagents makes it a considerable 45 challenge to dissect out and characterize endogenous P2 receptors in different biologic systems, and

5 even more of a challenge to identify for each of these a physiological or pathophysiological role. The molecular presence of P2Y1R on human islet cells has never been documented in the literature. In the present application, the presence of P2Y1R on human islet cells has been demonstrated in samples from two different patients, supporting a role for this receptor in insulin secretion. An agonist of P2Y1R will be a treatment to augment insulin secretion in Type 2 diabetes.

10 The present invention describes the specific molecular characterization of a P2Y1 receptor on human islet cells and the use of an agonist for that receptor as a treatment for the augmentation of insulin secretion in Type 2 diabetes. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify such therapeutic antibodies and/or therapeutic small molecules.

15 In particular the invention relates to the use of P2Y1R protein, gene or antibody as a target in diagnostic applications and/or target for small molecule drugs and antibody therapeutics. The inventors have shown the expression of the P2Y1R gene on human islets, supporting a role for this receptor in insulin secretion. An agonist of P2Y1R will be a treatment to augment insulin secretion in Type 2 diabetes. The present invention describes the specific molecular characterization of a P2Y1
20 receptor (P2Y1R) on human islet cells and the use of an agonist for that receptor as a treatment for the augmentation of insulin secretion in Type 2 diabetes. In a particular embodiment of the invention, P2Y1R is a target for screening. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify P2Y1R agonist therapeutic antibodies and/or therapeutic small molecules beneficial in the treatment of Type 2
25 diabetes.

Not to be limited by a particular mechanism of action, the inventor nevertheless has discovered that activation of P2Y1R may have beneficial effects for treating diabetes by acting in many metabolic tissues, including pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Specifically, in the islet cells, P2Y1R activation
30 leads to an increase in insulin secretion that may modulate hyperglycemia in type II diabetes. Therefore, an agonist of P2Y1R is useful for the treatment of diabetes.

Discovery Process

35 The following sections describe the study design(s) used to identify the P2Y1R-encoded protein and any variants, thereof, as being suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for treatment of Obesity and/or Diabetes.

Example C1. Human CG197218-01 Sequence Identification

40 The sequence of Human P2Y1R (Acc. No. CG197218-01) was derived by laboratory screening of cDNA library. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were sequenced. In silico prediction was based on sequences available in CuraGen Corporation's proprietary sequence databases or in the public human sequence
45 databases, and provided either the full-length DNA sequence, or some portion thereof. The protocol for identification of human sequence(s) is disclosed in Example Q8.

- 5 An alignment of the protein sequences of the human (CG197218-01), rat and mouse homologs of the P2Y1R is shown in Table C1.

Table C1. Protein alignment (ClustalW) of the human (CG197218-01; U42029_Hs_P2Y1R_short) (SEQ ID NO: 10), mouse (SEQ ID NO: 71) and rat homologs (SEQ ID NO: 72) of the P2Y1R protein.

P2Y1_R_rat_P49651	1	MTEVPWSAVPNGTDAAFLAGLGLSLWGNSTIASTAAVSSSFRCALIKTGFQFYLLPAVYIL	60
NM_008772_Mm_P2Y1_R	1	MTEVPWSAVPNGTDAAFLAGLGLSLWGNSTVASTAAVSSSFQCALTKTGFQFYLLPAVYIL	60
U42029_Hs_P2Y1R_short	1	MTEVLPWPAVPNGTDAAFLAGPSSWGNSTVASTAAVSSSFKCALTKTGFQFYLLPAVYIL	60
P2Y1_R_rat_P49651	61	VFIIGFLGNSVAIWMFVFMKPWSGISVYMFNLALADFLYVLTLPALIFYFYNKTDWIFG	120
NM_008772_Mm_P2Y1_R	61	VFIIGFLGNSVAIWMFVFMKPWSGISVYMFNLALADFLYVLTLPALIFYFYNKTDWIFG	120
U42029_Hs_P2Y1R_short	61	VFIIGFLGNSVAIWMFVFMKPWSGISVYMFNLALADFLYVLTLPALIFYFYNKTDWIFG	120
P2Y1_R_rat_P49651	121	DVMCKLQRFIFHVNLYGSLFLTCISAHRYSGVVYPLKSLGRLKKKNAIYVSVLWVLIIV	180
NM_008772_Mm_P2Y1_R	121	DAMCKLQRFIFHVNLYGSLFLTCISAHRYSGVVYPLKSLGRLKKKNAIYVSVLWVLIIV	180
U42029_Hs_P2Y1R_short	121	DAMCKLQRFIFHVNLYGSLFLTCISAHRYSGVVYPLKSLGRLKKKNAICTSVLWVLIIV	180
P2Y1_R_rat_P49651	181	VAISPILFYSGTGIKKNKTVTCTDSTSEYLRSYFIYSMCTTVAMFCIPLVLILGCYGLI	240
NM_008772_Mm_P2Y1_R	181	VAISPILFYSGTGTKKNKTVTCTDSTSDYLRSYFIYSMCTTVAMFCIPLVLILGCYGLI	240
U42029_Hs_P2Y1R_short	181	VAISPILFYSGTGVKKNKTITCTDSTSEYLRSYFIYSMCTTVAMFCIPLVLILGCYGLI	240
P2Y1_R_rat_P49651	241	VREALIYKDLDNSPLRRKSIYLVII VLTVFAVSYPFHVMKTMNLRARLDFQTPMCD FND	300
NM_008772_Mm_P2Y1_R	241	VREALIYNLDLNSPLRRKSIYLVII VLTVFAVSYPFHVMKTMNLRARLDFQTPMCD FND	300
U42029_Hs_P2Y1R_short	241	VREALIYKDLDNSPLRRKSIYLVII VLTVFAVSYPFHVMKTMNLRARLDFQTPMCA FND	300
P2Y1_R_rat_P49651	301	RVYATYQVTRGLASLNSCVDPILYFLAGDTFRRLSRATRKASRRSEANLQSKSE EMTLN	360
NM_008772_Mm_P2Y1_R	301	RVYATYQVTRGLASLNSCVDPILYFLAGDTFRRLSRATRKASRRSEANLQSKSE EMTLN	360
U42029_Hs_P2Y1R_short	301	RVYATYQVTRGLASLNSCVDPILYFLAGDTFRRLSRATRKASRRSEANLQSKSE DMTLN	360
P2Y1_R_rat_P49651	361	ILSEFKQNGDTSI	373
NM_008772_Mm_P2Y1_R	361	ILSEFKQNGDTSI	373
U42029_Hs_P2Y1R_short	361	ILPEFKQNGDTSI	373

- 10 The laboratory cloning was performed using one or more of the methods summarized in Example Q8. The NOV3 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table C2.

Table C2. NOV3 Sequence Analysis			
NOV3a,	CG197218-01	SEQ ID NO: 9	1410 bp
DNA Sequence		ORF Start: ATG at 47	ORF Stop: TGA at 1166
CCGCCTCTACCCCTCGGAGCCGCCGCTAAGTCGAGGAGGAGAGAAATGACCGAGGTGCTGTGGCCGG CTGTCCCCAACGGGACGGACGCTGCCTTCCTGGCCGGTCCGGGTTTCGTCTCTGGGGGAACAGCACGGTC GCCTCCACTGCCGCCGTCTCCTCGTCGTTCAAATGCGCCTTGACCAAGACGGGCTTCCAGTTTACTA CCTGCCGGCTGTCTACATCTTGGTATTCATCATCGGCTTCTCGGCAACAGCGTGGCCATCTGGATGT TCGTCTCCACATGAAGCCCTGGAGCGGCATCTCCGTGTACATGTTCAATTTGGCTCTGGCCGACTTC TTGTAGCTGCTGACTCTGCCAGCCCTGATCTTCTACTACTTCAATAAAACAGACTGGATCTTCGGGGA TGCCATGTGTAAACTGCAGAGGTTTCATCTTTCATGTGAACCTCTATGGCAGCATCTTGTTCCTGACAT GCATCAGTGCCACCGGTACAGCGGTGTGGTGTACCCCTCAAGTCCCTGGGCGGGCTCAAAAAGAAG AATGCGATCTGTATCAGCGTGTGGTGTGGCTCATTTGTGGTGGTGGCGATCTCCCCATCCTCTTCTA CTCAGGTACCGGGGTCCGCAAAAACAAACCATCACCTGTTTACGACACCACCTCAGACGAGTACCTGC GAAGTTATTTTCATCTACAGCATGTGCACGACCGTGGCCATGTTCTGTGTCCCCTTGGTGTGATCTGT GGCTGTTACGGATTAATTGTGAGAGCTTTGATTTACAAAGATCTGGACAACCTCTCCTCTGAGGAGAAA ATCGATTTACCTGGTAATCATTTGACTGACTGTTTTTGTGCTGTGTCTTACATCCCTTTCCATGTGATGA AAACGATGAACCTTGAGGGCCGGCTTGATTTTCAGACCCAGCAATGTGTGCTTTCAATGACAGGGTT TATGCCACGTATCAGGTGACAAGAGGTCTAGCAAGTCTCAACAGTTGTGTGGACCCATTCTCTATTT CTTGGCGGGAGATACTTTTCAAGAGGAGACTCTCCCGAGCCACAAGGAAAGCTTCTAGAAGAAGTGAGG CAAATTTGCAATCCAAGAGTGAAGACATGACCCCTCAATATTTTACCTGAGTTCAAGCAGAATGGAGAT ACAAGCCTGTGAAGGCACAAGAATCTCCAAACACCTCTCTGTTGTAATATGGTAGGATGCTTAAACAGA ATCAAGTACTTTTCCCTCTTTAACTTTTCTAGTTTGAAGAAAAAATCAAACCAAGAAAAATAGTGAGTTA AAAAAATAATAGAAGTAGAAATGCCACATCCACACTTAGCTTGTGTTGGGTTTGCTTTCACAGTCTCT CTTCCTTCTGACTAGAAGTATGTATAATAAAACAATACTACCTAGTTAAA			
NOV3a,	CG197218-01	SEQ ID NO: 10	373 aa MW at 42071.0kD

Protein Sequence
MTEVLWPAVPNGTDAAFLAGPGSSWGNSTVASTAAVSSSFKCALTKTGFQFYLPVAVYILVFIIGFLG NSVAIWMFVFHMKPWSGISVYMFNLALADFLYVLTLPALIFYFYNKTDWIFGDAMCKLQRFIFHVNLY GSILFLTCSAHRYSGVVYPLKSLGRLKKKNAICISVLVWLIVVVAISPILFYSGTGVVRKNKTITCYD TTSDEYLRSYFIYSMCTTVAMFCVPLVLILGCYGLIVRALIYKDLDNSPLRRKSIYLVIIIVLTVFAVS YIPFHVMTMNLRLRDLFQTPAMCAFNDRVYATYQVTRGLASLNSCVDPILYFLAGDTFRRRLSRATR KASRRSEANLQSKSEDMTLNILPEFKQNGDTSL

- 5 Pfam analysis predicts that the NOV3a protein contains the domains shown in the Table C3.

Table C3. Domain Analysis of NOV3a			
Pfam Domain	NOV3a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_1	68..324	82/279 (29%) 184/279 (66%)	8.9e-51

5

Example C2. Expression Profile of the Human P2Y1R Gene (CG197218-01)

The protocol for quantitative expression analysis is disclosed in Example Q9.

- 10 Expression of gene CG197218-01 was assessed using the primer-probe sets Ag7643 and Ag7643, described in Tables C4 and C5. Results of the RTQ-PCR runs are shown in Tables C6, C7 and C8.

Table C4. Probe Name Ag7643

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aaagaagaatgcgatctgtatcag-3'	24	538	73
Probe	TET-5'-accacaatgagccacaccagcac-3'- TAMRA	23	563	74
Reverse	5'-accccggtacctgagtagaag-3'	21	607	75

Table C5. Probe Name Ag7643

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aaagaagaatgcgatctgtatcag-3'	24	538	76
Probe	TET-5'-accacaatgagccacaccagcac-3'- TAMRA	23	563	77
Reverse	5'-accccggtacctgagtagaag-3'	21	607	78

Table C6. General_screening_panel_v1.7

Column A - Rel. Exp.(%) Ag7643, Run 318601495					
Column B - Rel. Exp.(%) Ag7643, Run 318841727					
Tissue Name	A	B	Tissue Name	A	B
Adipose	15.9	16.0	Gastric ca. (liver met.) NCI-N87	0.0	0.0
HUVEC	43.5	38.7	Stomach	3.3	4.0
Melanoma* Hs688(A).T	0.0	0.0	Colon ca. SW-948	16.6	17.9
Melanoma* Hs688(B).T	2.8	2.7	Colon ca. SW480	0.0	0.0
Melanoma (met) SK-MEL-5	28.5	33.9	Colon ca. (SW480 met) SW620	0.0	0.0
Testis	8.2	10.7	Colon ca. HT29		0.7
Prostate ca. (bone met) PC-3	0.0	0.0	Colon ca. HCT-116	22.5	29.7
Prostate ca. DU145	25.9	28.9	Colon cancer tissue		0.4
Prostate pool	35.1	37.1	Colon ca. SW1116	0.0	0.0
Uterus pool	6.3	7.7	Colon ca. Colo-205	0.1	0.1
Ovarian ca. OVCAR-3		0.5	Colon ca. SW-48	0.6	0.4
Ovarian ca. (ascites) SK-OV-3	0.0	0.0	Colon	75.8	80.1
Ovarian ca. OVCAR-4	0.2	0.2	Small Intestine	23.8	30.1
Ovarian ca. OVCAR-5	13.2	18.7	Fetal Heart	5.4	8.1
Ovarian ca. IGROV-1	0.5	0.5	Heart	11.8	16.5

Ovarian ca. OVCAR-8	11.6	6.2	Lymph Node pool 1	10.5	18.6
Ovary	59.5	66.9	Lymph Node pool 2	17.2	15.6
Breast ca. MCF-7	0.1	0.2	Fetal Skeletal Muscle	7.3	7.0
Breast ca. MDA-MB-231		0.5	Skeletal Muscle pool	5.7	12.6
Breast ca. BT-549	0.1	0.0	Skeletal Muscle	32.5	30.6
Breast ca. T47D	0.0	0.0	Spleen	6.6	7.5
Breast pool	0.0	13.3	Thymus	7.2	10.7
Trachea	35.6	48.6	CNS cancer (glio/astro) SF-268	0.1	0.3
Lung	100.0	100.0	CNS cancer (glio/astro) T98G	0.0	0.0
Fetal Lung	78.5	81.2	CNS cancer (neuro;met) SK-N-AS	0.0	0.0
Lung ca. NCI-N417		0.1	CNS cancer (astro) SF-539	28.3	31.6
Lung ca. LX-1		0.3	CNS cancer (astro) SNB-75	15.7	14.5
Lung ca. NCI-H146	1.0	0.7	CNS cancer (glio) SNB-19	6.3	6.6
Lung ca. SHP-77	0.0	0.0	CNS cancer (glio) SF-295	0.0	0.0
Lung ca. NCI-H23	4.1	7.4	Brain (Amygdala)	19.6	27.7
Lung ca. NCI-H460	18.9	28.3	Brain (Cerebellum)	28.3	29.9
Lung ca. HOP-62	5.1	7.2	Brain (Fetal)	34.2	35.1
Lung ca. NCI-H522	0.7	2.7	Brain (Hippocampus)	31.2	39.2
Lung ca. DMS-114	0.0	0.2	Cerebral Cortex pool	27.5	21.8
Liver	1.8	2.3	Brain (Substantia nigra)	11.2	16.2
Fetal Liver	6.0	10.7	Brain (Thalamus)	28.3	32.5
Kidney pool	12.7	11.3	Brain (Whole)	70.7	88.9
Fetal Kidney	2.3	3.5	Spinal Cord	14.3	13.3
Renal ca. 786-0	4.8	5.8	Adrenal Gland	32.1	12.8
Renal ca. A498	4.9	7.0	Pituitary Gland	11.8	9.5
Renal ca. ACHN	0.0	0.0	Salivary Gland	1.6	2.3
Renal ca. UO-31	2.6	4.3	Thyroid	6.9	8.5
Renal ca. TK-10	0.0	0.0	Pancreatic ca. PANC-1	0.0	0.2
Bladder	15.1	16.8	Pancreas pool	5.7	4.4

5 Table C7. Panel 5 Islet

Column A - Rel. Exp.(%) Ag7643, Run 311303517					
Column B - Rel. Exp.(%) Ag7643, Run 312303208					
Tissue Name	A	B	Tissue Name	A	B
97457_Patient-02go_adipose	12.9	11.3	94709_Donor 2 AM - A_adipose	0.0	0.0
97476_Patient-07sk_skeletal muscle	0.0	0.0	94710_Donor 2 AM - B_adipose	0.0	0.0
97477_Patient-07ut_uterus	42.9	24.7	94711_Donor 2 AM - C_adipose	0.0	0.0
97478_Patient-07pl_placenta	51.1	39.5	94712_Donor 2 AD - A_adipose	0.0	0.0
99167_Bayer Patient 1	15.8	10.5	94713_Donor 2 AD - B_adipose	1.0	0.9
97482_Patient-08ut_uterus	7.4	6.1	94714_Donor 2 AD - C_adipose	0.0	0.4
97483_Patient-08pl_placenta	47.0	40.6	94742_Donor 3 U - A_Mesenchymal	0.0	0.0

			Stem Cells		
97486_Patient-09sk_skeletal muscle	1.7	1.5	94743_Donor 3 U - B_Mesenchymal Stem Cells	1.5	0.0
97487_Patient-09ut_uterus	78.5	79.0	94730_Donor 3 AM - A_adipose	0.0	0.0
97488_Patient-09pl_placenta	48.3	49.3	94731_Donor 3 AM - B_adipose	0.0	0.4
97492_Patient-10ut_uterus	21.3	20.2	94732_Donor 3 AM - C_adipose	0.0	0.0
97493_Patient-10pl_placenta	66.0	48.6	94733_Donor 3 AD - A_adipose	0.8	0.0
97495_Patient-11go_adipose	11.0	7.1	94734_Donor 3 AD - B_adipose	2.5	0.0
97496_Patient-11sk_skeletal muscle	3.1	5.4	94735_Donor 3 AD - C_adipose	0.0	0.0
97497_Patient-11ut_uterus	98.6	57.8	77138_Liver_HepG2untreated	0.8	0.0
97498_Patient-11pl_placenta	42.6	36.9	73556_Heart_Cardiac stromal cells (primary)	0.0	0.0
97500_Patient-12go_adipose	13.4	9.4	81735_Small Intestine	63.7	32.1
97501_Patient-12sk_skeletal muscle	22.5	17.8	72409_Kidney_Proximal Convoluted Tubule	0.0	0.0
97502_Patient-12ut_uterus	99.3	53.2	82685_Small intestine_Duodenum	46.3	6.7
97503_Patient-12pl_placenta	100.0	100.0	90650_Adrenal_Adrenocortical adenoma	0.0	1.2
94721_Donor 2 U A_Mesenchymal Stem Cells	0.0	0.0	72410_Kidney_HRCE	2.8	1.3
94722_Donor 2 U B_Mesenchymal Stem Cells	0.0	0.0	72411_Kidney_HRE	0.0	0.6
94723_Donor 2 U C_Mesenchymal Stem Cells	0.0	0.0	73139_Uterus_Uterine smooth muscle cells	3.3	1.5

5 Table C8. Human Metabolic

Column A - Rel. Exp.(%) Ag7643, Run 323695062					
Column B - Rel. Exp.(%) Ag7643, Run 388353499					
Tissue Name	A	B	Tissue Name	A	B
137857_psoas-AA.M.Diab.-hi BMI-6	0.6	0.4	139523_pancreas-HLM.Norm-hi BMI-31	3.5	6.4
135760_psoas-HLM.Diab.-hi BMI-21	1.9	2.6	139520_pancreas-CC.M.Norm-hi BMI-29	3.5	5.0
134827_psoas-CC.M.Diab.-hi BMI-4	0.2	0.7	142744_pancreas-HLM.Norm-med BMI-35	0.6	0.9
137860_psoas-AA.M.Diab.-med BMI-8	0.5	0.8	139545_pancreas-AA.M.Norm-med BMI-47	4.7	7.9
137834_psoas-CC.M.Diab.-med BMI-2	9.3	14.3	139531_pancreas-AA.M.Norm-med BMI-37	2.0	1.9
137828_psoas-CC.M.Diab.-med BMI-1	21.5	28.9	137871_pancreas-CC.M.Norm-med BMI-26	0.6	1.2
135763_psoas-HLM.Diab.-med BMI-23	10.0	14.4	139541_pancreas-Hi.M.Norm-low BMI-41	2.7	6.2

142740_psoas-AS.M.Diab.-low BMI-20	18.6	28.7	139537_pancreas-CC.M.Norm-low BMI-40	3.3	5.7
134834_psoas-AA.M.Diab.-low BMI-17	5.8	4.9	139533_pancreas-CC.M.Norm-low BMI-39	1.1	0.9
137850_psoas-AS.M.Norm-hi BMI-34	12.8	17.4	137845_pancreas-AS.M.Norm-low BMI-28	1.5	1.6
135769_psoas-HI.M.Norm-hi BMI-31	0.9	1.2	143530_small intestine-AA.M.Diab.-hi BMI-6	3.3	5.0
135766_psoas-AA.M.Norm-hi BMI-25	3.7	4.7	143529_small intestine-CC.M.Diab.-hi BMI-4	25.3	33.2
142746_psoas-AA.M.Norm-med BMI-37	8.2	8.1	143538_small intestine-HI.M.Diab.-med BMI-23	30.8	33.9
142745_psoas-HI.M.Norm-med BMI-35	11.2	19.3	143531_small intestine-AA.M.Diab.-med BMI-8	31.0	58.6
137855_psoas-AA.M.Norm-med BMI-47	2.5	3.4	143528_small intestine-CC.M.Diab.-med BMI-2	21.9	29.5
137844_psoas-CC.M.Norm-med BMI-26	0.5	0.9	143537_small intestine-HI.M.Diab.-low BMI-22	39.0	39.0
142742_psoas-CC.M.Norm-low BMI-40	7.8	14.1	143535_small intestine-AS.M.Diab.-low BMI-20	21.9	18.8
137873_psoas-AS.M.Norm-low BMI-28	0.4	0.6	143534_small intestine-AA.M.Diab.-low BMI-17	0.4	0.6
137853_psoas-HI.M.Norm-low BMI-41	2.9	4.1	143544_small intestine-AS.M.Norm-hi BMI-34	6.1	11.2
135775_psoas-CC.M.Norm-low BMI-39	0.2	0.3	143543_small intestine-HI.M.Norm-hi BMI-31	2.3	2.4
137858_diaphragm-AA.M.Diab.-hi BMI-6	0.3	0.5	143542_small intestine-CC.M.Norm-hi BMI-29	9.2	11.4
135772_diaphragm-AS.M.Diab.-hi BMI-9	8.2	12.2	143539_small intestine-AA.M.Norm-hi BMI-25	8.0	15.3
135761_diaphragm-HI.M.Diab.-hi BMI-21	4.1	5.6	143548_small intestine-AA.M.Norm-med BMI-47	11.4	14.6
134828_diaphragm-CC.M.Diab.-hi BMI-4	0.3	0.7	143547_small intestine-AA.M.Norm-med BMI-37	6.8	9.2
137835_diaphragm-CC.M.Diab.-med BMI-2	6.6	9.6	143540_small intestine-CC.M.Norm-med BMI-26	6.2	7.4
135764_diaphragm-HI.M.Diab.-med BMI-23	1.2	1.8	143550_small intestine-CC.M.Norm-low BMI-40	3.4	4.0
134835_diaphragm-AA.M.Diab.-low BMI-17	0.4	0.6	143549_small intestine-CC.M.Norm-low BMI-39	15.2	20.9
142738_diaphragm-CC.M.Norm-hi BMI-29	3.4	4.3	143546_small intestine-HI.M.Norm-low BMI-41	5.7	9.3
139517_diaphragm-AS.M.Norm-hi BMI-34	2.4	4.2	143525_hypothalamus-HI.M.Diab.-hi BMI-21	37.6	36.6
137848_diaphragm-HI.M.Norm-hi BMI-31	4.2	3.6	143515_hypothalamus-CC.M.Diab.-hi BMI-4	10.7	12.1

137843_diaphragm-AA.M.Norm-hi BMI-25	14.4	20.9	143513_hypothalamus-AA.M.Diab.-hi BMI-6	6.6	10.5
137879_diaphragm-AA.M.Norm-med BMI-47	5.7	6.8	143507_hypothalamus-AS.M.Diab.-hi BMI-9	12.7	13.5
137872_diaphragm-CC.M.Norm-med BMI-26	0.7	0.5	143506_hypothalamus-CC.M.Diab.-med BMI-1	14.3	22.5
135773_diaphragm-HI.M.Norm-med BMI-35	5.8	5.6	143505_hypothalamus-HI.M.Diab.-med BMI-23	24.0	25.9
139542_diaphragm-HI.M.Norm-low BMI-41	2.9	3.0	143509_hypothalamus-AA.M.Diab.-low BMI-17	16.3	16.4
137877_diaphragm-CC.M.Norm-low BMI-39	0.4	0.5	143508_hypothalamus-CC.M.Diab.-low BMI-13	11.7	14.6
137874_diaphragm-AS.M.Norm-low BMI-28	0.3	0.6	143503_hypothalamus-AS.M.Diab.-low BMI-20	10.8	16.4
141340_subQadipose-AA.M.Diab.-hi BMI-6	0.5	0.6	143522_hypothalamus-HI.M.Norm-hi BMI-31	1.4	1.4
137836_subQadipose-HI.M.Diab.-hi BMI-21	0.4	0.6	143516_hypothalamus-AS.M.Norm-hi BMI-34	2.2	3.8
135771_subQadipose-AS.M.Diab.-hi BMI-9	0.5	0.5	143511_hypothalamus-CC.M.Norm-hi BMI-29	7.3	9.5
141329_subQadipose-AA.M.Diab.-medbmi-8	0.2	0.3	143504_hypothalamus-AA.M.Norm-hi BMI-25	5.1	6.4
137862_subQadipose-CC.M.Diab.-med BMI-1	0.5	0.7	143517_hypothalamus-AA.M.Norm-med BMI-47	19.6	22.2
135762_subQadipose-HI.M.Diab.-med BMI-23	0.3	0.5	143514_hypothalamus-HI.M.Norm-med BMI-35	2.8	3.3
141338_subQadipose-AS.M.Diab.-low BMI-20	0.2	0.2	143521_hypothalamus-AS.M.Norm-low BMI-28	9.4	13.7
139547_subQadipose-HI.M.Diab.-low BMI-22	2.4	3.2	143512_hypothalamus-CC.M.Norm-low BMI-40	3.2	4.6
135757_subQadipose-CC.M.Diab.-low BMI-13	0.2	0.5	145454_Patient-25pl (CC.Diab.low BMI.no insulin)	30.8	39.5
134832_subQadipose-AA.M.Diab.-low BMI-17	0.2	0.1	110916_Patient-18pl (HI.Diab.obese.no insulin)	9.8	12.9
141332_subQadipose-HI.M.Norm-hi BMI-31	0.2	0.1	110913_Patient-18go (HI.Diab.obese.no insulin)	4.8	4.9
135767_subQadipose-CC.M.Norm-hi BMI-29	0.0	0.1	110911_Patient-17pl (CC.Diab.low BMI.no insulin)	10.9	19.2
135765_subQadipose-AS.M.Norm-hi BMI-34	0.1	0.3	110908_Patient-17go (CC.Diab.low BMI.no insulin)	5.2	7.7
141339_subQadipose-HI.M.Norm-med BMI-35	0.3	0.2	100752_Patient-15sk (CC.Diab.obese.no insulin)	8.2	14.4
141334_subQadipose-CC.M.Norm-med BMI-26	3.3	4.8	97828_Patient-13pl (CC.Diab.overwt.no insulin)	20.6	24.0
139544_subQadipose-AA.M.Norm-med BMI-47	0.2	0.4	160114_Patient (CC.Diab.obese.insulin) 27-ut	23.0	33.4

137875_subQadipose-AA.M.Norm-med BMI-37	0.3	0.4	160113_Patient (CC.Diab.obese.insulin) 27-pl	36.9	46.0
141331_subQadipose-AS.M.Norm-low BMI-28	7.7	9.7	160112_Patient (CC.Diab.obese.insulin) 27-sk	3.9	4.3
137878_subQadipose-HI.M.Norm-low BMI-41	0.4	1.0	160111_Patient (CC.Diab.obese.insulin) 27-go	7.2	14.0
137876_subQadipose-CC.M.Norm-low BMI-39	0.1	0.2	145461_Patient-26sk (CC.Diab.obese.insulin)	0.9	0.8
137859_vis.adipose-AA.M.Diab.-hi BMI-6	0.1	0.1	145441_Patient-22sk (CC.Diab.low BMI.insulin)	1.6	1.9
135770_vis.adipose-AS.M.Diab.-hi BMI-9	0.7	0.4	145438_Patient-22pl (CC.Diab.low BMI.insulin)	24.7	35.4
135759_vis.adipose-HI.M.Diab.-hi BMI-21	0.1	0.1	145427_Patient-20pl (CC.Diab.overwt.insulin)	87.7	83.5
143502_vis.adipose-CC.M.Diab.-med BMI-2	0.5	1.0	97503_Patient-12pl (CC.Diab.unknown BMI.insulin)	3.6	3.6
139510_vis.adipose-AA.M.Diab.-med BMI-8	13.8	12.6	145443_Patient-23pl (CC.Non-diab.overwt)	68.8	54.3
137861_vis.adipose-CC.M.Diab.-med-1	0.2	0.4	145435_Patient-21pl (CC.Non-diab.overwt)	62.0	69.3
137839_vis.adipose-HI.M.Diab.-med BMI-23	0.4	0.6	110921_Patient-19pl (CC.Non-diab.low BMI)	29.5	36.1
139546_vis.adipose-HI.M.Diab.-low BMI-22	0.4	1.2	110918_Patient-19go (CC.Non-diab.low BMI)	5.4	7.2
137831_vis.adipose-CC.M.Diab.-low BMI-13	0.5	1.1	97481_Patient-08sk (CC.Non-diab.obese)	1.4	1.7
139522_vis.adipose-HI.M.Norm-hi BMI-31	0.2	0.3	97478_Patient-07pl (CC.Non-diab.obese)	16.8	22.5
139516_vis.adipose-AS.M.Norm-hi BMI-34	15.3	23.0	160117_Human Islets-male, obese	58.2	66.0
137846_vis.adipose-CC.M.Norm-hi BMI-29	0.2	0.2	145474_PANC1 (pancreas carcinoma) 1	0.0	0.2
137841_vis.adipose-AA.M.Norm-hi BMI-25	0.0	0.1	154911_Capan2 (pancreas adenocarcinoma)	0.8	1.5
139543_vis.adipose-AA.M.Norm-med BMI-47	100.0	100.0	141190_SW579 (thyroid carcinoma)	3.8	4.3
139532_vis.adipose-AA.M.Norm-med BMI-37	0.5	0.5	145489_SK-N-MC (neuroblastoma) 1	23.3	31.2
139530_vis.adipose-HI.M.Norm-med BMI-35	5.9	8.1	145495_SK-N-SH (neuroblastoma) 1	0.0	0.0
139539_vis.adipose-HI.M.Norm-low BMI-41	0.4	0.5	145498_U87 MG (glioblastoma) 2	14.6	28.5
139535_vis.adipose-CC.M.Norm-low BMI-40	24.1	31.2	145484_HEp-2 (larynx carcinoma) 1	0.0	0.1
137852_vis.adipose-CC.M.Norm-low BMI-39	0.1	0.0	145479_A549 (lung carcinoma)	2.0	4.2

135768_vis.adipose- AS.M.Norm-low BMI-28	0.2	0.0	145488_A427 (lung carcinoma) 2	0.2	0.3
141327_liver-CC.M.Diab.-hi BMI-4	0.4	0.5	145472_FHs 738Lu (normal lung) 1	1.8	3.5
139514_liver-HI.M.Diab.-hi BMI-21	1.6	1.8	141187_SKW6.4 (B lymphocytes)	0.1	0.1
139526_liver-CC.M.Diab.- med BMI-2	1.0	1.2	154644_IM-9 (immunoglobulin secreting lymphoblast)	6.4	7.0
139511_liver-AA.M.Diab.- med BMI-8	0.9	1.2	154645_MOLT-4 (acute lymphoblastic leukemia derived from peripheral blood)	3.8	3.6
137840_liver-HI.M.Diab.- med BMI-23	0.4	0.9	154648_U-937 (histiocytic lymphoma)	0.0	0.0
137827_liver-CC.M.Diab.- med BMI-1	1.3	1.1	154647_Daudi (Burkitt's lymphoma)	0.0	0.0
137838_liver-HI.M.Diab.- low BMI-22	0.5	0.3	145494_SK-MEL-2 (melanoma) 2	15.2	25.7
135758_liver-CC.M.Diab.- low BMI-13	0.5	0.7	141176_A375 (melanoma)	2.0	2.5
139519_liver-CC.M.Norm-hi BMI-29	0.8	1.7	154642_SW 1353 (humerus chondrosarcoma)	0.2	0.1
139518_liver-AA.M.Norm- hi BMI-25	1.0	1.3	141179_HT-1080 (fibrosarcoma)	7.8	10.7
137849_liver-AS.M.Norm-hi BMI-34	0.5	0.9	145491_MG-63 (osteosarcoma) 1	0.0	0.1
137847_liver-HI.M.Norm-hi BMI-31	0.6	1.0	141186_MCF7 (breast carcinoma)	6.2	8.1
142741_liver-AA.M.Norm- med BMI-37	1.8	3.3	141193_T47D (breast carcinoma)	0.0	0.1
141341_liver-HI.M.Norm- med BMI-35	0.2	0.4	154641_BT-20 (breast carcinoma)	0.0	0.0
141335_liver-CC.M.Norm- med BMI-26	0.2	0.5	141175_293 (kidney transformed with adenovirus 5 DNA)	1.8	2.4
139540_liver-HI.M.Norm- low BMI-41	1.1	0.9	141182_HUH hepatoma 1	0.4	0.5
139534_liver-CC.M.Norm- low BMI-39	1.4	1.7	141184_HUH7 hepatoma 1	1.5	2.3
139521_liver-AS.M.Norm- low BMI-28	0.5	1.0	145478_HT1376 (bladder carcinoma)	0.5	0.4
141328_pancreas- CC.M.Diab.-hi BMI-4	80.7	14.3	145481_SCaBER (bladder carcinoma)	1.6	3.3
139525_pancreas- AS.M.Diab.-hi BMI-9	2.1	3.0	141192_SW620 (lymph node metastasis, colon carcinoma) 2	0.0	0.0
137856_pancreas- AA.M.Diab.-hi BMI-6	1.4	1.6	141180_HT29 (colon carcinoma) 1	1.6	2.3
137837_pancreas-	1.4	1.7	141188_SW480 (colon carcinoma) 1	0.0	0.2

HL.M.Diab.-hi BMI-21					
141337_pancreas- CC.M.Diab.-med BMI-2	1.9	4.2	154646_CAOV-3 adenocarcinoma) (ovary	7.9	12.1
139527_pancreas- CC.M.Diab.-med BMI-1	4.1	3.7	141194_HeLa (cervix carcinoma)- 2	0.0	0.1
139515_pancreas- HL.M.Diab.-med BMI-23	31.6	36.9	145482_HeLa S3 (cervix carcinoma) 1	0.0	0.0
139512_pancreas- AA.M.Diab.-med BMI-8	29.5	33.0	145486_DU145 (prostate carcinoma)	8.3	14.6
142739_pancreas- AS.M.Diab.-low BMI-20	7.2	8.5	154643_PC-3 adenocarcinoma) (prostate	25.3	32.1
139513_pancreas- CC.M.Diab.-low BMI-13	18.3	27.2	154649_HCT-8 adenocarcinoma) (ileocecal	22.1	27.9
142743_pancreas- AA.M.Norm.-hi BMI-25	4.5	8.4			

- 5 **General_screening_panel_v1.7 Summary:** Ag7643 Highest expression of this gene is detected in lung (CTs=25-26). High to moderate expression of this gene is seen in brain and tissues with metabolic/endocrine functions including pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, 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,
- 10 such as obesity and diabetes.

Panel 5 Islet Summary: Ag7643 Highest expression of this gene is seen in placenta of a diabetic patient (CTs=30-31). Significant expression of this gene is also seen in islet cells (Bayer patient 1). Thus, the presence of P2Y1R on human islet cells supports a role for this receptor in insulin secretion. An agonist of P2Y1R will be a treatment to augment insulin secretion in Type 2 diabetes.

- 15 **Human Metabolic Summary:** Ag7643 Highest expression of this gene is detected in adipose (CTs=24-25.6). This gene shows widespread expression with high to moderate expression in metabolic tissues. Significantly high expression of this gene is also seen in islet cells from an obese male patient (CTs=25-26). Thus, the presence of P2Y1R on human islet cells supports a role for this receptor in insulin secretion. An agonist of P2Y1R will be a treatment to augment insulin secretion in
- 20 Type 2 diabetes.

Example C3. Screening Assay Formulation

- Assays for screening for antibody therapeutics or small molecule drugs targeting human P2Y1R can be formulated utilizing the non-exhaustive list of cell lines that express the P2Y1R gene
- 25 from the RTQ-PCR results shown above. To assay the activity of P2Y1R, radioligand binding assay using 35S ADP-beta S (Simon et al., 1995. Eur J Pharmacol. 291(3):281-9), as well as, increase in intracellular calcium measurement as described by Palmer et al. (Palmer et al., 1998. Mol Pharmacol 54: 1118-1123) can be utilized.

5 **D. NOV4 -- HUMAN PROTEIN KINASE C, NU (PRKCN) - LIKE PROTEINS, NUCLEIC ACIDS
ENCODING THE SAME & METHODS OF USE THEREOF**

Protein kinase C, nu (PRKCN, PKD3) belongs to a diacylglycerol-dependent protein kinase C subfamily of serine/threonine kinases. It contains an N-terminal hydrophobic sequence, a cysteine-rich motif, a pleckstrin homology domain, and a C-terminal catalytic region (Hayashi et al., 1999. Biochim Biophys Acta. 1450(1):99-106). The closest human homologue of PRKCN is diacylglycerol-dependent protein kinase C mu (PRKCM, PKD) that has been implicated in inhibition of c-Jun N-terminal kinase (JNK) and the protein transport from the trans-Golgi network to the cell surface (Baron et al., 2002. Science 295(5553):325-8). PRKCN localizes both in cell cytoplasm and nuclear, and cell treatment with GPCR agonists induced a rapid phosphorylation and subsequent activation of PRKCN followed by its a nuclear translocation (Rey et al., 2003. J Biol Chem. 278(26):23773-23785).

The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with various diseases or pathologies. The proteins and related proteins that are similar to them, are encoded by a cDNA and/or by genomic DNA. The proteins, polypeptides and their cognate nucleic acids were identified by the inventors in certain cases. In particular, the PRKCN protein encoded by CG197755-01 and any variants, thereof, are suitable as diagnostic markers, targets for an antibody therapeutic and targets for small molecule drugs.

The inventors have discovered that PRKCN interacts with nuclear hormone receptor PPAR gamma, an important regulator of adipose tissue functions (Lowell BB, 1999. Cell 99(3):239-42). Agonists of PPAR gamma known to improve insulin sensitivity are utilized in clinics for the treatment of diabetes (Wang et al., J Cell Biochem. 2003 May 1;89(1):38-47). The inventors have proposed that PRKCN phosphorylates PPAR gamma. Phosphorylation of PPAR gamma is known to decrease its transcriptional activity (Camp et al., 1997. J Biol Chem. 272(16):10811-6; Adams et al., 1997. J Biol Chem. 272(8):5128-32). Inhibition of PRKCN might lead to PPAR gamma activation and subsequent improvement of insulin sensitivity and hyperglycemia. In addition, the inventors have discovered that PRKCN is up-regulated in obese adipose. Activation of PPAR gamma promotes adipogenesis and weight gain (Larsen et al., 2003. Int J Obes Relat Metab Disord. 27(2):147-61). The finding that PRKCN similarly to PPAR gamma overexpressed in obese adipose tissue further supports the role of PRKCN in regulation of PPAR gamma function and adipose metabolism. In one embodiment of the invention, PRKCN is used for identifying a therapeutic compound that would be beneficial in the treatment of diabetes or obesity. As such, the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify such therapeutic antibodies and/or therapeutic small molecules.

In particular the invention relates to the use of PRKCN protein, gene or antibody thereto as a target in diagnostic applications and/or target for small molecule drugs and antibody therapeutics.

The inventors have discovered through PathCalling® (disclosed in Example Q10) that PRKCN interacted with nuclear hormone receptor PPAR gamma, a critical mediator of adipose metabolism. This interaction identifies PRKCN as an important regulator of PPAR gamma activity. Inhibition of PRKCN is expected to improve insulin sensitivity and hyperglycemia by increasing PPAR gamma activity. In experiments described in detail below, PRKCN was up-regulated in the adipose of

5 "genetically obese relative to the normal weight mice using CuraGen's GeneCalling® (disclosed in Example Q7) method of differential gene expression and competitive PCR. Identified positive correlation between PRKCN expression and body weight further supports the role of this kinase in peripheral metabolism and development of metabolic disorder. Therefore, an antagonist of PRKCN is useful for the treatment of obesity and/or diabetes.

10 In a particular embodiment of the invention, PRKCN is a target for screening. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify PRKCN antagonist therapeutic antibodies and/or therapeutic small molecules beneficial in the treatment of obesity and/or diabetes. Not to be limited by a particular mechanism of action, the inventors nevertheless have discovered that inhibition of PRKCN
15 has beneficial effects for treating obesity and/or diabetes by acting in many metabolic tissues, including pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Specifically, in adipose PRKCN inhibition leads to improvement of insulin sensitivity and hyperglycemia in obesity/diabetes.

20 **Discovery Process**

The following sections describe the study design(s) used to identify the PRKCN-encoded protein and any variants thereof, as being suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for treatment of Obesity and Diabetes.

25 **Example D1. Genetically Obese Mice vs. Genetically Lean Mice Study**

Protocol for Genetically Obese Mice vs. Genetically Lean Mice Study is disclosed in Example Q6.

A large number of mouse strains have been identified that differ in body mass and composition. The AKR and NZB strains are obese, the SWR, C57L and C57BL/6 strains are of average weight whereas the SM/J and Cast/Ei strains are lean. Understanding the gene expression
30 differences in the major metabolic tissues from these strains will elucidate the pathophysiological basis for obesity. These specific strains of rat were chosen for differential gene expression analysis because quantitative trait loci (QTL) for body weight and related traits had been reported in published genetic studies. Tissues included whole brain, skeletal muscle, visceral adipose, and liver.

Results

35 A fragment of the mouse PRKCN gene was initially found to be up-regulated by **1.7 fold** in the adipose tissue of obese NZB mice relative to the adipose of normal weight C57BL/6J mice using CuraGen's GeneCalling® (disclosed in Example Q7) method of differential gene expression. A differentially expressed mouse gene fragment migrating, at approximately **225** nucleotides in length (shown in Table D1) was identified as a component of the mouse PRKCN cDNA. The method of
40 competitive PCR was used for confirmation of the gene assessment. The electropherographic peaks corresponding to the gene fragment of the mouse PRKCN are ablated when a gene-specific primer (as shown in Table D1) competes with primers in the linker-adaptors during the PCR amplification. The peaks at **225** nt in length were ablated in the sample from both the adipose.

Table D1. The sequence of the gene fragment (from 47 to 271 band size: 225) and the gene-specific primers used for competitive PCR are indicated on the cDNA sequence of the mouse PRKCN and
--

are shown below in bold. The gene-specific primers at the 5' and 3' ends of the fragment are underlined. Gene length is 486, only region from 1 to 486 shown.

1 TAGCTGTGTG ACTTTGGGTT TGCCCGCATC ATTGGTGAGA AGTCGTT**CCG**
GAGGTCAGTG

61 **G**TAGGA**ACTC** CAGCCTACTT AGCCCCTGAG GTTCTCAGGA GCAAAGGCTA
 CAACCGTTCT

121 CTAGACATGT GGTCCGTGGG CGTCATCGTT TATGTGAGCC TCAGTGGCAC
 ATTCCCATTC

181 AATGAAGATG AAGATATAAA TGACCAAATC CAAAACGCTG CGTTTATGTA
 CCCACCAAAT

241 CCATGGCGAG AAATTT**CCAG TGAAGCAATT** GACTTGATAA ACAACTTGCT
 TCAAGTGAAG

301 ATGAGAAAAC GATACAGTGT TGACAAATCT CTCAGTCATC CTTGGCTACA
 GGATTATCAG

361 ACTTGCTTG ACCTCAGAGA ATTTGAAATT CGCATTGGAG AACGTTACAT
 TACACACGAA

421 AGTGATGATG CTCGCTGGGA AATACACGCA TACACGCACA ATCTGGAATA
 CCCAAAGCAT

481 TTCATT (SEQ ID NO: 79)

Example D2. Human CG197755-01 Sequence Identification

The sequence of Human PRKCN (Acc. No. CG197755-01) was derived by laboratory screening of cDNA library. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were 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 protocol for identification of human sequence(s) is disclosed in Example Q8.

An alignment of the protein sequences of the human (CG197755-01), mouse homologs of the PRKCN is shown in Table D2.

Table D2. Protein alignment (ClustalW) of the human (CG197755-01) PRKCN (SEQ ID NO: 14) and mouse PRKCN fragments NP083515 (SEQ ID NO: 80) and scm_gb-be847119_1 (SEQ ID NO: 81).

```

NP_083515
CG197755-01 MSANNPPSAQKSVLPTAIPAVLPAAAPCSPTKGLSARLSNGSF SAPSLTNSRGSVHTV
scm_gb-bc347119_1

NP_083515
CG197755-01 SFLLQIGLTRESVTIEAQELSLSAVKDLVCSIVYQKMFPECGFFGMVDKILLFRHDMNSEN
scm_gb-bc347119_1

NP_083515
CG197755-01 ILQLITSADEIHEGDLVEYVLSALATVEDFQIRPHALYVHSYKAPTCDYCGEMLWGLVR
scm_gb-bc347119_1

NP_083515
CG197755-01 QQLKCEGCQLNYHKRCFAKIPNNCSQVRKRRLSNVSLPGPGLSVP RPLQPECVPLLSEES
scm_gb-bc347119_1

NP_083515
CG197755-01 HTHQEPSEKIPFSWSGRPIWMEKMVMCRVKVPHTFAVHSYGRPTICQYCKRLLEGLFRQGM
scm_gb-bc347119_1

NP_083515
CG197755-01 QCKDCKFNCHKRCASKVPRDCLGEYTFNGEPSSLTDTADMPMDIDSSDUNSDSSRGIDDS
scm_gb-bc347119_1

NP_083515
CG197755-01 REPSPPEKMFLLDPSOLDVERDEEAVKTIKSPSTSMNIFLMRWVQSIKHKRKSSTMVKE
scm_gb-bc347119_1

NP_083515
CG197755-01 QWMVHYTSRDNLKRRHYWRLDSEKCLTLFQNESGSKYKKEIPLSEILRISSPADFTNISQQ
scm_gb-bc347119_1

NP_083515
CG197755-01 SNPHCFEII TDMVYVFGENNGDSSEMPVLAATGVGLDVAQSWEKALRQALMPVTPQASV
scm_gb-bc347119_1

NP_083515
CG197755-01 CTSPQQRDKDKLSTSSISVSNQIQENVDSITVYQIFADEVLGSQQFQIVYQGHKRAKTR
scm_gb-bc347119_1

NP_083515
CG197755-01 DVAIKVIDKMRFPPTQESQLRNEVAIQLNLHHPQIVNLECMFETPERVFFVMEKLGDMLE
scm_gb-bc347119_1

NP_083515
CG197755-01 EMILSSEKSRLEPRTKFMVTQILVALRNLFHFNIVHCDLKPENVLLASAEFPFQVKLCD
scm_gb-bc347119_1

NP_083515
CG197755-01 FQFARIIGKSFRRSVVGTPAYLAPVLRSGYMRSLDMWSVGVIVVWSLSTGTFPPNEDE
scm_gb-bc347119_1

NP_083515
CG197755-01 DINDQIQNAAFMYPPNPWREISSEAIDLINLLQVKMRKRYSDKSLSHPWLDQYQTWLD
scm_gb-bc347119_1

NP_083515
CG197755-01 LREFEIRIGERYITHESDDARWEIHAYTHNLVYPKHFIIMAPNPDDMEEDP
scm_gb-bc347119_1

```

5

The laboratory cloning was performed using one or more of the methods summarized in Example Q8. The NOV4 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table D3.

Table D3. NOV4 Sequence Analysis

NOV4a,	CG197755-01	SEQ ID NO: 13	5792 bp
DNA Sequence		ORF Start: ATG at 556	ORF Stop: TAA at 3226
AAAGTTCATCCCCCAGAATGAAAAATGAGGACATTTGAGAAGGTGATTAAAGGTGTGGACATTTGAGA			
AGGTGTCCTATCAAATAGTAAACCAAGGAAAAGTACTGAATAGATTAATCCAAACACTTACTGTT			

TTTTAAACGAAGAGGATTTACCTTGACAGAAAAACAACCTTTATTCAATATGTATTTCTGAAATTA
AAGAGACAAGTACAGACTGAAAGGAAAAATAGATTTCGTAAATAAGCTACGTCAACTCTATCCTGCCTGAG
GATAGCTCAGTGATGTTAAATCCTTTACAAATCCTCGGTTGTCTTCTCCTACAGACAAGACTGCTTTTTC
ATGGGACTGATATTAAGAGAAAATAGGACCTTTGGGGCATTCAACTCCTTGATAAAACTTAAAGTATC
GGCATGAGTGGCTTAACAGAGGAAAATAAGAAGTTTTCAACTAAATCCAAAAGTGCGGTCAATTTCTT
TACTGCTGTTATTTTAAAAACCTCTTCATAACCATGAAAAAGAATCGACAACATTTTAAAAAGATTA
AAGAAAGGCAGATGCTGCAATAAATCCCCCTCCATCAGCCCAGAAGCTGTATTACCCACAGCTATT
CCTGCTGTGCTTCCAGCTGCTTCTCCGTGTTCAAGTCTAAGACGGGACTCTCTGCCGACTCTCTAA
TGGAAGCTTCAGTGCACCATCACTACCAACTCCAGAGGCTCAGTGCATACAGTTTCATTTCTACTGTC
AAATTGGCCTCACACGGGAGAGTGTATACCATTAAGGCCAGGAAGTGTCTTTATCTGCTGTCAAGGAT
CTTGTGTGCTCCATAGTTTATCAAAAGTTTCCAGAGTGTGGATTCTTTGGCATGTATGACAAAATTCCT
TCTCTTTTCGCCATGACATGAAGTACAGAAAACATTTTGCAGCTGATTACCTCAGCAGATGAAATACATG
AAGGAGACCTAGTGAAGTGGTTCCTTTCAGCTTTAGCCACAGTAGAAGACTTCCAGATTTGCCATC
ACTCTCTATGTACATCTTACAAAAGCTCCTACTTTCTGTGATTACTGTGGTGAGATGCTGTGGGGATT
GGTACGTCAGGACTGAAATGTGAAGGCTGTGGATTAAATTACCATAAACGATGTGCCTTCAAGATTC
CAAAATACTGTAGTGGAGTAAGAAAGAGACGCTGTCAAAATGTATCTTTACCAGGACCCGGCCTCTCA
GTTCCAAGACCCCTACAGCTGAATATGTAGCCCTTCCCAGTGAAGAGTCACATGTCCACCAGGAACC
AAGTAAGAGAATTCCTTCTTGGAGTGGTCGCCCAATCTGGATGGAAGAAGATGGTAATGTGCAGAGTGA
AAGTTCACACACATTTGCTGTTCACTCTTACACCCGTCCCACGATATGTCAGTACTGCAAGCGGTTA
CTGAAAGCCCTCTTTTCGCCAAGGAATGCAGTGTAAAGATTGCAAAATTCAACTGCCATAAACGCTGTGC
ATCAAAAGTACCAAGAGACTGCCTTGGAGAGGTTACTTTCAATGGAGAACCTTCCAGTCTGGGAACAG
ATACAGATATACCAATGGATATTGACAATAAGTGAATAGTAGTGTAGTGGGGTTTGGATGAC
ACAGAAGAGCCATCACCCCCAGAAGATAAGATGTTCTTCTTGGATCCATCTGATCTCGATGTGGAAG
AGATGAAGAAGCCGTTAAACAATCAGTCCATCAACAAGCAATAATATTCGCTAATGAGGGTTGTAC
AATCCATCAAGCACACAAAGAGGAAGAGCAGCACAATGGTGAAGGAAGGGTGGATGGTCCATTACACC
AGCAGGGATAACCTGAGAAAGAGGCATTATTGGAGACTTGACAGCAAAATGTCTAACATTTATTAGAA
TGAATCTGGATCAAAAGTATTATAAGGAAATTCACCTTTTCAGAAATTCCTCCGATATCTTACCACGAG
ATTTACAAAACATTTACAAGGCAGCAATCCACACTGTTTGAATCATTACTGATACTATGGTATAC
TTCGTTGGTGAGAACAATGGGGACAGCTCTCATATCCTGTTCTTGTCTGCCACTGGAGTTGGACTTGA
TGTAGCACAGAGCTGGGAAAAAGCAATTCGCCAAGCCCTCATGCCTGTTACTCCTCAAGCAAGTGTTC
GCACCTTCCAGGCAAGGGAAGATGCACAAAGATTTGTCTACAAGTATCTCTGTATCTAATTTGTCAG
ATTCAGGAGAATGTGGATATCAGTACTGTTTACCAGATCTTTGCAGATGAGGTGCTTGGTTCCAGGCA
GTTTGGCATCGTTTATGGAGGAAAACATAGAAAGACTGGGAGGGATGTGGCTATTAAAGTAATTTGATA
AGATGAGATTTCCCAACAAAACAAAGAAAGTCAACTCCGTAATGAAGTGGCTATTTTACAGAAATTCGAC
CATCTTGGGATGTAAACCTGGAATGTATGTTTGAACCCCAAGAGTCTTTGTGATTAATTTGGA
GCTGCATGGAGATGTTTGGAAATGATTCTATCCAGTGAGAAAAGTCGGCTTCCAGAACGAATTAATA
AATTCATGGTCACACAGATACTTGTGCTTTGAGGAATCTGCATTTTAAGAATATTGTGCACTGTGAT
TTAAAGCCAGAAAATGTGCTGCTTGCATCAGCAGAGCCATTTCTCAGGTGAAGCTGTGTGACTTTGG
ATTTGCACGCATATTGGTGAAGAGTCATTAGGAGATCTGTGGTAGGAACCTCCAGCATACTTAGCCC
CTGAAGTTCTCCGGAGCAAAAGGTTACAACCGTTCCCTAGATATGTGGTCAGTGGGAGTTATCATCTAT
GTGAGCCTCAGTGGCACATTTCTCTTTAATGAGGATGAAGATATAAATGACCAAAATCCAAAATGCTGC
ATTTATGTATCCCAACAAATCCATGGAGAGAAATTTCTGGTGAAGCAATTTGATCTGATAACAAATCTGC
TTCAAGTGAAGATGAGAAAACGTTACAGTGTGACAAATCTCTTAGTCATCCCTGGCTACAGGACTAT
CAGCATTTGGCTTATGCTTAGAATTTGAAACTCCGATTTGGAGAACGTTACATTACACATGAAGTGA
TGATGCTCGCTGGGAAATACATGCATACACATAACCTTGTATACCCAAAGCACTTCATTATGGCTC
CTAATCCAGATGATATGGAAGAAGATCCTTAATCACTGAGCTAACCTAAATAAGGAAGGATTTTCATTT
TATGGACTGATATTTTGTGCTGTAACCTTGTCTTTCGTAGATTGTCTGTCAGTGTCTGCAAGATATG
AAGAAATATGATAACGAATAAGTGACACCAGTACTGTAGTTCATAATGAGTAGGTACAGGCGGGAAC
TGAATAATAAGAAGTCATAATGGAATCAAGGTGAAGCTTTTATAAACTTTTTCAGCCTAAGCAATAA
CTGGTTTTGTATTTTCTTAAATCCTTCAATATACAAATAGGCTCACTTAATTTGTCTTCCATTT
CTCTTTATATATATATATATATATAAAAAAATATAAATATATGTTTGTGTTGTTGTTTTTTAAGGAA
AAACAAGTCAAGCTAGCATCCAGTTACTATATAGCTTGGCTAAATTTATACAAGACTTACAAGATTGAT
TACTCGACAGGCTTGTATTTAAGAGATAACTGTGAGGTTACCATTATGTGATGTTACTATAAGGACTT
TTAACAATGGTTTAAACAAACCATAGAGGCATTGAAGGGTTTTTCTTAGATGCCTAGAAAAGCACACT
GGGCTGTTTTTACCTTTCTTTTAGGTCAATCAAGACTCCAAAATAGTGATTTCTTAACCTTTTGGAG
TTGCTCTGCTACTCTGAATATGTTCTATACAGCATAAGGATTGTCACCTTCTGTGTGTTGCAACAGCT
TCTAAGATAATTAGGGACAAATGATGTTACAAAAGGAAGAGTACTGCTGGTCTAAGTGCTGAGTTGTA
TGTCTTTGCATAGCTCCACTCTGCTGCTAAATATGCATGTTCTGACTGACACCATCTTGATGCCAGTA
CTGGATTCCAGCATTCAGCAGGTGCAGATCTCGGCTTTACACAATTTATCTTTACCTAGGGTTTCAGTC
AGTAATTTCTGCTTTTTCAGCCAGGGCCAGTGCAGGGTCAGTTAATGCTACAGTTACTGTATAGCAAC
AGTATCCTTTTTTCTCCTTCCCTTAGCCTATTGGGCTTTGCAGATATCTGGAGTGTTTTAAAGTCAAT
TATTTTAAAGCAGTTTGGGGGATGTGTAGGAGTGGAGCATGAAACAGTTTATAAGGCTGGGGCTGTAT
TATCAGCAGCAGCAATTAAGAATGAAGAAGTACTTCTTTTACATTTACAGTCCAGCAGCCAGCTA
TTTTAAAAAATATTTTTTAAATATCTTCCCAAGTTTAAAGATGTGGGACCATCTACTTGTAAAGAAACAG
TGGCTTATTTCTTCATTTCTCAGTAATCATTGTAAACCTTTTTTTTTTTCTCTCCACTCTAACAAAA
GTAAAGAAACAAACTTTTTGCTGAGGTCTTGGACTCCACTACTTGTAGTTATGTTGGCCTGGACAGA
TGGCAGGTCTGTGGCAGTTCTCTCAGTCTGTAAATGAAGATTATCATTTGCTCTCTCCCTTCTGTC
CAGTGTGTGCTGAAGATCAGATATGATGAATGCAGTCAAAATAAACTAAAAACTAGGAAAGTGTAA
CTATCGTTGCCACCGAATTTGAGGTAGCAAAAAAAGAAATGTTCTGGTACATGAACAGACTGA
GAAGGCCTTTCGACATCCCTGAAGCTAGTCTCTGCTAGACTAAAAATATCCTTGAAAGAATAGTAG

CAAACAGTATGCGAGTACTCCTGTGTTCTAAGTTCCTTTATGTTTGTAGTTTACCCCTCACAGCAGCCATA GTGTTTAACTCATTATGAGCTTCATCTTAAAGATAAGAAAAGTGGGCAAGGGGAGATTAACTTACTGC CAAATTTCAAACAGTTAGTAGTATTTGAGACCAGTAGTATGGCTCTAATCTTACCCCTAGCCCTCCCA ACCATTTAGCTGTACTATCTACCCATAAATACCTCCAGGAAACATTCCCTCTTGGTAATTTGTCTCTT CTATAGAAGCTTCAAAGATTAAATCAAATTAATTTTCAAAAAAATTTTCTTTAGAAAATGTTATTTT CCCTGTGATAGAGGATGACTTCCAGTTTACCAAAGTCTGTTTATATCAACACACACAAAATGGAAT AATTCTGAGTCACTAGGCAATCAATCTACTGTGGTTTACTATGTAAGGTGAAAATTAAGTGAACGA TGTGTTGTTGCTACTATACATAGTCAAACCTTACAAGCCATGAAATTAATTGCACTCTTTGTATTG TTGTTAAATGCTAAGAAGTTTCTAAAAAATTTGTAAAGGCACTGTGAGAGAATCTGGAGTTGAATG ATTATCCAGATACTGTATAACCTGCATAAATTTTGTCTTTAAGTCGTGTTGTAAAAGAAGTAATT GCTAGAACATTTGATAATGTACAAAGTAGTCTATAATGACTGTTCACTACATTTTAAATATTTTCT GGTATATCAAACTTTTGTAAATATACTGGAAGCTTGATAATAAAATGTATTTCTATCACCATACT TTTCCATGTGAAAACCTGAGCCTATTTCTAGTATAAGTATCCAAAGAAAAGTTTACCTGGTTGTGTT ATTTTACCAGAC			
NOV4a, Protein Sequence	CG197755-01	SEQ ID NO: 14	890 aa MW at 100469.6kD
MSANNSPPSAQKSVLPTAIPAVLPAASPCSSPKTGLSARLSNGSFSAPSLTNSRGSVHTVSVFLLQIGL TRESVTIEAQELSLSAVKDLVCSIVYQKFPECGFFGMYDKILLFRHDMNSENILQLITSADEIHEGDL VEVVLSALATVEDFQIRPHTLYVHSYKAPTFCDYCGEMLWGLVRQGLKCEGCGLYNHRKCAFKIPNNC SGVRKRRLSNVSLPGPGLSVPRPLQPEYVALPSEESHVHQEPSKRIPSWSGRPIWMEKMVMCRVKVPH TFAVHSYTRPTICQYCKRLKGLFRQGMQCKDKCFNCHKRCASKVPRDCLGEVTFNGEPSSLGTDTDI PMDIDNNINDSSRGLDDTEEPSPPEDKMFLLDPSDLDERDEEAVKTIPTSTNNIPLMRVVSQSIK HTKRKSSMTMVEGWMVHYTSRDNLKRHYWRDLSKCLTLFQNESGSKYKEIPLSEILRISSPRDFTN ISQGSNPHCFEIIITDMVYFVGENNGDSSHNPLAATGVGLDVAQSWEKAIRQALMPVTPQASVCTSP GQKDKHDLSTSISSVNCQIQENVDISTVYQIFADEVLGSGQFIVYGGKHKRKTGRDVAIKVIDKMRP PTKQESQLRNEVAILQNLHHPGIVNLECMFETPERVFMVEKLHGDMLEMILSSEKSRLEPERITKFMV TQILVALRNLHFKNIVHCDLKPENVLLASAEFPQVKLCDFGFARIIGKSFRRSVVGTPAYLAPEVL RSKGYNRSLDMWSVGVIIVVSLSGTFPFNEDEDINDQIQNAAFMYPPNPWREISGEAIDLINLLQVK MRKRYSVDKSLSHPWLDQYQTWLWDLREFETRIGERYITHESDDARWEIHAYTHNLVYPKHFIAPNPD DMEEDP			
NOV4b, DNA Sequence	CG197755-02	SEQ ID NO: 15	2697 bp
		ORF Start: ATG at 4	ORF Stop: TAA at 2695
ACCATGGGACATCATCACCACCATCACTCTGCAATAATTCCCTCCATCAGCCAGAAAGTCTGTATT ACCCAGAGCTATTCTGTGCTGTGCTTCCAGCTGCTTCTCCGTGTTCAAGTCCTAAGACGGGACTCTCTG CCCGACTCTCTAATGGAAGCTTCAGTGCACCATCACTCACCACCTCCAGAGGCTCAGTGCATACAGTT TCATTTCTACTGCAAAATTGGCCTCACACGGGAGAGTGTACCATTTGAAGCCCAGGAAGTGTCTTTATC TGCTGTCAAGGATCTTGTGTGCTCCATGTTTATCAAAAGTTTCCAGAGTGTGGATTCTTTGGCATGT ATGACAAAATTCTTCTCTTTCGCCATGACATGAAGTCAAGAAACATTTTGCAGCTGATTACCTCAGCA GATGAAATACATGAAGGAGACCTAGTGAAGTGGTTCTTTTCAAGCTTTAGCCACAGTAGAAGACTTCCA GATTTCGTCCACATACCTCTATGTACATTTCTACAAAGCTCCTACTTTCTGTGATTACTGTGGTGAGA TGCTCTGGGGATTGGTACGTCAAGGACTGAAATGTGAAGGCTGTGGATTAAATTACCATAAACGATGT GCCTTTCAAGATTCCAAATAACTGTAGTGGAGTAAAGAAAGAGACGCTCTGTCAAATGTATCTTTACGAG ACCCGGCTCTCAGTTCCAAGACCCCTACAGCTGAATATGTAGCCCTTCCAGTGAAGAGTCACATG TCCACCAGGAACCAAGTAAGAGAATTCCTTCTTGGAGTGGTCGCCAATCTGGATGGAAGAGATGGTA ATGTGCAGAGTGAAAGTTCCACACACATTTGCTGTTCACTCTTACACCCGTCCCACGATATGTCAGTA CTGCAAGCGGTTACTGAAAGGCTCTTTTCGCCAAGGAATGCAGTGTAAAGATTGCAAAATTCACGTCC ATAAACGCTGTGCATCAAAAGTACCAAGAGACTGCCTTGGAGAGGTTACTTTCAATGGAGAACCTTCC AGTCTGGGAACAGATACAGATATACCAATGGATATTGACAATAATGACATAAATAGTGATAGTAGTCG GGGTTTGGATGACACAGAAGAGCCATCACCCCAAGATAAGATGTTCTTCTTGGATCCATCTGATC TCGATGTGGAAGAGATGAAGAAGCCGTAAAAACAATCAGTCCATCAACAAGCAATAATATCCGCTA ATGAGGGTTGTACAATCCATCAAGCACACAAAGAGGAAGAGCAGCAATGGTGAAGGAAGGTGGAT GGTCCATTACACCAGCAGGAGATAACCTGAGAAAGAGGCATTATTTGAGACTTGACAGCAATGTCTAA CATTTATTTTCAAGATGAATCTGGATCAAGTATTTATAAGGAATTTCCACTTTTCAAGAAATCTCCGCATA TCTTACCACGAGATTTCAAAACATTTCAAGGCAGCAATCCACACTGTTTGAATCATTAAGTGA TACTATGGTATATCTCGTTGGTGAGAACAATGGGGACAGCTCTCATATCCTGTTCTTGTGCCACTG GAGTTGGACTTGATGTAGCACAGAGCTGGGAAAAAGCAATTCGCCAAGCCCTCATGCCTGTACTCCT CAAGCAAGTGTGTTGCACTTCTCCAGGGCAAGGGAAGATCACAAAGATTTGTCTACAAGTATCTCTGT ATCTAATTTGTGAGATTGAGGAGATGTGGATATCAGTACTGTTTACCAGATCTTTGCAGATGAGGTGCT TTGGTTTCAAGCCAGTTTGGCATCGTTTATGGAGGAAAAACATAGAAAGACTGGGAGGGATGTGGCTATT AAAGTAATTGATAAGATGAGATTCCCCACAAAACAAGAAAGTCAACTCCGTAACGAAGTGGCTATTTT ACAGAATTTGCACCATCTGGGATTGTAAACCTGGAATGTATGTTTGAACCCCAAGAGTCTTTG TAGTAATGGAAGCTGCATGGAGATATGTTGGAATGATTTCTATCCAGTGAGAAAAGCCGGCTTCCA GAACGAATTACTAAATTCATGGTCACACAGATACTGTGTTGCTTTGAGGAATCTGCATTTTAAGAATAT TGTGCATGTGATTTAAAGCCAGAAAATGTGCTGCTTGCATCAGCAGAGCCATTTCTCAGGTGAAGC TGTGTGACTTTGGATTGTCACGCATCATTTGGTGAAGAGTCATTGAGGAGATCTGTGGTAGGAAGTCCA GCATACTTAGCCCTGAAAGTTCTCCGGAGCAAGGTTACAACCGTTCCCTAGATATGTGGTCAAGTGGG AGTTATCATCTATGTGAGCCTCAGTGGCACATTTCTTTTAAATGAGGATGAAGATATAAATGACCAAA			

TCCAAAATGCTGCATTTATGTACCCACCAATCCATGGAGAGAAATTTCTGGTGAAGCAATTGATCTG
ATAACAATCTGCTTCAAGTGAAGATGAGAAAACGTTACAGTGTGACAAATCTCTTAGTCATCCCTG
GCTACAGGACTATCAGACTTGGCTTGACCTTAGAGAATTTGAACTCGCATTTGGAGAACGTTACATTA
CACATGAAAGTGATGATGCTCGCTGGGAAATACATGCATACACATAACCTTGTATACCCAAAGCAC
TTCATTATGGCTCCTAATCCAGATGATATGGAAGAAGATCCTTAA

NOV4b,	CG197755-02	SEQ ID NO: 16	897 aa	MW at 101349.5kD
Protein Sequence				

MGHHHHHHSANNSPPSAQKSVLPPTAIPAVLPAASPCSSPKTGLSARLSNGSFSAPSLTNSRGSVHTVS
FLLQIGLTRESVTIEAQELSLSAVKDLVCSIVYQKFPCEGFFGMYDKILLFRHDMNSENILQLITSAD
EIEHGDLEVVLSALATVEDFQIRPHTLYVHSYKAPTFCDYCGEMLWGLVRQGLKCEGCGLNYHKRCA
FKIPNNCSGVRKRRLSNVSLPGPGLSVPRPLQPEYVALPSEESHVHQEPSKRIPSWSGRPIWMEKMVM
CRVKVPHTFAVHSYTRPTICQYCKRLLKGLFRQGMQCKDKFNCHKRCASKVPRDCLGEVTFNGEPSS
LGTDTDIPMDIDNNDINSDSSRGLDDTEEPSPPEDKMFLLDPSDLDERDEEAVKTISPSTSNIPLM
RVVQSIKHTKRKSSTMVKEGWMVHYTSRDLNRKRHYRWLDSKCLTLFQNESGSKYKEIPLSEILRIS
SPRDFNTNISQGSNPHCFEIIITDMVYFVGENNGDSSHNPLAATGVGLDVAQSWEKAIRQALMPVTPQ
ASVCTSPGQKGDHKLSTISVSNCQIQENVDISTVYQIFADEVLGSGQFGIVYGGKHKRKTGRDVAIK
VIDKMRFPKQESQLRNEVAILQNLHHPGIVNLECMFETPERVFWMEKLHGDMLEMILSSEKSRLEPE
RITKFMVTOILVALRNLHFKNIVHCDLKPENVLASAEPFPQVKLCDFGFARIIEKSFRRSVVGTPE
YLAPEVLRSGYNRSLDMWSVGVIYVSLSGTFPFNEDEDINDQIQNAAFMYPNPNPWEISGEAIDL
NNLLQVKMRKRYSDKSLSHPWLQDYQTLWDLREFETRIGERYITHESDDARWEIHAYTHNLVYPKHF
IMAPNPDDMEEDP

NOV4c,	CG197755-03	SEQ ID NO: 17	2317 bp
DNA Sequence		ORF Start: at 2	ORF Stop: TAA at 2315

TACCATGGGACATCATCACCACCATCACTCTGCAAATAATTCCTCCATCAGCCCAGAACTCTGTAT
TACCACAGCTATTCCCTGCTGTGCTTCCAGCTGCTTCTCCGTGTTCAAGTCTAAGACGGGACTCTCT
GCCCGACTCTCTAATGGAAGCTTCAGTGCACCATCACTCACCACCTCCAGAGGCTCAGTGCATACAGT
TTCATTTCTACTGCAAATTTGGCCTCACACGGGAGAGTGTACCATTGAAGCCCAGGAACGTCTTTTAT
CTGCTGTCAAGGATCTTGTGTGCTCCATAGTTTATCAAAAGTTTCCAGAGTGTGGATTCTTTGGCATG
TATGACAAAATCTTCTCTTTCGCATGACATGAACACTCAGAAAACATTTTGCAGCTGATTACCTCAGC
AGATGAAATACATGAAGGAGACCTAGTGAAGTGGTTCTTTCAGCTTTAGCCACAGTAGAAGACTTCC
AGATTCGTCCACATACCTCTATGTACATTCTTACAAAGCTCCTACTTTCTGTGATTACTGTGGTGAG
ATGCTCTGGGGATTGGTACGTCAAGGACTGAAATGTGAAGGCTGTGGATTAAATTACCATAAACGATG
TGCCCTTCAAGATTCCAAATAACTGTAGTGGAGTAAGAAAGAGACGTCTGTCAAATGTATCTTTACCAG
GACCCGGCCTCTCAGTTCCAAAGACCCCTACAGCTGAATATGTAGCCCTTCCAGTGAAGAGTACAT
GTCCACCAGGAACCAAGTAAGAGAATTCCTTCTTGGAGTGGTCGCCCAATCTGGATGGAAGATGGT
AATGTGCAGAGTGAAAGTTCCACACACATTTGCTGTTCACTCTTACACCCGTCCCACGATATGTCACT
ACTGCAAGCGGTTACTGAAAGGCCTCTTTCGCCAAGGAATGCAGTGTAAAGATTGCAAATTCACATGC
CATAAACCGTGTGATGATCAAAAGTACCAAGAGACTGCTTGGGGAGGTTACCTTCAATGGAGAACCTTC
CAGTCTGGGAACAGATACAGATATACCAATGGATATTGACAATAATGACATAAATAGTGATAGTAGTC
GGGGTTTGGATGACACAGAAGAGCCATCACCACCAGGAAGATAAGATGTCTTCTTGGATCCATCTGTAT
CTCGATGTGGAAGAGATGAAGAAGCCGTTAAACAATCAGTCCATCAACAAGCAATAATATTCGGCT
AATGAGGGTTGTACAATCCATCAAGCACACAAGAGGAAGAGCAGCAATGGTGAAGGAAGGTTGGA
TGAAGAGATTTGTCTACAAGTATCTGTATCTAATTGTGATGATTCAGGAGAAATGTGGATATCAGTACT
GTTTACCAGATCTTTGCAGATGAGGTGCTTGGTTTCAAGCCAGTTTGGCATCGTTTATGGAGGAAAACA
TAGAAAGACTGGGAGGGATGTGGCTATTAAAGTAATTGATAAGATGAGATTCCCCACAAAACAAGAAA
GTCAACTCCGTAACGAAGTGGCTATTTTACAGAATTTGCACCATCCTGGGATTGTAAACCTGGAATGT
ATGTTTGAACACCCAGAACGAGCTTTTGTAGTAAATGGAAGAAAGCTGCATGGAGATATGTTGGAATGAT
TCTATCCAGTGAGAAAAGCCGGCTTCCAGAACGAATTACTAAATTCATGGTCACACAGATACTTGTG
CTTTGAGGAATCTGCATTTTAAAGAATATTGTGCACTGTGATTTAAAGCCAGAAAATGTGCTGCTTGCA
TCAGCAGAGCCATTTCTCAGGTGAAGCTGTGTGACTTTGGATTGTCACGCATCATTGGTGAAAAGTC
ATTCAGGATAGTGTGGTAGGAACCTCCAGCATACTTAGCCCTGAAGTTCTCCGGAGCAAAGGTTACA
ACCGTTCCCTAGATATGTGGTCAGTGGGAGTTATCATCTATGTGAGCCTCAGTGGCACATTTCCCTTTT
AATGAGGATGAAGATATAAATGACCAATCCAAATGCTGCATTTATGTACCCACCAAATCCATGGAG
AGAAATTTCTGGTGAAGCAATTGATCTGATAAACAATCTGCTTCAAGTGAAGATGAGAAAACGTTACA
GTGTTGACAAATCTCTTAGTCATCCCTGGCTACAGGACTATCAGACTTGGCTTGACCTTAGAGAATTT
GAAACTCGCATTTGGAGAACGTTACATTACATGAAAGTGATGATGCTCGCTGGGAAATACATGCATA
CACACATAACCTTGTATACCCAAAGCACTTCATTATGGCTCCTAATCCAGATGATATGGAAGAAGATC
CTTAA

NOV4c,	CG197755-03	SEQ ID NO: 18	771 aa	MW at 87211.8kD
Protein Sequence				

TMGHHHHHHSANNSPPSAQKSVLPPTAIPAVLPAASPCSSPKTGLSARLSNGSFSAPSLTNSRGSVHTV
SFLQIGLTRESVTIEAQELSLSAVKDLVCSIVYQKFPCEGFFGMYDKILLFRHDMNSENILQLITSAD
DEIHEGDLVEVVLSALATVEDFQIRPHTLYVHSYKAPTFCDYCGEMLWGLVRQGLKCEGCGLNYHKRCA
AFKIPNNCSGVRKRRLSNVSLPGPGLSVPRPLQPEYVALPSEESHVHQEPSKRIPSWSGRPIWMEKMV
MCRVKVPHTFAVHSYTRPTICQYCKRLLKGLFRQGMQCKDKFNCHKRCASKVPRDCLGEVTFNGEPSS

SLGTDTDIPMDIDNNDINSRSSRLDDTEEPSPPEDKMFLLDPSDLDERDEEAVKTIISPSTSNNIPL MRVVQSIKHTKRKSSSTMVKEGWMKDLSTISVSNQCIIQENVDISTVYQIFADEVLGSGQFGIVYGGKH RKTGRDVAIKVIDKMRFPKQESQLRNEVAILQNLHHPGIVNLECMFETPERVVFVMEKLGDMLEMI LSSEKSRLPERITKFMVTQILVALRNLHFKNIVHCDLKPENVLASAEPFPQVKLCDFGFARIIGES FRRSVVGTPAYLAPEVLRSGYNRSLDMWSVGVIIVVSLSGTFPFNEDEDINDQIQNAAFMYPPNPWR EISGEAIDLINLLQVKMRKRYSDVKSLSHPWLQDYQTWLTLREFETRIGERYITHESDDARWEIHAY THNLVYPKHFIMAPNPDDMEEDP			
NOV4d, DNA Sequence	CG197755-04	SEQ ID NO: 19	2711 bp
		ORF Start: ATG at 4	ORF Stop: at 2674
ACCATGTCTGCAAATAATTCCCCTCCATCAGCCAGAGTCTGTATTACCCACAGCTATTCTGCTGT GCTTCCAGCTGCTTCTCCGTGTTCAAGTCTAAGACGGGACTCTCTGCCGACTCTCTAATGGAAGCT TCAGTGCACCATCACTCACCAACTCCAGAGGCTCAGTGCATACAGTTTCATTTCTACTGCAAATTGGC CTCACACGGGAGAGTGTACCATTGAAGCCAGGAAGTGTCTTTATCTGCTGTCAAGGATCTTGTGTG CTCCATAGTTTTATCAAAGTTTCCAGAGTGTGGATTCTTTGGCATGTATGACAAAATTTCTTCTTTTC GCCATGACATGAAGTCAAGAAACATTTTGACGTGATTACCTCAGCAGATGAAATACATGAAGGAGAC CTAGTGGAGAGTGGTTCTTTTCAGCTTTAGCCACAGTAGAAGACTTCCAGATTCTGTCACATACTCTCTA TGTACATTCTTACAAAGCTCCTACTTTCTGTGATTACTGTGGTGAGATGCTCTGGGGATTGGTACGTC AAGGACTGAAATGTGAAGGCTGTGGATTAAATTACCATAAACGATGTGCCTTCAAGATTCCAAATAAC TGTAGTGGAGTAAGAAAGAGACGTCTGTCAAATGTATCTTTACCAGGACCCGGCTCTCAGTTCCAAG ACCCCTGAGCCTGAAATATGTAGCCCTTCCAGTGAAGAGTCAATGTCCACCAGGAACCAAGTAAGA GAATTCCTTCTTGGAGTGGTCGCCCAATCTGGATGGAAAAGATGGTAATGTGCAGAGTGAAAGTTCCA CACATTTGCTGTTCACTCTTACACCCGTCCACGATATGTCACTACTGCAAGCGGTTACTGAAAGG CCTCTTTTCGCCAAGGAATGCAGTGTAAAGATTGCAAATTCACCTGCCATAAACGCTGTGCATCAAAG TACCAAGAGACTGCCCTGGAGAGGTTACTTTCAATGGAGAACCCTCCAGTCTGGGAACAGATACAGAT ATACCAATGGATATTGACAATAATGACATAAATAGTGATAGTAGTCGGGGTTTGGATGACACAGAAGA GCCATCACCCCCAGAAGATAAGATGTTCTTCTTGGATCCATCTGATCTCGATGTGGAAGAGATGAAG AAGCGGTAAACAATCAGTCCATCAACAAGCAATAATATTCGGCTAATGAGGGTGTACAATCCATC AAGCACACAAAGAGGAAGAGCAGCACAAATGGTGAAGGAAGGGTGGATGGTCCATTACACAGCAGGA TAACCTGAGAAAGAGGCATTATTGGAGACTTGACAGCAATGTCTAACATTTATTCAGAAATGAATCTG GATCAAAGTATTATAAGGAAATTCACCTTTCAGAAATTCCTCGCATATCTTACCACGAGATTTCACA AACATTTACAAAGGCAGCAATCCACACTGTTTGTGAAATCATTACTGATACTATGGTATACTTCGTTGG TGAGAACAATGGGGACAGCTCTCATAATCCTGTTCTTCTGCTGCCACTGGAGTTGGACTTGATGTAGCAC AGAGCTGGGAAAAGCAATTCGCCAAGCCCTCATGCTGTTACTCTCAAGCAAGTGTGTTGCACTTCT CCAGGGCAAGGGAAAGATCACAAAGATTGTCTACAAGTATCTCTGTATCTAATTGTCAGATTCAGGA GAATGTGGATATCAGTACTGTTTACCAGATCTTTCAGATGAGGTGCTTGGTTTCAGGCCAGTTTGGCA TCGTTTATGGAGGAAACATAGAAAGACTGGGAGGGATGTGGCTATTAAAGTAATTGATAAGATGAGA TTCCCCACAAACAAGAAAGTCAACTCCGTAACGAAGTGGCTATTTTACAGAATTTGCACCATCCTGG GATGTGTAACCTGGTAATGTATGTTTGAACCCAGAACGAGTCTTGTAGTAATGGAAGCTGCATG GAGATATGTTGGAATGATTCATCCAGTGAGAAAAGCCGGCTTCCAGAACGAATTACTAAATTCATG GTCACACAGATACTTGTGCTTTGAGGAATCTGCATTTTAAGAATATTGTGCACTGTGATTAAAGCC AGAAAATGTGCTGCTTGCATCAGCAGAGCCATTTCTCAGGTGAAGCTGTGTGACTTTGGATTGAC GCATCATTTGGTGAAAAGTCACTCAGGAGATCTGTGGTAGGAACCTCCAGCATACTTAGCCCTGAAAGTT CTCCGGAGCAAAGGTTACAACCGTTCCCTAGATATGTGGTCAGTGGGAGTTATCATCTATGTGAGCCT CAGTGGCACATTTCTTTTAAATGAGGATGAAGATATAAATGACCAATCCAAAATGCTGCATTTATGT ACCCACCAAAATCCATGGAGAGAAATTTCTGGTGAAGCAATTGATCTGATAAACAATCTGCTTCAAGTG AAGATGAAAACCTGTTACAGTGTGACAAATCTCTAGTCATCCCTGGCTACAGGACTTACAGACTTG GCTTGACCTTAGAGAAATTTGAAACTCGCATTTGGAGAACGTTACATTACACATGAAAGTGATGATGCTC GCTGGGAAATACATGCATACACATAACCTTGTATACCCAAAGCACTTCATTATGGCTCCTAATCCA GATGATATGGAAGAAGATCCTCATCATCACCACCATCACTAAGCGGCCGCTTTTCAATC			
NOV4d, Protein Sequence	CG197755-04	SEQ ID NO: 20	890 aa MW at 100469.6kD
MSANNSPPSAQKSVLPATPAVLPAASPCSSPKTGLSARLSNGSFSAPSLTNSRGSVHTVSFLLQIGL TRESVTIEAQELSLSAVKDLVCSIVYQKFPCEGFFGMYDKILLFRHDMNSENILQLITSADEIHEGDL VEVVLALATVEDFQIRPHTLYVHSYKAPTFCDYCGEMLWGLVRQGLKCEGCGLNHYHKRAFKIPMNC SGVRKRRLSNVSLPGPGLSVPRPLQPEYVALPSESHVHQEPSKRIPSWSGRPIWMEKVMCMRVKVP TFVAVHSYTRPTICQYCKRLLKGLFRQGMQCKDKFNCHKRCASKVPRDCLGEVTFNGEPSSSLGTDTDI PMDIDNNDINSRSSRLDDTEEPSPPEDKMFLLDPSDLDERDEEAVKTIISPSTSNNIPLMRVVQSIK HTKRKSSSTMVKEGWMVHYTSRDNLKRHRWRLDSKCLTLFQNESGSKYKEIPLSEILRISSPRDFTN ISQGSNPHCFEIIITDMVYFVGENNGDSSHNPLAATGVGLDVAQSWEKAIQALMPVTPQASVCTSP GQGDHDKDLSTISVSNQCIIQENVDISTVYQIFADEVLGSGQFGIVYGGKHRTGRDVAIKVIDKMRFP PTKQESQLRNEVAILQNLHHPGIVNLECMFETPERVVFVMEKLGDMLEMI LSSEKSRLPERITKFMV TQILVALRNLHFKNIVHCDLKPENVLASAEPFPQVKLCDFGFARIIGESFRRSVVGTPAYLAPEVL RSKGYNRSLDMWSVGVIIVVSLSGTFPFNEDEDINDQIQNAAFMYPPNPWREISGEAIDLINLLQVK MRKRYSDVKSLSHPWLQDYQTWLTLREFETRIGERYITHESDDARWEIHAYTHNLVYPKHFIMAPNP DMEEDP			

- 5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table D4.

Table D4. Comparison of the NOV4 protein sequences.

NOV4a	-----MSANNSPPSAQKSVLPTAIPAVLPAASPCSSPKTGLSARLSNGSFSAPSLTN
NOV4b	MGHHHHHHSANNSPPSAQKSVLPTAIPAVLPAASPCSSPKTGLSARLSNGSFSAPSLTN
NOV4c	MGHHHHHHSANNSPPSAQKSVLPTAIPAVLPAASPCSSPKTGLSARLSNGSFSAPSLTN
NOV4d	-----MSANNSPPSAQKSVLPTAIPAVLPAASPCSSPKTGLSARLSNGSFSAPSLTN
NOV4a	SRGSVHTVSFLLQIGLTRESVTIEAQELSLSAVKDLVCSIVYQKFPECGFFGMYDKILLF
NOV4b	SRGSVHTVSFLLQIGLTRESVTIEAQELSLSAVKDLVCSIVYQKFPECGFFGMYDKILLF
NOV4c	SRGSVHTVSFLLQIGLTRESVTIEAQELSLSAVKDLVCSIVYQKFPECGFFGMYDKILLF
NOV4d	SRGSVHTVSFLLQIGLTRESVTIEAQELSLSAVKDLVCSIVYQKFPECGFFGMYDKILLF
NOV4a	RHDMNSENILQLITSADEIHEGDLVEVVL SALATVEDFQIRPHTLYVHSYKAPTFCDYCG
NOV4b	RHDMNSENILQLITSADEIHEGDLVEVVL SALATVEDFQIRPHTLYVHSYKAPTFCDYCG
NOV4c	RHDMNSENILQLITSADEIHEGDLVEVVL SALATVEDFQIRPHTLYVHSYKAPTFCDYCG
NOV4d	RHDMNSENILQLITSADEIHEGDLVEVVL SALATVEDFQIRPHTLYVHSYKAPTFCDYCG
NOV4a	EMLWGLVRQGLKCEGCGLN YHKRCAFKIPNNCSGVRKRRLSNVSLPGPGLSVPRPLQPEY
NOV4b	EMLWGLVRQGLKCEGCGLN YHKRCAFKIPNNCSGVRKRRLSNVSLPGPGLSVPRPLQPEY
NOV4c	EMLWGLVRQGLKCEGCGLN YHKRCAFKIPNNCSGVRKRRLSNVSLPGPGLSVPRPLQPEY
NOV4d	EMLWGLVRQGLKCEGCGLN YHKRCAFKIPNNCSGVRKRRLSNVSLPGPGLSVPRPLQPEY
NOV4a	VALPSEESHVHQEPSKRIPSWSGRPIWMEKMVMCRVKVPHTFAVHSYTRPTICQYCKRLL
NOV4b	VALPSEESHVHQEPSKRIPSWSGRPIWMEKMVMCRVKVPHTFAVHSYTRPTICQYCKRLL
NOV4c	VALPSEESHVHQEPSKRIPSWSGRPIWMEKMVMCRVKVPHTFAVHSYTRPTICQYCKRLL
NOV4d	VALPSEESHVHQEPSKRIPSWSGRPIWMEKMVMCRVKVPHTFAVHSYTRPTICQYCKRLL
NOV4a	KGLFRQGMQCKDCKFNCHKRCASKVPRDCLGEVTFNGEPSSLGTDTDIPMDIDNNDINSD
NOV4b	KGLFRQGMQCKDCKFNCHKRCASKVPRDCLGEVTFNGEPSSLGTDTDIPMDIDNNDINSD
NOV4c	KGLFRQGMQCKDCKFNCHKRCASKVPRDCLGEVTFNGEPSSLGTDTDIPMDIDNNDINSD
NOV4d	KGLFRQGMQCKDCKFNCHKRCASKVPRDCLGEVTFNGEPSSLGTDTDIPMDIDNNDINSD
NOV4a	SSRGLDDTEEPSPPEDKMFFLDPSDL DVERDEEAVKTISPSTSNNIPLMRVVQSIKHTKR
NOV4b	SSRGLDDTEEPSPPEDKMFFLDPSDL DVERDEEAVKTISPSTSNNIPLMRVVQSIKHTKR
NOV4c	SSRGLDDTEEPSPPEDKMFFLDPSDL DVERDEEAVKTISPSTSNNIPLMRVVQSIKHTKR
NOV4d	SSRGLDDTEEPSPPEDKMFFLDPSDL DVERDEEAVKTISPSTSNNIPLMRVVQSIKHTKR
NOV4a	KSSTMVKEGWMVHYTSRDNLKRHYWR L DSKCLTLFQNESGSKYYKEIPLSEILRISSPR
NOV4b	KSSTMVKEGWMVHYTSRDNLKRHYWR L DSKCLTLFQNESGSKYYKEIPLSEILRISSPR
NOV4c	KSSTMVKEGWMK-----
NOV4d	KSSTMVKEGWMVHYTSRDNLKRHYWR L DSKCLTLFQNESGSKYYKEIPLSEILRISSPR
NOV4a	DFTNISQGSNPHCFEIITDTMVYFVG ENNGDSSHNPVLAATGVGLDVAQSWEKAIRQALM
NOV4b	DFTNISQGSNPHCFEIITDTMVYFVG ENNGDSSHNPVLAATGVGLDVAQSWEKAIRQALM
NOV4c	-----
NOV4d	DFTNISQGSNPHCFEIITDTMVYFVG ENNGDSSHNPVLAATGVGLDVAQSWEKAIRQALM
NOV4a	PVTPQASVCTSPGQGKD HKDLSTSISVSNCQIQENVDISTVYQIFADEVLGSGQFGIVYG
NOV4b	PVTPQASVCTSPGQGKD HKDLSTSISVSNCQIQENVDISTVYQIFADEVLGSGQFGIVYG
NOV4c	-----DLSTSISVSNCQIQENVDISTVYQIFADEVLGSGQFGIVYG
NOV4d	PVTPQASVCTSPGQGKD HKDLSTSISVSNCQIQENVDISTVYQIFADEVLGSGQFGIVYG
NOV4a	GKHKRTGRDVAIKVIDKMRFP TKQESQLRNEVAILQNLHHPGIVNLECMFETPERVFVVM
NOV4b	GKHKRTGRDVAIKVIDKMRFP TKQESQLRNEVAILQNLHHPGIVNLECMFETPERVFVVM

NOV4c GKHRKTGRDVAIKVIDKMRFP TKQESQLRNEVAILQNLHHPGIVNLECMFETPERVFVVM
 NOV4d GKHRKTGRDVAIKVIDKMRFP TKQESQLRNEVAILQNLHHPGIVNLECMFETPERVFVVM

 NOV4a EKLHGDMLEMILSSEKSRLPERITKFMVTQILVALRNLHFKNIVHCDLKPENVLLASAEP
 NOV4b EKLHGDMLEMILSSEKSRLPERITKFMVTQILVALRNLHFKNIVHCDLKPENVLLASAEP
 NOV4c EKLHGDMLEMILSSEKSRLPERITKFMVTQILVALRNLHFKNIVHCDLKPENVLLASAEP
 NOV4d EKLHGDMLEMILSSEKSRLPERITKFMVTQILVALRNLHFKNIVHCDLKPENVLLASAEP

 NOV4a FPQVKLCDFGFARIIGEKSFRRSVVGTPAYLAPEVLRSGYNRSLDMWSVGVIYVSLSG
 NOV4b FPQVKLCDFGFARIIGEKSFRRSVVGTPAYLAPEVLRSGYNRSLDMWSVGVIYVSLSG
 NOV4c FPQVKLCDFGFARIIGEKSFRRSVVGTPAYLAPEVLRSGYNRSLDMWSVGVIYVSLSG
 NOV4d FPQVKLCDFGFARIIGEKSFRRSVVGTPAYLAPEVLRSGYNRSLDMWSVGVIYVSLSG

 NOV4a TFPFNEDEDINDQIQNAAFMYPPNPWREISGEAIDLINLLQVKMRKRYSDVKSLSHPWL
 NOV4b TFPFNEDEDINDQIQNAAFMYPPNPWREISGEAIDLINLLQVKMRKRYSDVKSLSHPWL
 NOV4c TFPFNEDEDINDQIQNAAFMYPPNPWREISGEAIDLINLLQVKMRKRYSDVKSLSHPWL
 NOV4d TFPFNEDEDINDQIQNAAFMYPPNPWREISGEAIDLINLLQVKMRKRYSDVKSLSHPWL

 NOV4a QDYQTWLDLREFETRIGERYITHESDDARWEIHAYTHNLVYPKHFIMAPNPDDMEEDP
 NOV4b QDYQTWLDLREFETRIGERYITHESDDARWEIHAYTHNLVYPKHFIMAPNPDDMEEDP
 NOV4c QDYQTWLDLREFETRIGERYITHESDDARWEIHAYTHNLVYPKHFIMAPNPDDMEEDP
 NOV4d QDYQTWLDLREFETRIGERYITHESDDARWEIHAYTHNLVYPKHFIMAPNPDDMEEDP

 NOV4a (SEQ ID NO: 14)
 NOV4b (SEQ ID NO: 16)
 NOV4c (SEQ ID NO: 18)
 NOV4d (SEQ ID NO: 20)

- 5 Pfam analysis predicts that the NOV4a protein contains the domains shown in the Table D5.

Table D5. Domain Analysis of NOV4a			
Pfam Domain	NOV4a Match Region	Identities/ Similarities for the Matched Region	Expect Value
DAG_PE-bind	155..204	31/51 (61%) 42/51 (82%)	3.7e-18
DAG_PE-bind	272..321	23/51 (45%) 44/51 (86%)	7.4e-20
PH	417..532	27/116 (23%) 83/116 (72%)	8.5e-10
Kdo	579..753	48/227 (21%) 111/227 (49%)	0.14
Pkinase	576..832	98/278 (35%) 195/278 (70%)	1.2e-70

Example D3. Expression Profile of the Human PRKCN Gene (CG197755-01)

- 10 The protocol for quantitative expression analysis is disclosed in Example Q9.

Expression of gene CG197755-01 was assessed using the primer-probe set Ag8180, described in Table D6. Results of the RTQ-PCR runs are shown in Table D7.

5 Table D6. Probe Name Ag8180

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aaacagtttagtagtatttgagaccagtagtatg-3'	33	5041	82
Probe	TET-5'-aatcttacccttagccctcccaacca-3'- TAMRA	26	5079	83
Reverse	5'-gaggtatttatgggtagatagtagacagcata-3'	30	5106	84

Table D7. Panel 5 Islet

Column A - Rel. Exp.(%) Ag8180, Run 323591161			
Tissue Name	A	Tissue Name	A
97457 Patient-02go adipose	12.1	94709 Donor 2 AM - A adipose	52.9
97476 Patient-07sk skeletal muscle	0.0	94710 Donor 2 AM - B adipose	30.4
97477 Patient-07ut uterus	9.0	94711 Donor 2 AM - C adipose	18.2
97478 Patient-07pl placenta	4.2	94712 Donor 2 AD - A adipose	50.7
99167 Bayer Patient 1	2.5	94713 Donor 2 AD - B adipose	71.2
97482 Patient-08ut uterus	2.6	94714 Donor 2 AD - C adipose	65.5
97483 Patient-08pl placenta	13.0	94742 Donor 3 U - A Mesenchymal Stem Cells	14.2
97486 Patient-09sk skeletal muscle	4.8	94743 Donor 3 U - B Mesenchymal Stem Cells	18.0
97487 Patient-09ut uterus	3.9	94730 Donor 3 AM - A adipose	46.3
97488 Patient-09pl placenta	11.7	94731 Donor 3 AM - B adipose	52.9
97492 Patient-10ut uterus	18.3	94732 Donor 3 AM - C adipose	54.7
97493 Patient-10pl placenta	12.9	94733 Donor 3 AD - A adipose	94.0
97495 Patient-11go adipose	10.2	94734 Donor 3 AD - B adipose	100.0
97496 Patient-11sk skeletal muscle	10.7	94735 Donor 3 AD - C adipose	31.6
97497 Patient-11ut uterus	31.0	77138 Liver HepG2untreated	44.8
97498 Patient-11pl placenta	0.0	73556 Heart Cardiac stromal cells (primary)	18.4
97500 Patient-12go adipose	21.0	81735 Small Intestine	38.7
97501 Patient-12sk skeletal muscle	50.3	72409 Kidney Proximal Convoluted Tubule	44.8
97502 Patient-12ut uterus	24.3	82685 Small intestine Duodenum	20.7
97503 Patient-12pl placenta	10.7	90650 Adrenal Adrenocortical adenoma	32.3
94721 Donor 2 U - A Mesenchymal Stem Cells	41.5	72410 Kidney HRCE	27.0
94722 Donor 2 U - B Mesenchymal Stem Cells	28.9	72411 Kidney HRE	14.4
94723 Donor 2 U - C Mesenchymal Stem Cells	42.9	73139 Uterus Uterine smooth muscle cells	35.4

5 **Panel 5 Islet Summary:** Highest expression of PRKCN this gene was detected in fully differentiated adipocytes (CT=30.7). Based on RTQ-PCR data PRKCN expression is induced upon adipocyte differentiation. Similar expression pattern has been reported for PPAR gamma in literature [Lowell BB. 1999 PPARgamma: an essential regulator of adipogenesis and modulator of fat cell function. Cell 99(3):239-42, PMID: 10555139]. This finding further supports the hypothesis about connection
10 between PRKCN and PPAR gamma.

Example D4. Biochemistry/Cell Line Expression/ Screening Assay Formulation

Assays for screening for antibody therapeutics or small molecule drugs targeting human PRKCN can be formulated utilizing the non-exhaustive list of cell lines that express the PRKCN gene
15 from the RTQ-PCR results shown above. To assay the enzymatic activity of PRKCN the measurement[s] of ^{32}P incorporation to generic protein kinase substrate like syntide 2, myelin basic protein could be utilized. Lipid-independent onstitutively active mutant PRKCN (A161E) could be used in the assay.

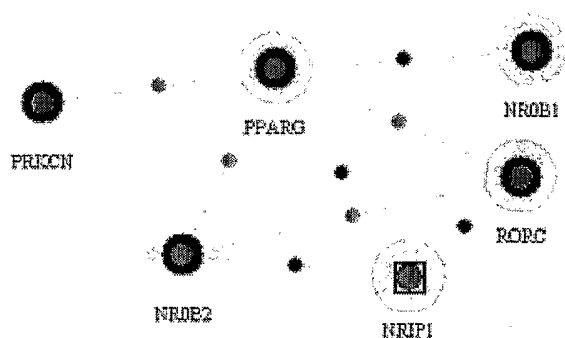
Example D5. Pathways Relevant to the Etiology and Pathogenesis of Obesity and/or Diabetes

Protocol for PathCalling® is disclosed in Example Q10.

CuraGen Corporation's PathCalling® data showed that PRKCN interacted with DNA-binding domain of PPAR gamma (Figure D1). Table D8 summarizes the amino acid positions of the bait and
25 prey polypeptides used to detect this novel interaction. The DNA-binding domain of PPAR gamma contains at least one phosphorylation site (S112). It has been demonstrated that phosphorylation of PPAR gamma at S112 inhibits its transcriptional activity (Adams et al., 1997 J Biol Chem. 272(8):5128-32). The interaction of PRKC nu with PPAR gamma suggested that PRKC nu may inhibit PPAR gamma function by phosphorylation. PPAR gamma also interacted with a network of
30 nuclear hormone receptors (RORC, NR1P1, NROB1, NROB2) that may be involved in modulation of PPAR gamma functions (Figure D1). For instance, it has been shown that NR0B2 acts as an endogenous enhancer of PPAR gamma (Nishizawa et al., 2002. J Biol Chem;277(2):1586-92).

Figure D1.

35



5

Table D11. Yeast Two-hybrid Interaction Information

Interaction Frame	PPAR gamma Interaction Domain (amino acid positions)	PRKC nu (amino acid positions)	Number of Yeast Colonies Observed
3 (+)	Prey: 1-298	Bait: 295-381	1

5 **E. NOV5 --HUMAN FATTY ACID DESATURASE 2 - LIKE PROTEINS, NUCLEIC ACIDS
NCODING THE SAME & METHODS OF USE THEREOF**

The de novo unsaturation of fatty acids is regulated by individual desaturase enzymes that introduce double bonds between defined carbons of the fatty acyl chain. Fatty Acid Desaturase 2 (delta 6 desaturase, FADS2) is known to catalyze the first step in biosynthesis of polyunsaturated fatty acids (PUFA), e.g. the desaturation of linoleic and alpha linolenic acids (Horrobin et al., Am J Clin Nutr. 1993 May; 57(5 Suppl):732S-736S; discussion 736S-737S). Several PUFAs have significant effect on adipogenesis, insulin sensitivity and beta-oxidation via regulation of the activity of important nuclear hormone receptors such as PPAR alpha, NF-kB, SREBP1c and LXR (Jump et al., Curr Opin Lipidol. 2002 Apr;13(2):155-64). Although many literature reports agree that FADS2 might play a role in metabolic disorders, there is significant disagreement about positive or negative contribution of FADS2 to obesity and diabetes. Some literature repots demonstrated the increase in FADS2 expression in genetically obese animals (Hughes et al., Biochem J 1985 Jan 15;225(2):307-13) and the decrease in FADS2 expression and activity in animal models of type 2 diabetes (Rimoldi et al., 2001; 283(2):323-6), the other argues that the effect is not significant (Brown et al., Lipids 2000 Dec;35(12):1319-23). Coguen et al. demonstrated that treatment with fish oil known to have beneficial effect on diabetes/obesity causes significant inhibition of FADS2 (Goguen et al., Curr Med Res Opin. 2002;18 Suppl 1:s58-74; Garg et al., Biochem J 1988 Jan 15;249(2):351-6). Moreover recently, it has been shown that delta6 desaturase index activity is negatively correlated with insulin resistance in liver and muscle (Vessby et al., Ann N Y Acad Sci. 2002 Jun;967:183-95).

The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with various diseases or pathologies. The proteins and related proteins that are similar to them, are encoded by a cDNA and/or by genomic DNA. The proteins, polypeptides and their cognate nucleic acids were identified by the inventors in certain cases. In particular, the FADS2 protein encoded by CG184446-01 and any variants, thereof, are suitable as diagnostic markers, targets for an antibody therapeutic and targets for small molecule drugs. The inventors discover the up-regulation of FADS2 gene in liver of both genetically obese and diet-induced obese animals. In addition, FADS2 gene was found to be up-regulated in obese hyperglycemic animals. The inventors argue that up-regulation of FADS2 observed in animals with obese/diabetic phenotype originated from different condition indicate the direct contribution of the enzyme to the development and progression of obesity/diabetes. Moreover, the inventors showed that FADS2 gene is significantly up-regulated in liver from patients with Type II Diabetes compared to healthy subjects. The inventors further propose that FADS2 inhibition would lead to changes of cellular PUFA composition which would promote the transcriptional activity of PPAR alpha, nuclear hormone receptor involved in lipid oxidation, and on other hand attenuate the transcription activity function SREBP1c, LXR and NF-KB known to be involved in adipogenesis and insulin resistance. In summary, an antagonist of FADS2 would decrease lipid storage and increase lipid burning, promoting lean and insulin sensitive phenotype. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify such therapeutic antibodies and/or therapeutic small molecules.

5 In particular the invention relates to the use of FADS2 protein, gene or antibody thereto as a target in diagnostic applications and/or target for small molecule drugs and antibody therapeutics. The inventors have discovered that FADS2 is up-regulated in both genetic and diet-induced obesity animal models. In addition, FADS2 is up-regulated in obese hyperglycemic animals. The data suggests that up-regulation of FADS2 directly contributes to obese/diabetic phenotype. In a particular embodiment of the invention, FADS2 is a target for screening. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify FADS2 antagonist therapeutic antibodies and/or therapeutic small molecules beneficial in the treatment of obesity and/or diabetes.

15 Not to be limited by a particular mechanism of action, the inventors nevertheless have discovered that inhibition of FADS2 has beneficial effects for treating obesity and/or diabetes by acting in many metabolic tissues, including pancreas, adipose, adrenal gland, thyroid, skeletal muscle, heart, liver, colon and small intestine.

Specifically, in liver inventors have found that FADS2 gene is up-regulated in several animal model of obesity and diabetes. Inventors have further demonstrated the up-regulation of FADS2 gene in liver of patient with Type II Diabetes compared to healthy subjects. These findings indicate the negative role of FADS2 in peripheral metabolism. Inhibition of FADS2 resulted in change of cellular polyunsaturated fatty acid composition, e.g. decrease in Δ^6 -C18:3/C18:2 ratio, increase in linoleic acid and decrease in arachidonic acid level, would favorably modulate the activity of important transcription factors mediated adipogenesis, lipid oxidation and insulin sensitivity. The inventors have proposed that inhibition of FADS2 would facilitate lipid burning associated and attenuate lipid synthesis causing the decrease in lipid storage and improve insulin sensitivity.

Therefore, an antagonist of FADS2 is useful for the treatment of obesity and/or diabetes.

Discovery Process

The following sections describe the study design(s) used to identify the FADS2-encoded protein and any variants, thereof, as being suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for treatment of Obesity and Diabetes.

Example E1. Mouse Dietary – Induced Obesity Study

35 Protocol for Mouse Dietary – Induced Obesity Study is disclosed in Example Q1.

The predominant cause for obesity in clinical populations is excess caloric intake. This so-called diet-induced obesity (DIO) is mimicked in animal models by feeding high fat diets of greater than 40% fat content. The DIO study was established to identify the gene expression changes contributing to the development and progression of diet-induced obesity. In addition, the study design seeks to identify the factors that lead to the ability of certain individuals to resist the effects of a high fat diet and thereby prevent obesity. The sample groups for the study were selected from C57BL/6J mice and had body weights +1 S.D. (sd1), + 4 S.D. (sd4) and + 7 S.D. of the chow-fed controls (below). In addition, the biochemical profile of the + 7 S.D. mice revealed a further stratification of these animals into mice that retained a normal glycemic profile in spite of obesity (ngsd7) and mice that demonstrated hyperglycemia (hgsd7). Tissues examined included hypothalamus, brainstem, liver, retroperitoneal white adipose tissue (WAT), epididymal WAT, brown adipose tissue (BAT),

- 5 **gastrocnemius muscle (fast twitch skeletal muscle) and soleus muscle (slow twitch skeletal muscle).**
The differential gene expression profiles for these tissues should reveal genes and pathways that can be used as therapeutic targets for obesity.

Results

- 10 A fragment of the mouse FADS2 gene was initially found to be upregulated by **2.1** fold in the liver of mice fed a high fat diet who reach 7 standard deviations (ngsd7) of body weight and mice fed a high fat diet who in addition to reaching 7 standard deviations of body weight are hyperglycemic (hgsd7) relative to chow fed mice using CuraGen's GeneCalling® method (disclosed in Example Q7) of differential gene expression. A differentially expressed mouse gene fragment migrating, at
15 approximately **171** nucleotides in length (shown in Table E1) was identified as a component of the mouse FADS2 cDNA. The method of competitive PCR was used for confirmation of the gene assessment. The electropherographic peaks corresponding to the gene fragment of the mouse FADS2 were ablated when a gene-specific primer (shown in Table E1) competed with primers in the linker-adaptors during the PCR amplification. The peaks at **171.5** nt in length were ablated in the
20 sample from both the liver.

Table E1. Competitive PCR primer for the Mouse FADS2:
The sequence of the **171** nucleotide-long gene fragment (from 2196 to 2366 band size: 171) and the gene-specific primers used for competitive PCR are indicated on the cDNA sequence of the mouse FADS2 and are shown below in bold. The gene-specific primers at the 5' and 3' ends of the fragment are underlined. (gene length is 3109, only region from 1715 to 2846 shown)

1715 GCTGCCTCAC ATTCCCCGGG CTTGCATCCT TTCTTGTTAT TCCTCTTCCC
TCACCCCTAC

1775 CTA~~CT~~CCATT CTCATGGAGT TCCCCCTTG TCAGCCCTGA CCCCATCCCC
TCTGCCTCCC

1835 AGTCCCTTCT CCCAGAGAGC CGGAGAGGTG GTTGTCTCAG GTGGTTACGC
GTCTGCCTTT

1895 AACTTCTGCA CCTAAAGATT CCTGGTATGA GATTCTGCTC CTGGCAGCCT
AGTCCCCACC

1955 CCCTTCTGTC CCACAGACAC TCTGGGGGCC TGTGGAACAG GAGTCCAGGA
CTTGGCTCTA

2015 CACTCCCTGT CAGCATTATG GTGTAGTCAT TTGGGTGAAG ACATGCTTCG
TTCTGAGTGT

2075 CTTTCCCATG GCTGGGATTG GTACTTGAGG CTTCTTTCCA GCTGCCAAGG
ACCACCTTAG

2135 ACTTTGAGGA CCAAGGAAAA GTTCTCTCAT GCCCACCATC ACAACTCCAG
ACCATAGGGT

2195 TTGTACAGGT TCTATTTCAT CATCTACCCA ACTCCAGAAA CCAAAGGGAG
TGTTCTTGT

2255 GCGGGCTCAG GAGAGGCAGT AGCCAGGTTG AAGGAGGGAC TGACCAGCCT
GGAACTTAA

2315 **CCTCTCCAC TACAGTCAGG TGCTTCTGAG TAAAAGGCC CAAGATTCTA**
GAAGGATCAG

2375 ATCTTTTCC AGATCCTCTG GCCAGTTCCA GATGCTTGGT CTAGAGAATG
TGGACCTTGC

2435 AGGAGAGCCC CTGAGCAGCT CTCCCAGGA TAGAGCTGCC AATCTAGCCA
GCACTTGGCT

2495 TCTTTTGGC TCATCCCTGT CTTTACAGAT GTTTGTTTGC TTGTTTGGT TAGCTGTTGT

2555 GTTAACATTT CCTACAGAAT TCTTCTTAAT AGGAGAGTAG CAGAGGGCTG
GCTGGGTCTT

2615 AGGACAATCC ATCTTTACTA GATGTCCCTT GCCCTTCAGG TAGAGCTAGG
CTTCGCCTTC

2675 AAGGCCACAG TATACGCTGT GGCTTGAGGT TGCTGCTATC AGAAAGGCAG
GAAAAACAGG

2735 GTAGGGTGTG AGGTCTCAGG AGGGCTGTTG GGGGTACCCA GTGAGGCTCA
AGCAGGAGTG

2795 TAGAGGGAAG AGGAGGCTGA GGGCATGAGG GAAATGTGAG GGCAGGTGGG AG
(SEQ ID NO: 85)

5

Example E2. Genetically Obese Mice vs. Genetically Lean Mice Study (MB.04)

Protocol for Genetically Obese Mice vs. Genetically Lean Mice Study is disclosed in Example Q6.

10

A large number of mouse strains have been identified that differ in body mass and composition. The AKR and NZB strains are obese, the SWR, C57L and C57BL/6 strains are of average weight whereas the SM/J and Cast/Ei strains are lean. Understanding the gene expression differences in the major metabolic tissues from these strains will elucidate the pathophysiological basis for obesity. These specific strains of rat were chosen for differential gene expression analysis because quantitative trait loci (QTL) for body weight and related traits had been reported in published genetic studies. Tissues included whole brain, skeletal muscle, visceral adipose, and liver.

15

MB.04 AKR vs. SWR liver

20

A fragment of the mouse FADS2 gene was initially found to be upregulated by 1.9 fold in the liver of obese mice (AKR) relative to normal weight mice (SWR) using CuraGen's GeneCalling® method of differential gene expression. A differentially expressed mouse gene fragment migrating, at approximately 287.8, 171.5, 182.9, 203.7 nucleotides in length was identified as a component of the mouse FADS2 cDNA. The method of TraPping was used for confirmation of the gene assessment. A table depicting the fragment lengths, dysregulation, TraPping score and actual sequence data is displayed below (Table E2). The numerator in the score represents the number of nucleotides matched in the gene fragment. The actual nucleotide sequence is displayed in the column labeled "Fragment TraP Data". The denominator in the score represents the total number of

25

5 TraP nucleotides available for this fragment with the actual nucleotide sequence presented in the column labeled "Predicted Trap Nucleotide Sequence". A score of 3/3 or 4/4 is treated with high confidence that the band belongs to that gene. The trapping data shows that these differentially expressed gene fragments in Discovery Study MB.04 are from the mouse FADS2. Protocol for TrapPing is disclosed in Example Q7, Part (C).

10

Table E2. The results of the trapping data that show that the bands found to be dysregulated in Discovery Study MB.04 study are from the mouse FADS2.

Length of fragment in nucleotides	Fold Difference	Score	Fragment TraP Data	Predicted Trap Nucleotide Sequ
287.8	1.8	4/4	ATCT	ATC CT
171.5	1.9	3/3	ATGG	TGG
182.9	1.9	4/4	GGGA	G AG GA
203.7	1.8	4/4	TGTT	TGTT

15

MB.04 AKR vs. C57

A fragment of the mouse FADS2 gene was initially found to be upregulated by 1.9 fold in the liver of obese mice (AKR) relative to normal weight mice (C57) using CuraGen's GeneCalling® method of differential gene expression. A differentially expressed mouse gene fragment migrating, at approximately 287.8, 171.5, 182.9, 203.7, 203.8, 203.1 nucleotides in length was identified as a component of the mouse FADS2 cDNA. The method of TraPing was used for confirmation of the gene assessment. A table depicting the fragment lengths, dysregulation, TraPing score and actual sequence data is displayed below (Table E3). The numerator in the score represents the number of nucleotides matched in the gene fragment. The actual nucleotide sequence is displayed in the column labeled "Fragment TraP Data". The denominator in the score represents the total number of TraP nucleotides available for this fragment with the actual nucleotide sequence presented in the column labeled "Predicted Trap Nucleotide Sequence". A score of 3/3 or 4/4 is treated with high confidence that the band belongs to that gene. The trapping data shows that these differentially expressed gene fragments in Discovery Study MB.04 are from the mouse FADS2.

25

Table E3. The results of the trapping data that show that the bands found to be dysregulated in Discovery Study MB.04 study are from the mouse FADS2.

Length of fragment in nucleotides	Fold Difference	Score	Fragment TraP Data	Predicted Trap Nucleotide Sequ
287.8	1.8	4/4	ATCT	ATC CT
171.5	1.6	3/3	ATGG	TGG
182.9	1.9	4/4	GGGA	G AG GA
203.7	1.7	4/4	TGTT	TGTT
203.8	2.7	1/1	TGTT	- - - T
203.1	2.1	2/2	TTCC	TT - -

30

MB.04 C57 vs. Cast/Ei liver

A fragment of the mouse FADS2 gene was initially found to be upregulated by 2.0 fold in the liver normal weight mice (C57) relative to lean mice (Cast/Ei) using CuraGen's GeneCalling® method

5 of differential gene expression. A differentially expressed mouse gene fragment migrating, at approximately **287.9 and 47.0** nucleotides in length was identified as a component of the mouse FADS2 cDNA. The method of TraPping was used for confirmation of the gene assessment. A table depicting the fragment lengths, dysregulation, TraPping score and actual sequence data is displayed below (Table E4). The numerator in the score represents the number of nucleotides matched in the gene fragment. The actual nucleotide sequence is displayed in the column labeled "Fragment TraP Data" (Table E4). The denominator in the score represents the total number of TraP nucleotides available for this fragment with the actual nucleotide sequence presented in the column labeled "Predicted Trap Nucleotide Sequence". A score of 3/3 or 4/4 is treated with high confidence that the band belongs to that gene. The trapping data shows that these differentially expressed gene fragments in Discovery Study MB.04 are from the mouse FADS2.

Table E4. The results of the trapping data that show that the bands found to be dysregulated in Discovery Study MB.04 study are from the mouse FADS2.

Length of fragment in nucleotides	Fold Difference	Score	Fragment TraP Data	Predicted Trap Nucleotide Sequence
287.9	2.0	4/4	ATCT	ATC CT
47.0	2.6	4/4	CCTT	CCTT

Example E3. Human CG184446-01 Sequence Identification

The sequence of Human FADS2 gene (Acc. No. CG184446-01) was derived by laboratory screening of cDNA library. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were 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 protocol for identification of human sequence(s) is disclosed in Example Q8.

An alignment of the protein sequences of the human (CG184446-01), rat and mouse homologs of the FADS2 is shown in Table E5.

Table E5. Protein alignment (ClustalW) of the human (CG184446-01) (SEQ ID NO: 22), mouse (SEQ ID NO: 86) and rat (SEQ ID NO: 87) orthologs of the FADS2.

FADS2_rat	MGKGGNQEGSTELQAPMPTRWEEIQKHNLRTDRWLVIDRKVYNVTKWSQRHPGGHRVI
CG184446-01	MGKGGNQEGCAERREVSVPPTFSWEEIQKHNLRTDRWLVIDRKVYNVTKWSIQHPGGQRVI
FADS2_mouse	MGKGGNQEGSTERQAPMPTRWEEIQKHNLRTDRWLVIDRKVYNVTKWSQRHPGGHRVI
FADS2_rat	GHYSGEDATDAFR AFHLDLDFVGKFLKPLLIGELAPEEPSLDROKSSQITEDFRALKKTA
CG184446-01	GHYSGEDATDAFR AFHLDLDFVGKFLKPLLIGELAPEEPSLDROKSSQITEDFRALKKTA
FADS2_mouse	GHYSGEDATDAFR AFHLDLDFVGKFLKPLLIGELAPEEPSLDROKSSQITEDFRALKKTA
FADS2_rat	EDMNLFKTNHLFFFLLLSHIIVMESIAWFILSYFGNGWIFPTITAFVLATSQAQAGWLQH
CG184446-01	EDMNLFKTNHLFFFLLLSHIIVMESIAWFILSYFGNGWIFPTITAFVLATSQAQAGWLQH
FADS2_mouse	EDMNLFKTNHLFFFLLLSHIIVMESIAWFILSYFGNGWIFPTITAFVLATSQAQAGWLQH
FADS2_rat	DYGHLSVYKKS IWNHIVHKFVIGHLKGASANWWNHRHFQHHAKPNI FHKDPDIKSLHVFV
CG184446-01	DYGHLSVYKKS IWNHIVHKFVIGHLKGASANWWNHRHFQHHAKPNI FHKDPDVNMLHVFV
FADS2_mouse	DYGHLSVYKKS IWNHIVHKFVIGHLKGASANWWNHRHFQHHAKPNI FHKDPDIKSLHVFV
FADS2_rat	LGEWQPLEVGKKLKYLPYNHQQEYFFLIGFP LLIPMYFQYQIIMTMI RRRDWVDLAWAI
CG184446-01	LGEWQPLEYGGKKLKYLPYNHQQEYFFLIGFP LLIPMYFQYQIIMTMI VHNWVDLAWAI
FADS2_mouse	LGEWQPLEYGGKKLKYLPYNHQQEYFFLIGFP LLIPMYFQYQIIMTMI SRRDWVDLAWAI
FADS2_rat	SYVIA RFFFTYIPFYGILGALVFLNFI RFLSHWFVWVTQMNHIVMEIDLHRYDWFS SQL
CG184446-01	SYVIA RFFFTYIPFYGILGALVFLNFI RFLSHWFVWVTQMNHIVMEIDLHRYDWFS SQL
FADS2_mouse	SYVIA RFFFTYIPFYGILGALVFLNFI RFLSHWFVWVTQMNHIVMEIDLHRYDWFS SQL
FADS2_rat	AATCNVEQSFFNDWFSGLNFIQIEHHLFPTMPRHNLHKIAPLVKSLCAKHGIEYQEKPLL
CG184446-01	AATCNVEQSFFNDWFSGLNFIQIEHHLFPTMPRHNLHKIAPLVKSLCAKHGIEYQEKPLL
FADS2_mouse	AATCNVEQSFFNDWFSGLNFIQIEHHLFPTMPRHNLHKIAPLVKSLCAKHGIEYQEKPLL
FADS2_rat	RALLDIVSSLKKS GELWLDAYLHK
CG184446-01	RALLDIVSSLKKS GELWLDAYLHK
FADS2_mouse	RALLDIVSSLKKS GELWLDAYLHK

5

The laboratory cloning was performed using one or more of the methods summarized in Example Q8. The NOV5 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table E6.

Table E6. NOV5 Sequence Analysis			
NOV5a,	CG184446-01	SEQ ID NO: 21	3149 bp
DNA Sequence		ORF Start: ATG at 151	ORF Stop: TGA at 1483
AGGGGGCGCGGTGGGAGGAGTAGGAGAAGACAAAGCCGAAAGCGAAGAGGGCCCCGGGCTGCACACAC CGGCTGGGAGGCGAGCCGCTCTGTGCAGCGAGCAGCCGGCGCGGGGAGGCCGAGTGCACGGGGCGTCA AGTCGCGAGGCGAGCATG GGG GAAGGGAGGGAACCAAGGCGAGGGGGCCGCCGAGCGCGAGGTGTCTGGTG CCCACCTTCAGCTGGGAGGAGATTGAGAAGCATAACCTGCGCACCGACAGGTGGCTGGTCAATGACCG CAAGGTTTACAACATCACCAAAATGGTCCATCCAGCACCCGGGGGGGCCAGCGGGTTCATCGGGCACTACG CTGGAGAAGATGCAACGGATGCCTTCCGCGCCTTCCACCTGACCTGGAATTCGTGGGCAAGTTCCTG AAACCCCTGCTGATTGGTGAACCTGGCCCGGAGGAGCCAGCCAGGACCAAGGCAAGAAGTCAAAGAT CACTGAGGACTTCCGGGGCCCTGAGGAAGACGGCTGAGGACATGAACCTGTTCAAGACCAACCAGTGT TCTTCTCTCTCTCTGGCCACATCATCGCCCTGGAGAGCATTGCATGGTTCAGTGTCTTTTACTTT GGCAATGGCTGGATTCTTACCCTCATACGGCCCTTTGTCTTGTCTACCTTTCAGGCCCAAGCTGGATG GCTGCAACATGATTATGGCCACCTGTCTGTCTACAGAAACCAAGTGAACCAACCTTGTCCACAAAT TCGTCAATTGGCCACTTAAAGGGTGCCTCTGCCAATCGTGGAAATCATCGCCACTTCCAGCACCAAGCC AAGCCTAACATCTTCCACAAGGATCCCGATGTGAACATGCTGCACGTGTTTGTCTGGGCGAATGGCA GCCCATCGAGTACGGCAAGAAGAAGCTGAAATACCTGCCCTACAATCACCAGCACGAATACTTCTTCC TGATTGGGCGCGCGCTGCTCATCCCCATGTATTTCCAGTACCAGATCATCATGACCATGATCGTCCAT AAGAAGCTGGGTGGACCTGGCCCTGGGCGCTCAGCTACTACATCCGGTCTTTCATCACCTACATCCGTTT CTACGGCATCTCTGGGAGCCCTCCTTTTCTCAACTTCATCAGGTTCCTGGAGAGCCACTGGTTTGTGT GGGTACACAGATGAATCACATCGTCATGGAGATTGACCAGGAGGCCCTACCGTGACTGGTTTCAGTAGC CAGCTGACAGCCACCTGCAACGTGGAGCAGTCCCTTCTTCAACGACTGGTTCACTGGACACCTTAACCT CCAGATTGAGCACCACCTCTTCCCCACCATGCCCGGCACAACTTACACAAGATCGCCCCGCTGGTGA AGTCTCTATGTGCCAAGCATGGCATTTGAATACCCAGGAGAAGCCGCTACTGAGGGCCCTGCTGGACATC ATCAGGTCCCTGAAGAAGTCTGGGAAGCTGTGGCTGGACGCTACCTTACAAAT TGA AGCCACAGCCC CCGGGACACCGTGGGGAAGGGGTGCAGGTGGGGTGATGGCCAGAGGAATGATGGGCTTTTGTCTGAG GGGTGTCCGAGAGGCTGGTGTATGCACTGCTCAGGACCCCATGTTGGATCTTTTCTCCCTTTCTCTCTC TCCTTTTCTCTTCACTCTCCCCATAGCACCTTGCCCTCATGGGACCTGCCCTCCCTCAGCCGTCAC GCCATCAGCCATGGCCCTCCCAAGTGCCTTCTAGCCCTTCTTCCAAGGAGCAGAGAGGTGGCCACCGG GGGTGGCTCTGTCTTACCTCCACTCTCTGCCCTTAAAGATGGGAGGAGACCAGCGGTCCATGGGTCTG GCCTGTGAGTCTCCCTTGCAGCCTGGTCACTAGGCATCACCCCGCTTTGGTTCTTTCAGATGCTCTT GGGGTTTCATAGGGGAGGTTCCTAGTCGGGAGGGGCCCTGACCTCCCGGCTGGCTTCACTCTCCCT GACGGCTGCCATTGGTCCACCTTTCTCATAGAGAGGCGTCTTGTGTACAAAGCTCGGGTCTCCCTCT GCAGCTCGGTTAAGTACCCGAGGCTCTCTTAAGATGTCCAGGGCCCCAGGCCCCGGGGCACAGCCAG CCCAAACTTGGGCGCTGGAAGAGTCTTCCACCCATCACTAGAGTGTCTTGACCCTGGGCTTTTCACTG			

GGCCCCATTCCACCGCCTCCCCAATTGAGCCTGTGACCTTGGGACCAAGGGGGAGTCCCTCGTCTC TTGTGACTCAGCAGAGGCAGTGGCCACGTTAGGGAGGGGCGGCTGGCCTGGAGGCTCAGCCCACCC TCCAGCTTTTCTCAGGGTGTCTGAGGTCCAAGATTCTGGAGCAATCTGACCTTCTCCAAAGGCTC TGTATCAGCTGGGCAGTGCCAGCCAATCCCTGGCCATTGGCCCCAGGGGACGTGGGCCCTGCAGGC TGCAGGAGGGCACTGGAGCTGGGAGGTCTCGTCCCAGCCCTCCCCATCTCGGGGCTGCTGTGTGGACG GCGCTGCCTCAGGCACTCTCCTGTCTGAACCTGCCCTTACTGTGTTTAACTGTGCTCCAGGATGCA TTCTGATAGGAGGGGCGGCAGGGCTGGGCCTTGTGACAACTTGCCCTTTCACCACATGGCCCTTGCCCTC GGTGGCCCTGACTGTCAGGGAGGGCCAGGGAGGCAGAGCGGGAGGGAGTCTCAGGAGGAGGCTGCCCT GAGGGCTGGGGAGGGGGTACCTCATGAGGACCAGGTGGAGCTGAGAAGAGGAGGAGGTGGGGGCTG GAGGTGCTGGTAGCTGAGGGGACGGGCAAGTGAGAGGGGAGGGAGGGAAGTCTGGGAGGATCCTGAG CTGCTGTTGCACTAATCACTAATCAGTTCTTAGATTAGGGGAAGGGCAGGCACCAACAACCTCAG AATGGGGGCTTTTCGGGGAGGGCGCCTAGTCCCCCAGCTTAAGCAGCCAGGAGGACCTGCATCTAA GCATCTGGGTTGCCATGGCAATGGCATGCCCCCAGCTACTGTATGCCCCGACCCCGCAGAGGCAG AATGAACCCATAGGGAGCTGATCGTAATGTTTATCATGTTACTTCCCCACCCCTACATTTTGTAAAT AAAAATAAGGAATTTTATTCTC				
NOV5a, Protein Sequence	CG184446-01	SEQ ID NO: 22	444 aa	MW at 52258.9kD
MGKGGNQGEAAEREVSVPFWSWEEIQKHNLRTDRWLVIDRKVYNITKWSIQHPGGQRVIGHYAGEDA TDAFRAFHPDLEFVGKFLKPLLIGELAPEEPSQDHGKNSKITEDFRALRKTAEDMNLFKTNHVFLLLL LAHIIALESIAWFVTFYFGNGWIPTLITAFVLATSQAQAGWLQHDYGHLSVYRKPKNWNLVHKFVIGH LKGASANWNNHRHFQHHAKPNIFHKDPDVNMLHVFLVGEWQPIEYGKKKLKYLPLYNHQHEFFLIGPP LLIPMYFYQYQIIMTMIVHKNWVDLAWAVSYIRFFITYIPFYGILGALLFLNFIREFLESHWFVWVTQM NHIVMEIDQEAYRDWFSQLTATCNVEQSFFNDWFSGHLNFQIEHHLFPTMPRNLHKIAPLVKSLCA KHGIEYQEKPLLRALLDIIRSLKSGKLWLDAYLHK				
NOV5b, DNA Sequence	CG184446-02	SEQ ID NO: 23	1351 bp	
		ORF Start: ATG at 7	ORF Stop: TGA at 1339	
TCCACCATGGGGAAGGGAGGGAACAGGGCGAGGGGCGCCGAGCGCAGGTGTGGTGGCCACCTT CAGCTGGGAGGAGATTGAGAAGCATAACCTGCGCACCGACAGGTGGCTGGTCAATGACCGCAAGGTTT ACAACATCACCAATGGTCCATCCAGCACCCGGGGGGCCAGCGGGTCATCGGGCACTACGCTGGAGAA GATGCAACGGATGCCCTTCCGCGCCTTCCACCTGACCTGGAATTCTGTGGGCAAGTCTTGAACCCCT GCTGATTGTTGAATGGCCCCGAGGAGCCAGCCAGGACCACGCAAGAACTCAAAGATCACTGAGG ACTTCCGGGGCCCTGAGGAAGACGGCTGAGGACATGAACCTGTTCAAGACCAACCACGTGTTCTTCTC CTCCTCTGGCCACATCATCGCCCTGGAGAGCATTGCATGGTCTACTGTCTTTTACTTTGGCAATGG CTGGATTCTACCCCTCATCAGGCGCTTGTCTTGTCTACCTCTCAGGCCCCAAGCTGGATGGCTGCAAC ATGATTATGGCCACCTGTCTGTCTACAGAAAACCAAGTGAACACCTTGTCCACAAATTCGTCAAT GGCCACTTAAAGGGTGCTCTGCCAATGGTGAATCATCGCCACTTCCAGCACCACGCAAGCCTAA CATCTTCCACAAGGATCCCGATGTGAACATGCTGCACGTGTTTGTCTGGGCGAATGGCAGCCCATCG AGTACGGCAAGAAAGCTGAAATACCTGCCCTACAATCACCAGCACGAATACTTCTTCTGATTGGG CCGCGCTGCTCATCCCCATGTATTTCCAGTACCAGATCATCATGACCATGATCGTCCATAAGAACTG GGTGGACCTGGCCTGGGCGTCAGCTACTACATCGGTTCTTCATCACCTACATCCCTTCTACGGCA TCTTGGGAGCCCTCTTTTCTCAACTTCATCAGGTTCCTGGAGAGCCACTGGTTTGTGTGGGTGACA CAGATGAATCACATCGTCATGGAGATTGACAGGCGCTACCGTGACTGGTTCAGTAGCCAGCTGAC AGCCACCTGCAACGTGGAGCAGTCTTCTTCAACGACTGGTTCAGTGGACACCTTAACCTCCAGATTG AGCACCACCTCTTCCCCACCATGCCCCGGCACAACCTTACACAAGATCGCCCCGCTGGTGAAGTCTCTA TGTGCCAAGCATGGCATTTGAATACCAGGAGAAGCCGCTACTGAGGGCCCTGCTGGACATCATCAGGTC CCTGAAGAAGTCTGGGAAGCTGTGGCTGGACGCCTACCTTACAAAATGAGCGGCGCTCG				
NOV5b, Protein Sequence	CG184446-02	SEQ ID NO: 24	444 aa	MW at 52258.9kD
MGKGGNQGEAAEREVSVPFWSWEEIQKHNLRTDRWLVIDRKVYNITKWSIQHPGGQRVIGHYAGEDA TDAFRAFHPDLEFVGKFLKPLLIGELAPEEPSQDHGKNSKITEDFRALRKTAEDMNLFKTNHVFLLLL LAHIIALESIAWFVTFYFGNGWIPTLITAFVLATSQAQAGWLQHDYGHLSVYRKPKNWNLVHKFVIGH LKGASANWNNHRHFQHHAKPNIFHKDPDVNMLHVFLVGEWQPIEYGKKKLKYLPLYNHQHEFFLIGPP LLIPMYFYQYQIIMTMIVHKNWVDLAWAVSYIRFFITYIPFYGILGALLFLNFIREFLESHWFVWVTQM NHIVMEIDQEAYRDWFSQLTATCNVEQSFFNDWFSGHLNFQIEHHLFPTMPRNLHKIAPLVKSLCA KHGIEYQEKPLLRALLDIIRSLKSGKLWLDAYLHK				

A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table E7.

Table E7. Comparison of the NOV5 protein sequences.

NOV5a	MGKGGNQGEAAEREVSVPFWSWEEIQKHNLRTDRWLVIDRKVYNITKWSIQHPGGQRV
NOV5b	MGKGGNQGEAAEREVSVPFWSWEEIQKHNLRTDRWLVIDRKVYNITKWSIQHPGGQRV

NOV5a GHYAGEDATDAFRAFHPDLEFVGKFLKPLLIGELAPEEPSQDHGKNSKITEDFRALRKTA
 NOV5b GHYAGEDATDAFRAFHPDLEFVGKFLKPLLIGELAPEEPSQDHGKNSKITEDFRALRKTA

NOV5a EDMNLFKTNHVFFLLLLAHIIALESIAWFTVFYFGNGWIPTLITAFVLATSQAQAGWLQH
 NOV5b EDMNLFKTNHVFFLLLLAHIIALESIAWFTVFYFGNGWIPTLITAFVLATSQAQAGWLQH

NOV5a
 DYGHLSVYRKPKWNHLVHKFVIGHLKGASANWWNHRHFQHHAKPNIFHKDPDVNMLHVFV
 NOV5b
 DYGHLSVYRKPKWNHLVHKFVIGHLKGASANWWNHRHFQHHAKPNIFHKDPDVNMLHVFV

NOV5a LGEWQPIEYGKKKLKYL PYNHQHEYFFLIGPPLLIPMYFQYQIIMTMIVHKNWVDLAWAV
 NOV5b LGEWQPIEYGKKKLKYL PYNHQHEYFFLIGPPLLIPMYFQYQIIMTMIVHKNWVDLAWAV

NOV5a SYIIRFFITYIPFYGILGALLFLNFIRFLESHWFVWVTQMNHIVMEIDQEAYRDWFSSQL
 NOV5b SYIIRFFITYIPFYGILGALLFLNFIRFLESHWFVWVTQMNHIVMEIDQEAYRDWFSSQL

NOV5a TATCNVEQSFFNDWFSGHLNFQIEHHLFPTMPRHNHKKIAPLVKSLCAKHGIEYQEKPLL
 NOV5b TATCNVEQSFFNDWFSGHLNFQIEHHLFPTMPRHNHKKIAPLVKSLCAKHGIEYQEKPLL

NOV5a RALLDIIRSLKKSGKLWLDAYLHK
 NOV5b RALLDIIRSLKKSGKLWLDAYLHK

NOV5a (SEQ ID NO: 22)
 NOV5b (SEQ ID NO: 24)

- 5 Pfam analysis predicts that the NOV5a protein contains the domains shown in the Table E8.

Table E8. Domain Analysis of NOV5a			
Pfam Domain	NOV5a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Cyt-b5	20..95	26/80 (32%) 63/80 (79%)	9.2e-20
FA_desaturase	165..407	62/300 (21%) 196/300 (65%)	6.2e-56

Example E4. Human FADS2 Gene Variants and SNPs

- 10 Protocol for SNP identification is disclosed in Example Q11.

The variants of the human FADS2 were obtained from direct cloning and/or public databases. In addition to the human version of the FADS2 identified as mouse homolog that is differentially expressed in the experimental study, other variants have been identified by direct

15 sequencing of cDNAs derived from many different human tissues and/or alignment with sequences in public databases. Several amino acid-changing and non-amino acid-changing cSNPs were identified and are shown in Table E9 below, wherein the following abbreviations include: UCP=uncharged polar, NP=non polar, A=acidic, and B=basic. Those cSNPs with ID "cgsp" refer to CuraGen proprietary SNPs, whereas those labeled "hsnp" are from public databases. In one embodiment, a

20 variant used for screening purposes, is CG184446-01.

5 Table E9. CG184446-01 SNPs

SNP ID	SNP MATCH POSI	ALLELES	SNP MATCH POSI ON ORF	ORIG AA	NEW AA	AA CHANGE TYPE
cgsp:13382178	251	A/G	34	Asp	Gly	A to UCP
cgsp:cg111.3	1140	C/T	330	Glu	Glu	Silent
cgsp:cg111.13727	1159	C/T	337	Val	Ile	Conservative
cgsp:cg111.2	1173	A/G	341	Asn	Asn	Silent
cgsp:13382211	2094	C/T	UTR	N/A	N/A	Non-Coding(UTR)
cgsp:13382212	2271	C/T	UTR	N/A	N/A	Non-Coding(UTR)
cgsp:cg111.11850	3004	C/-	UTR	N/A	N/A	Non-Coding(UTR)

10 Example E5. Expression Profile of the Human FADS2 Gene (CG184446-01)

Protocol for quantitative gene expression analysis is disclosed in Example Q9. Protocol for CuraChip™ analysis is disclosed in Example Q12.

15 Expression of gene CG184446-01 was assessed using the primer-probe set Ag8306 for RTQ-PCR (Table E10). Approximately 137 samples of RNA from tissues obtained from surgically dissected disease- and non-disease tissues, and treated and untreated cell lines, were used to generate labeled nucleic acid which was hybridized to PTG Chip 1.2. An oligo, optg2_1202495 (Table E11) that corresponds to CG18446-01 on the PTG Chip 1.2 was analyzed for its expression profile. Results of the RTQ-PCR runs and the CuraChip data are shown in Table E12 and E13, respectively.

20 Table E10. Probe Name Ag8306

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-atcccttttctacggcatcct-3'	20	1081	88
Probe	TET-5'-agccctccttttccctcaacttcac-3'- TAMRA	26	1104	89
Reverse	5'-cacacaaaccagtggctctc-3'	20	1138	90

Table E11. Oligo optg2_1202495.

5'CTGTTGCAGTCTAACCCACTAATCAGTTCT 3' (SEQ ID NO: 91)

Table E12. Human Metabolic

Column A - Rel. Exp.(%) Ag8306, Run 325701630			
Tissue Name	A	Tissue Name	A
137857_psoas-AA.M.Diab.-hi	0.4	139523_pancreas-HI.M.Norm-hi BMI-31	3.7

BMI-6			
135760_psoas-HI.M.Diab.-hi BMI-21	0.2	139520_pancreas-CC.M.Norm-hi BMI-29	9.3
134827_psoas-CC.M.Diab.-hi BMI-4	0.3	142744_pancreas-HI.M.Norm-med BMI-35	0.6
137860_psoas-AA.M.Diab.-med BMI-8	0.2	139545_pancreas-AA.M.Norm-med BMI-47	16.6
137834_psoas-CC.M.Diab.-med BMI-2	0.9	139531_pancreas-AA.M.Norm-med BMI-37	6.8
137828_psoas-CC.M.Diab.-med BMI-1	1.3	137871_pancreas-CC.M.Norm-med BMI-26	0.1
135763_psoas-HI.M.Diab.-med BMI-23	0.6	139541_pancreas-HI.M.Norm-low BMI-41	4.6
142740_psoas-AS.M.Diab.-low BMI-20	1.8	139537_pancreas-CC.M.Norm-low BMI-40	4.3
134834_psoas-AA.M.Diab.-low BMI-17	0.5	139533_pancreas-CC.M.Norm-low BMI-39	2.0
137850_psoas-AS.M.Norm-hi BMI-34	1.0	137845_pancreas-AS.M.Norm-low BMI-28	0.1
135769_psoas-HI.M.Norm-hi BMI-31	0.3	143530_small intestine-AA.M.Diab.-hi BMI-6	0.7
135766_psoas-AA.M.Norm-hi BMI-25	0.4	143529_small intestine-CC.M.Diab.-hi BMI-4	2.3
142746_psoas-AA.M.Norm-med BMI-37	4.8	143538_small intestine-HI.M.Diab.-med BMI-23	9.3
142745_psoas-HI.M.Norm-med BMI-35	1.2	143531_small intestine-AA.M.Diab.-med BMI-8	6.9
137855_psoas-AA.M.Norm-med BMI-47	0.3	143528_small intestine-CC.M.Diab.-med BMI-2	4.8
137844_psoas-CC.M.Norm-med BMI-26	0.2	143537_small intestine-HI.M.Diab.-low BMI-22	5.4
142742_psoas-CC.M.Norm-low BMI-40	6.6	143535_small intestine-AS.M.Diab.-low BMI-20	3.6
137873_psoas-AS.M.Norm-low BMI-28	0.1	143534_small intestine-AA.M.Diab.-low BMI-17	0.3
137853_psoas-HI.M.Norm-low BMI-41	0.8	143544_small intestine-AS.M.Norm-hi BMI-34	2.6
135775_psoas-CC.M.Norm-low BMI-39	0.1	143543_small intestine-HI.M.Norm-hi BMI-31	0.3
137858_diaphragm-AA.M.Diab.-hi BMI-6	0.7	143542_small intestine-CC.M.Norm-hi BMI-29	4.3
135772_diaphragm-AS.M.Diab.-hi BMI-9	1.0	143539_small intestine-AA.M.Norm-hi BMI-25	8.4
135761_diaphragm-HI.M.Diab.-hi BMI-21	1.3	143548_small intestine-AA.M.Norm-med BMI-47	2.1
134828_diaphragm-CC.M.Diab.-	0.2	143547_small intestine-AA.M.Norm-med	1.9

hi BMI-4		BMI-37	
137835_diaphragm-CC.M.Diab.-med BMI-2	0.9	143540_small intestine-CC.M.Norm-med BMI-26	2.2
135764_diaphragm-HI.M.Diab.-med BMI-23	0.6	143550_small intestine-CC.M.Norm-low BMI-40	0.1
134835_diaphragm-AA.M.Diab.-low BMI-17	0.6	143549_small intestine-CC.M.Norm-low BMI-39	8.1
142738_diaphragm-CC.M.Norm-hi BMI-29	0.2	143546_small intestine-HI.M.Norm-low BMI-41	1.6
139517_diaphragm-AS.M.Norm-hi BMI-34	1.0	143525_hypothalamus-HI.M.Diab.-hi BMI-21	17.8
137848_diaphragm-HI.M.Norm-hi BMI-31	0.2	143515_hypothalamus-CC.M.Diab.-hi BMI-4	1.8
137843_diaphragm-AA.M.Norm-hi BMI-25	0.6	143513_hypothalamus-AA.M.Diab.-hi BMI-6	4.8
137879_diaphragm-AA.M.Norm-med BMI-47	0.5	143507_hypothalamus-AS.M.Diab.-hi BMI-9	6.5
137872_diaphragm-CC.M.Norm-med BMI-26	0.2	143506_hypothalamus-CC.M.Diab.-med BMI-1	9.2
135773_diaphragm-HI.M.Norm-med BMI-35	0.9	143505_hypothalamus-HI.M.Diab.-med BMI-23	5.3
139542_diaphragm-HI.M.Norm-low BMI-41	0.5	143509_hypothalamus-AA.M.Diab.-low BMI-17	13.1
137877_diaphragm-CC.M.Norm-low BMI-39	0.1	143508_hypothalamus-CC.M.Diab.-low BMI-13	7.2
137874_diaphragm-AS.M.Norm-low BMI-28	0.2	143503_hypothalamus-AS.M.Diab.-low BMI-20	7.0
141340_subQadipose-AA.M.Diab.-hi BMI-6	1.2	143522_hypothalamus-HI.M.Norm-hi BMI-31	0.3
137836_subQadipose-HI.M.Diab.-hi BMI-21	0.1	143516_hypothalamus-AS.M.Norm-hi BMI-34	0.0
135771_subQadipose-AS.M.Diab-hi BMI-9	0.4	143511_hypothalamus-CC.M.Norm-hi BMI-29	4.8
141329_subQadipose-AA.M.Diab-medbmi-8	0.5	143504_hypothalamus-AA.M.Norm-hi BMI-25	3.0
137862_subQadipose-CC.M.Diab.-med BMI-1	0.2	143517_hypothalamus-AA.M.Norm-med BMI-47	4.4
135762_subQadipose-HI.M.Diab.-med BMI-23	0.1	143514_hypothalamus-HI.M.Norm-med BMI-35	4.2
141338_subQadipose-AS.M.Diab.-low BMI-20	0.2	143521_hypothalamus-AS.M.Norm-low BMI-28	1.8
139547_subQadipose-HI.M.Diab.-low BMI-22	0.8	143512_hypothalamus-CC.M.Norm-low BMI-40	1.0
135757_subQadipose-CC.M.Diab.-low BMI-13	0.7	145454_Patient-25pl (CC.Diab.low BMI.no insulin)	3.3
134832_subQadipose-	0.2	110916_Patient-18pl (HI.Diab.obese.no	1.8

AA.M.Diab.-low BMI-17		insulin)	
141332_subQadipose-HI.M.Norm-hi BMI-31	0.2	110913_Patient-18go (HI.Diab.obese.no insulin)	1.4
135767_subQadipose-CC.M.Norm-hi BMI-29	0.1	110911_Patient-17pl (CC.Diab.low BMI.no insulin)	1.5
135765_subQadipose-AS.M.Norm-hi BMI-34	0.1	110908_Patient-17go (CC.Diab.low BMI.no insulin)	2.1
141339_subQadipose-HI.M.Norm-med BMI-35	0.3	100752_Patient-15sk (CC.Diab.obese.no insulin)	2.3
141334_subQadipose-CC.M.Norm-med BMI-26	1.2	97828_Patient-13pl (CC.Diab.overwt.no insulin)	1.1
139544_subQadipose-AA.M.Norm-med BMI-47	3.4	160114_Patient 27-ut (CC.Diab.obese.insulin)	8.1
137875_subQadipose-AA.M.Norm-med BMI-37	0.3	160113_Patient 27-pl (CC.Diab.obese.insulin)	3.0
141331_subQadipose-AS.M.Norm-low BMI-28	1.3	160112_Patient 27-sk (CC.Diab.obese.insulin)	1.0
137878_subQadipose-HI.M.Norm-low BMI-41	0.1	160111_Patient 27-go (CC.Diab.obese.insulin)	3.2
137876_subQadipose-CC.M.Norm-low BMI-39	0.1	145461_Patient-26sk (CC.Diab.obese.insulin)	0.6
137859_vis.adipose-AA.M.Diab.-hi BMI-6	0.0	145441_Patient-22sk (CC.Diab.low BMI.insulin)	1.0
135770_vis.adipose-AS.M.Diab.-hi BMI-9	0.8	145438_Patient-22pl (CC.Diab.low BMI.insulin)	3.1
135759_vis.adipose-HI.M.Diab.-hi BMI-21	0.0	145427_Patient-20pl (CC.Diab.overwt.insulin)	9.0
143502_vis.adipose-CC.M.Diab.-med BMI-2	0.6	97503_Patient-12pl (CC.Diab.unknown BMI.insulin)	0.2
139510_vis.adipose-AA.M.Diab.-med BMI-8	0.3	145443_Patient-23pl (CC.Non-diab.overwt)	1.4
137861_vis.adipose-CC.M.Diab.-med-1	0.1	145435_Patient-21pl (CC.Non-diab.overwt)	5.2
137839_vis.adipose-HI.M.Diab.-med BMI-23	0.1	110921_Patient-19pl (CC.Non-diab.low BMI)	2.8
139546_vis.adipose-HI.M.Diab.-low BMI-22	0.5	110918_Patient-19go (CC.Non-diab.low BMI)	2.2
137831_vis.adipose-CC.M.Diab.-low BMI-13	0.1	97481_Patient-08sk (CC.Non-diab.obese)	2.8
139522_vis.adipose-HI.M.Norm-hi BMI-31	0.2	97478_Patient-07pl (CC.Non-diab.obese)	1.3
139516_vis.adipose-AS.M.Norm-hi BMI-34	1.0	160117_Human Islets-male, obese	1.8
137846_vis.adipose-CC.M.Norm-hi BMI-29	0.0	145474_PANC1 (pancreas carcinoma) 1	39.8
137841_vis.adipose-	0.0	154911_Capan2 (pancreas adenocarcinoma)	12.9

AA.M.Norm-hi BMI-25			
139543_vis.adipose-AA.M.Norm-med BMI-47	2.0	141190_SW579 (thyroid carcinoma)	7.7
139532_vis.adipose-AA.M.Norm-med BMI-37	0.1	145489_SK-N-MC (neuroblastoma) 1	34.9
139530_vis.adipose-HI.M.Norm-med BMI-35	0.2	145495_SK-N-SH (neuroblastoma) 1	22.5
139539_vis.adipose-HI.M.Norm-low BMI-41	0.1	145498_U87 MG (glioblastoma) 2	3.0
139535_vis.adipose-CC.M.Norm-low BMI-40	1.9	145484_HEp-2 (larynx carcinoma) 1	6.4
137852_vis.adipose-CC.M.Norm-low BMI-39	0.2	145479_A549 (lung carcinoma)	92.7
135768_vis.adipose-AS.M.Norm-low BMI-28	0.1	145488_A427 (lung carcinoma) 2	33.9
141327_liver-CC.M.Diab.-hi BMI-4	0.5	145472_FHs 738Lu (normal lung) 1	64.6
139514_liver-HI.M.Diab.-hi BMI-21	2.3	141187_SKW6.4 (B lymphocytes)	19.9
139526_liver-CC.M.Diab.-med BMI-2	1.6	154644_IM-9 (immunoglobulin secreting lymphoblast)	19.8
139511_liver-AA.M.Diab.-med BMI-8	1.0	154645_MOLT-4 (acute lymphoblastic leukemia derived from peripheral blood)	8.6
137840_liver-HI.M.Diab.-med BMI-23	17.6	154648_U-937 (histiocytic lymphoma)	10.4
137827_liver-CC.M.Diab.-med BMI-1	3.0	154647_Daudi (Burkitt's lymphoma)	0.1
137838_liver-HI.M.Diab.-low BMI-22	8.9	145494_SK-MEL-2 (melanoma) 2	10.7
135758_liver-CC.M.Diab.-low BMI-13	11.4	141176_A375 (melanoma)	1.1
139519_liver-CC.M.Norm-hi BMI-29	1.2	154642_SW 1353 (humerus chondrosarcoma)	4.0
139518_liver-AA.M.Norm-hi BMI-25	0.6	141179_HT-1080 (fibrosarcoma)	6.2
137849_liver-AS.M.Norm-hi BMI-34	2.6	145491_MG-63 (osteosarcoma) 1	15.3
137847_liver-HI.M.Norm-hi BMI-31	2.4	141186_MCF7 (breast carcinoma)	4.8
142741_liver-AA.M.Norm-med BMI-37	1.1	141193_T47D (breast carcinoma)	4.6
141341_liver-HI.M.Norm-med BMI-35	0.2	154641_BT-20 (breast carcinoma)	0.2
141335_liver-CC.M.Norm-med BMI-26	0.5	141175_293 (kidney transformed with adenovirus 5 DNA)	3.1
139540_liver-HI.M.Norm-low	0.9	141182_HUH hepatoma 1	27.9

BMI-41			
139534_liver-CC.M.Norm-low BMI-39	1.3	141184_HUH7 hepatoma 1	34.2
139521_liver-AS.M.Norm-low BMI-28	0.5	145478_HT1376 (bladder carcinoma)	28.7
141328_pancreas-CC.M.Diab.-hi BMI-4	22.8	145481_SCaBER (bladder carcinoma)	2.0
139525_pancreas-AS.M.Diab.-hi BMI-9	2.1	141192_SW620 (lymph node metastasis, colon carcinoma) 2	1.8
137856_pancreas-AA.M.Diab.-hi BMI-6	0.1	141180_HT29 (colon carcinoma) 1	0.0
137837_pancreas-HI.M.Diab.-hi BMI-21	0.2	141188_SW480 (colon carcinoma) 1	4.0
141337_pancreas-CC.M.Diab.-med BMI-2	1.0	154646_CAOV-3 (ovary adenocarcinoma)	45.4
139527_pancreas-CC.M.Diab.-med BMI-1	8.8	141194_HeLa (cervix carcinoma)- 2	5.7
139515_pancreas-HI.M.Diab.-med BMI-23	12.9	145482_HeLa S3 (cervix carcinoma) 1	5.1
139512_pancreas-AA.M.Diab.-med BMI-8	12.9	145486_DU145 (prostate carcinoma)	22.8
142739_pancreas-AS.M.Diab.-low BMI-20	4.3	154643_PC-3 (prostate adenocarcinoma)	100.0
139513_pancreas-CC.M.Diab.-low BMI-13	8.0	154649_HCT-8 (ileocecal adenocarcinoma)	1.4
142743_pancreas-AA.M.Norm-hi BMI-25	3.9		

5 Table E13. CuraChip 1.2 analysis.

Sample	Foreground Mean	Background Mean	optg2_1202495
hypothalamus-AA.M.Diab-low-1	4461.09	0.03	5127.81
hypothalamus-AA.M.Norm-hi-1	2743.99	0	4721.12
hypothalamus-AA.M.Norm-med-2	3203.67	1.6	2149.75
hypothalamus-AS.M.Diab-hi-1	2903.35	0.35	5596.71
hypothalamus-AS.M.Diab-low-1	3141.26	4.11	5350.72
hypothalamus-AS.M.Norm-hi-1	3775.62	0.33	337.38
hypothalamus-AS.M.Norm-low-1	3492.22	8.2	2346.96
hypothalamus-CC.M.Diab-hi-1	3640.27	0.66	3484.99

hypothalamus-CC.M.Diab-low-1	2797.6	0.05	4860.27
hypothalamus-CC.M.Diab-med-1	3991.28	0.03	5927.62
hypothalamus-CC.M.Norm-hi-1	2356.93	0	2861.39
hypothalamus-CC.M.Norm-low-2	3572.68	3.94	2468.06
hypothalamus-HI.M.Diab-hi-1	3836.19	0.94	8018.48
hypothalamus-HI.M.Diab-med-1	3858.89	0.05	3438.35
hypothalamus-HI.M.Norm-hi-1	3990.48	1.74	3082.94
liver-AA.M.Diab-med-1	4097.25	0.01	487.01
liver-AA.M.Norm-hi-1	4188.33	0	240.31
liver-AA.M.Norm-med-1	3236.08	0	345.36
liver-AS.M.Norm.-hi-1	2041.19	1.4	2855.1
liver-AS.M.Norm-low-1	3953.34	1.35	252.37
liver-CC.M.Diab.-med-1	1812.72	0	3453.43
liver-CC.M.Diab-med-2	3997.83	0	389.06
liver-CC.M.Norm-hi-1	3863.42	0	421.96
liver-CC.M.Norm-low-1	3377.02	0.08	892.18
liver-CC.M.Norm-med-1	2806.5	0	717.26
liver-CMD-low-1	2243.37	0	10818.72
liver-HI.M.Diab.-low-1	1363.73	0	11053.8
liver-HI.M.Diab.-med-1	2387.37	0	11300.55
liver-HI.M.Diab-hi-1	4026.22	0	448.61
liver-HI.M.Norm-hi-1	1779.79	0	1793.58
liver-HI.M.Norm-low-1	2748.56	0.02	433.03
Mergen brain control	3183.65	0.02	19089.38
pancreas-AA.M.Diab.-hi-1	3367.25	0	4770.78
pancreas-AA.M.Diab-low-1	3125.18	0.02	4970.31
pancreas-AA.M.Diab-med-1	4001.4	0.15	5174.79
pancreas-AA.M.Norm-med-1	3214.77	0	468.09
pancreas-AA.M.Norm-med-2	4044.67	0	4400.36

pancreas-AS.M.Diab-hi-1	3616.61	0	350.99
pancreas-AS.M.Norm-low-1	4813.67	0.69	12901.09
pancreas-CC.M.Diab-hi-1	4143.79	0	8734.88
pancreas-CC.M.Diab-low-1	4368.7	0	5658.25
pancreas-CC.M.Diab-med-1	3003.42	0	265.9
pancreas-CC.M.Diab-med-2	3863.8	0	9143.51
pancreas-CC.M.Norm-hi-1	3523.53	0.04	1919.95
pancreas-CC.M.Norm-low-1	2751.11	0.08	127.15
pancreas-CC.M.Norm-low-2	4727.28	0.01	7433.34
pancreas-HI.M.Diab.-hi-1	3755.94	0.14	6095.78
pancreas-HI.M.Diab-med-1	3909.15	0.39	6698.8
pancreas-HI.M.Norm-hi-1	4935.28	0	619.84
pancreas-HI.M.Norm-low-1	4512.13	0.01	10099.86
Patient-06pl	1059.88	25.62	444.2
Patient-07pl	999.32	27.24	460.11
Patient-08pl	1924.57	43.31	1129.4
Patient-12go	759.57	45.57	72.41
Patient-12pl	1092.94	28.94	110.71
Patient-12sk	1148.71	23.63	60.33
psoas-AA.M.Diab.-hi-1	2126.8	45.03	50.17
psoas-AA.M.Diab.-med-1	1576.95	40.33	72.55
psoas-AA.M.Norm-med-2	1818.53	40.48	230.46
psoas-AAMD-low-1	1220.42	24.72	39.66
psoas-AAMND-hi-1	2995.58	0	2935.46
psoas-AS.M.Diab.-hi-1	2193.88	47.93	296.83
psoas-AS.M.Diab-low-1	1182.9	39.64	1002.45
psoas-AS.M.Norm.-hi-1	1951.61	26.36	129.07
psoas-AS.M.Norm-low-1	1808.85	41.18	131.35
psoas-CC.M.Diab.-low-1	1757.62	26.55	15.65
psoas-CC.M.Diab.-med-1	2267.87	45.76	189.65

psoas-CC.M.Diab.-med-2	1322.17	30.51	113.15
psoas-CC.M.Norm-low-1	1946.32	42.95	16.96
psoas-CC.M.Norm-low-2	1750.59	43.83	7255.04
psoas-CC.M.Norm-med-1	1050.93	40.63	35.59
psoas-CMD-hi-1	1080.96	42.01	4.07
psoas-CMND-low-1	2413.53	42.18	41.93
psoas-HLM.Diab.-low-1	1083.91	30.53	110.62
psoas-HLM.Norm.-low-1	2515.44	37.48	322.29
psoas-HMD-hi-1	1112.34	40.39	1.98
psoas-HMD-med-1	1755.49	26.38	68.93
psoas-HMND-hi-1	2429.06	36.59	67.93
small intestine-AA.M.Diab-hi-1	2413.02	0.01	717.98
small intestine-AA.M.Diab-low-1	1735.47	0.02	585.66
small intestine-AA.M.Diab-med-1	2133.48	0	2292.31
small intestine-AA.M.Norm-hi-1	2695.1	0	3685.58
small intestine-AA.M.Norm-med-1	2692.05	0.03	2372.39
small intestine-AA.M.Norm-med-2	2536.78	0	2648.12
small intestine-AS.M.Diab-low-1	2928.39	1.62	3270.26
small intestine-AS.M.Norm-hi-1	2834.91	0	2664.14
small intestine-AS.M.Norm-low-1	2945.45	0.26	4896.03
small intestine-CC.M.Diab-hi-1	2729.49	0	2593.74
small intestine-CC.M.Diab-med-2	2799.43	0.2	3081.81
small intestine-CC.M.Norm-hi-1	2965.93	0.13	4120.46
small intestine-CC.M.Norm-low-1	2631.59	0	5945.61
small intestine-CC.M.Norm-low-2	1506.5	0	389.18
small intestine-CC.M.Norm-med-1	2506.55	0.08	2367.6
small intestine-HLM.Diab-low-1	2901.51	0.31	2284.91
small intestine-HLM.Diab-med-1	2456.78	0	4275.03
small intestine-HLM.Norm-hi-1	2423	0.02	677.8

small intestine-HI.M.Norm-low-1	2862.37	0	2222.77
subQadipose-AA.M.Diab-hi-1	2318.76	3.54	1323.08
subQadipose-AA.M.Diab-med-1	2367.84	0.05	1816.89
subQadipose-AA.M.Norm-med-1	2020.03	0	518.41
subQadipose-AA.M.Norm-med-2	3086.76	17.63	3705.44
subQadipose-AAMD-low-1	1010.03	0	39.21
subQadipose-AMD-hi-1	2509.99	0.16	589.44
subQadipose-AMND-hi-1	1325.5	0	50.62
subQadipose-AS.M.Diab-low-1	2569.17	0.6	442.28
subQadipose-CC.M.Diab.-med-1	2061.84	0	276.89
subQadipose-CC.M.Diab-hi-1	2526.68	13.89	1092.3
subQadipose-CC.M.Norm-low-1	1567.95	3.04	215.38
subQadipose-CMD-low-1	1607.13	0	591.36
subQadipose-CMND-hi-1	2357.92	0.28	228.59
subQadipose-HI.M.Diab-low-1	2981.4	1.27	334.27
subQadipose-HI.M.Norm-hi-1	2354.58	0.75	923.6
subQadipose-HI.M.Norm-low-1	2319.79	3.48	289.72
subQadipose-HI.M.Norm-med-1	2539.53	0.81	869.77
subQadipose-HMD-med-1	1957.58	0	75.3
vis.adipose-AA.M.Diab-med-1	1845.86	47.71	1272.31
vis.adipose-AA.M.Norm-med-1	1060.42	31.39	182.57
vis.adipose-AA.M.Norm-med-2	1313.8	33.91	993
vis.adipose-AAMD-hi-1	1265.55	41.94	61.71
vis.adipose-AMD-hi-1	2029.17	41.29	597.39
vis.adipose-AMND-low-1	1411.8	41	63.89
vis.adipose-AS.M.Diab.-low-1	1233.6	34.8	69.55
vis.adipose-AS.M.Norm-hi-1	1333.35	43.96	3401.43

vis.adipose-CC.M.Diabe-hi-1	1819	33.99	907.09
vis.adipose-CC.M.Diab-med-2	875.78	41.28	243.67
vis.adipose-CC.M.Norm-low-1	1564.03	38.79	704.72
vis.adipose-CC.M.Norm-low-2	1613.4	40.56	2427.85
vis.adipose-CC.M.Norm-med-1	1959.84	39.23	3116.73
vis.adipose-CMD-low-1	960.41	23.72	151.19
vis.adipose-CMND-low-1	1137.43	42.85	0.97
vis.adipose-HLM.Diab.-med-1	1805.45	30.45	106.01
vis.adipose-HLM.Diab-low-1	1233.26	37.21	247.96
vis.adipose-HLM.Norm-hi-1	1040.98	34.05	406.83
vis.adipose-HLM.Norm-low-1	1067.92	31.3	2.06
vis.adipose-HLM.Norm-med-1	1550.86	39.6	1341.97
vis.adipose-HMD-hi-1	1269.32	40.69	0

- 5 **Summary of human FADS2 gene expression by RTQ-PCR:** FADS2 gene is highly expressed in tumor cell lines and in all metabolic tissues that presented in the panel. From metabolic tissues FADS2 is expressed significantly in liver, pancreas, small intestine and hypothalamus that is in agreement with GeneCalling® data and proposed disease association. Notably, significant up-regulation of FADS2 gene has been detected in liver from patients with Type II Diabetes compared to
- 10 healthy patients (Average ct value for diabetic patients 25.9; average for healthy subjects 27.8 with T-test value 0.017).

- Expression analysis of CG184446-01 by using PTG Chip 1.2:** Significant expression of this gene is seen in metabolic/endocrine related tissues including adipose, small intestine, skeletal muscle, liver, and hypothalamus. Notably, CuraChip analysis also showed that FADS2 gene expressed
- 15 significantly higher in liver from patients with Type II Diabetes compared with healthy subjects (Average intensity value for diabetic patients 5421; average for healthy subjects 883 with T-test value 0.024). Therefore, therapeutic modulation of this gene or its protein product may be useful in the treatment of diabetes and obesity.

20 **Example E6. Biochemistry/Cell Line Expression/ Screening Assay Formulation**

Assays for screening for antibody therapeutics or small molecule drugs targeting human FADS2 can be formulated utilizing the hepatocyte microsomal fraction and the non-exhaustive list of mammalian cell lines that express the FADS2 gene from the RTQ-PCR results shown above. FADS2 (delta 6 desaturase) desaturates essential fatty acids 18:2(n-6) and 18:3(n-3) to form 20:4(n-

5 6) (arachidonic acid) and 22:6(n-3) (docosahexaenoic acid). Inhibition of FADS2 has been shown to decrease in w6-C18:3/C18:2 ratio in all lipid classes (about 50%), increase linoleic acid level and slightly decrease arachidonic acid level. To assay the enzymatic activity of FADS2 the measurement of C16:1 n-10 can be utilized as described by Guillou et al (J Lipid Res. 2003, 44(3):450-4) and D'andrea et al. (Biochem J. 2002, 364(Pt 1):49-55). To assess the selectivity of the compounds, specific inhibitors such as SC-26196 can be used (Hansen-Petrik et al., Cancer Lett. 2002 Jan 25;175(2):157-63).

F. NOV6. HUMAN MONOCARBOXYLATE TRANSPORTER - LIKE PROTEINS, NUCLEIC ACIDS ENCODING THE SAME & METHODS OF USE THEREOF

15 Increased hepatic glucose production is one of the key contributors to the fasting hyperglycemia in type II diabetes. Lactate is a significant source of substrate for gluconeogenesis. There are estimates that 17% of glucose produced by liver gluconeogenesis is derived from lactate. Lactate is also converted to glycogen in the liver. Lactate metabolism is altered in diabetes. Hepatic glucose production from lactate is increased in diabetic humans. In addition, lactate uptake into liver is increased in fasted and postabsorptive diabetic rats. Lactate uptake, as well as pyruvate and ketone body uptake is mediated by a family of monocarboxylate transporters. There are 8 publicly cloned monocarboxylate transporters, SLC16A1, SLC16A2, SLC16A3, SLC16A4, SLC16A5, SLC16A6, SLC16A7 and SLC16A6, in addition to CG56918-02. (Mithieux et al., Diabetes. 2002 Jan; 20 51(1):139-43; Radziuk et al., Diabetes Metab Res Rev. 2001 Jul-Aug;17(4):250-72; Large et al., Diabetes. 1999 Jun;48(6):1251-7; Cusi et al., 1996 Nov;81(11):4059-67; Consoli et al., J Clin Invest. 1990 Dec;86(6):2038-45; Matsuda et al., Metabolism. 2002 Sep;51(9):1111-9; Jackson et al., J Biol Chem. 1996 Jan 12;271(2):861-8).

30 The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with various diseases or pathologies. The proteins and related proteins that are similar to them, are encoded by a cDNA and/or by genomic DNA. The proteins, polypeptides and their cognate nucleic acids were identified by the inventors in certain cases. In particular, the monocarboxylate transporter protein encoded by CG56918-02 and any variants, thereof, are suitable as diagnostic markers, targets for an antibody therapeutic and targets for small molecule drugs. CG56918-02 is the only family member of known eight human monocarboxylate transporters that is highly expressed in liver when compared to the other tissues. A preferred method of the invention is the use of the monocarboxylate transporter for identifying an antagonist that would be beneficial in the treatment of diabetes. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify such therapeutic antibodies and/or therapeutic small molecules.

40 In particular the invention relates to the use of monocarboxylate transporter protein as a diagnostic and/or target for small molecule drugs and antibody therapeutics. The inventors have discovered that CG56918-02 is the only transporter in the monocarboxylate transporter class showing higher expression in liver when compared to other normal tissues. Monocarboxylate transporters mediate the uptake of lactate as well as pyruvate and ketone bodies into the cells. Lactate is a significant source of substrate for liver gluconeogenesis leading to hepatic glucose

5 production. Increased hepatic glucose production is one of the key contributors to the fasting hyperglycemia in type II diabetes. Hepatic glucose production from lactate is increased in diabetic humans. In addition, lactate uptake into liver is increased in diabetic rat liver (Diabetes 48:1251, 1999). Inhibition of lactate/pyruvate uptake in the liver may result in reduced hepatic glucose production. Therefore, inhibition of the novel monocarboxylate transporter would be beneficial to the treatment of diabetes. In a particular embodiment of the invention, monocarboxylate transporter is a target for screening. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify monocarboxylate transporter antagonist, therapeutic antibodies and/or therapeutic small molecules beneficial in the treatment of obesity and/or diabetes.

15 Not to be limited by a particular mechanism of action, the inventors nevertheless have discovered that inhibition of monocarboxylate transporter has beneficial effects for treating diabetes by acting in many metabolic tissues, including adipose, liver, heart, skeletal muscle, adrenal, pituitary, thyroid, and pancreas. Specifically, in liver monocarboxylate transporter activation may lead to increase in lactose uptake that modulates hepatic glucose production. Therefore an antagonist of the novel monocarboxylate transporter may be beneficial for the treatment of diabetes.

Discovery Process

The following sections describe the study design(s) used to identify the monocarboxylate transporter -encoded protein and any variants, thereof, as being suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for treatment of Obesity and Diabetes.

Example F1. Human CG56918-02 Sequence Identification

The sequence of Human monocarboxylate transporter gene (Acc. No. CG56918-02) was derived by laboratory screening of cDNA library. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were 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 protocol for identification of human sequence(s) is disclosed in Example Q8.

An alignment of the protein sequences of the human (CG56918-02), rat and mouse orthologs of the monocarboxylate transporter is shown in Table F1.

35 Table F1. Protein alignment (ClustalW) of the human (CG56918-02) (SEQ ID NO: 26), and rat (SEQ ID NO: 92) and mouse (SEQ ID NO: 93) orthologs of the monocarboxylate transporter.

CG56918-02	MAARRTEPPDGGWGWVYVLSAFFQSALVFGVLRSEGVFFVEFVAAFDEQAARVSWIASIGI
aw917289_m	
bf782328_m	MVRRLEPPDGGWGWVYVLSAFFQSALVFGVLRSEGVFFVEFVAAFDEQAARVSWIASIGI
CG56918-02	AVQQFGSPFDSALSTKFGPRPVYMTGGILAAALOMLLASFATSLTHLYLSIGLLSGSGWAL
aw917289_m	
bf782328_m	AVQQFGSPFDSALSTKLGPRPVYMTGGILAAALOMLLASFATSLTHLYLSIGLLSGSG---
CG56918-02	TFAPTLACLSCYFSRRRSLATGLALTGVGLSSFTFAFFQWLLSHYAWRGSLLIVSALS
aw917289_m	TGVGHSFAFAPLQWLLNHYAWRGALLIVSALS
bf782328_m	
CG56918-02	HLVACGALLRFPPLAEDPAVGGPRAQLTSLLHHGFPLRYTVALTINTGYFIPYHLVAH
aw917289_m	
bf782328_m	HLVACGALLRPLSKTIEDTAVGGPWACITSLLCGGPFLRYTVALTINTGYFIPYHLVAH
CG56918-02	LQDLQWDPLFAAFLLSVMAISDLVGRVVSQWLGDVAPGPVIRLLMLWTTLTGVSIALFPV
aw917289_m	
bf782328_m	LQDLQWDPLFAAFLLSVMAISDLVGRVVSQWLGDVAPGPVIRLLMLWTTLTGVSIALFPV
CG56918-02	AQAPTALVALAVAYGFTSGALPLAFSVLPFLGTRRIYCGLGILQMVESISGLLGFPLS
aw917289_m	
bf782328_m	AQASTTLVVALAVAYGFTSGALTPVAFSVLPFLGTRRIYCGLGILQMVESISGLLGFPLS
CG56918-02	GYLKRDVTONTASFVVAGAFLLSGSCLLTLPHFF-CSSITTSQFQDLVTEALDTKVFPLP
aw917289_m	
bf782328_m	GYLKRDVTONTASFVVAGAFLLSGSCLLTLPHFFSCISLTSRQDLVTEALDTKVFPLP
CG56918-02	KEEGLLED
aw917289_m	
bf782328_m	KEEGLGEN

5

Protein alignment (ClustalW) of the human monocarboxylate transporter (CG56918-02), and human homologs of the monocarboxylate transporter is shown in Table F2.

Table F2. Protein alignment (ClustalW) of the human (CG56918-02), and human homologs of the monocarboxylate transporter.

SLC16A1 (SEQ ID NO:94)
 SLC16A7 (SEQ ID NO:95)
 SLC16A3 (SEQ ID NO:96)
 SLC16A8 (SEQ ID NO:97)
 CG56918-02 (SEQ ID NO:26)
 SLC16A5 (SEQ ID NO:98)
 SLC16A6 (SEQ ID NO:99)
 SLC16A4 (SEQ ID NO:100)
 CG93088-01 (SEQ ID NO:101)
 SLC16A2 (SEQ ID NO:102)

[illegible]

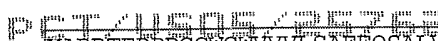
5

5 PCT/US2005/025762 The laboratory cloning was performed using one or more of the methods summarized in

Example Q8. The NOV6 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table F3.

Table F3. NOV6 Sequence Analysis				
NOV6a,	CG56918-02	SEQ ID NO: 25	1294 bp	
DNA Sequence		ORF Start: ATG at 1	ORF Stop: TGA at 1279	
ATG CGCGCGCAGGACAGAGCCCCCGACGGGGGCTGGGGATGGGTGGTGGTGCCTCTCAGCGTTCTTCCA GTCGGCGCTTGTGTTTGGGGTGCCTCCGCTCCTTTGGGGTCTTCTTCGTGGAGTTTGTGGCGGCGTTTG AGGAGCAGGCAGCGCGCGTCTCCTGGATCGCTCCATAGGAATCGCGGTGCAGCAGTTTGGGAGCCCCG GTAGGCAGTGCCCTGAGCACGAAGTTCGGGCCCAGGCCCGTGGTGATGACTGGAGGCATCTTGGCTGC GCTGGGGATGCTGCTCGCCTCTTTTGCTACTTCTTTGACCCACCTATACCTGAGTATTGGGTGCTGT CAGGCTCTGGCTGGGCTTTGACCTTCGCTCCGACCTGGCCTGCCTGCTGTTATTTTCTCGCCGA CGATCCCTGGCCACCGGGCTGGCACTGACAGGCGTGGGCCCTCTCCTCCTCACATTTGCCCCCTTTTT CCAGTGGCTGCTCAGCCACTACGCCTGGAGGGGGTCCCTGCTGCTGGTGTCTGCCCTCTCCCTCCACC TAGTGGCCTGTGGTGTCTCCTCCGCCACCCCTCCCTGGCTGAGGACCTGCTGTGGGTGGTCCCAGG GCCCAACTCACCTCTCTCCATCATAGGCCCTTCTCCGTTACACTGTTGCCCTCACCTGATCAA CACTGGCTACTTCAATCCCTACCTCCACCTGGTGGCCCATCTCCAGGACCTGGATTGGGACCCACTAC CTGCTGCCTTCCCTACTCTCAGTTGTTGCTATTTCTGACCTCGTGGGGCGTGTGGTCTCCGGATGGCTG GGAGATGCAGTCCCAGGGCCTGTGACACGACTCCTGATGCTCTGGACCACCTTGACTGGGGTGTCACT AGCCCTGTTCCCTGTAGCTCAGGCTCCCACAGCCCTGGTGGCTCTGGCTGTGGCCCTACGGCTTCACAT CAGGGGCTCTGGCCCCACTGGCCTTCTCCGTGCTGCCTGAACATAAGGACTAGAAAGGATTTACTGT GGCCTGGGACTGTTGCAGATGGTAGAGAGCATCGGGGGGCTGCTGGGGCCTCCTCTCTCAGGCTACCT CCGGGATGTGACAGGCAACTACACGGCTTCTTTTGTGGTGGCTGGGGCCTTCTCTTTCAGGGAGTG GCATTCTCCTCACCTGCCCCACTTCTTCTGCTCCTCAACTACTACCTCCGGGCCCCAGGACCTTGTA ACAGAAGCACTAGATACTAAAGTTCCCTACCCAAGGAGGACTGGAAGAGGACT GAACTCCACAGAG TC				
NOV6a,	CG56918-02	SEQ ID NO: 26	426 aa	MW at 44916.8kD
Protein Sequence				
MARRTEPPDGGWGVVVLSAFFQSALVFGVLRSGVFFVEFVAAFEEQAARVSWIASIGIAVQQFGSP VGSALSTKFGPRPVMTGGILAAALGMLLASFATSLTHLYLSIGLLSGSGWALTFAPTLACLSCYFSRR RSLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVLSALSLHLVACGALLRPPSLAEDPAVGGPR AQLTSLHHGPFLRYTVALTLINTGYFI PYLHLVAHLQDLWDPLPAAFLLSVVAISDLVGRVVSGLW GDAVPGPVTRLLMLWTTLTGVSLALFPVAQAPTALVALAVAYGFTSGALAPLAFSVLPFLIGTRRIYC GLGLLQMVESIGLLGPPLSGYL RDTGNYTASFVVGAFLLSGSGILLTLPHFFCSSTTSGPQDLV TEALDTKVPLPKEGLEED				
NOV6b,	CG56918-01	SEQ ID NO: 27	1302 bp	
DNA Sequence		ORF Start: ATG at 9	ORF Stop: TGA at 1287	
CCGCTTAG ATGG CGCGCAGGACAGAGCCCCCGACGGGGGCTGGGGATGGGTGGTGGTGCCTCTCAGCG TTCTTCCAGTCCGCGCTTGTGTTTGGGGTGCCTCGCTCCTTTGGGGTCTTCTTCGTGGAGTTTGTGGC GGCGTTTGGAGGACAGGCAGCGCGCTCTCCTGGATCGCCTCCATAGGAATCGCGGTGCAGCAGTTTG GGAGCCCGGTAGGCAGTGCCCTGAGCACGAAGTTCGGGCCCAGGCCCGTGGTGATGACTGGAGGCATC TTGGCTGCGCTGGGGATGCTGCTCGCCTCTTTTGCTACTTCTTGGACCCACCTATACCTGAGTATTGG GTTGCTGTCAAGCTCTGGCTGGGCTTTGACCTTCGCTCCGACCTGGCCTGCCTGCTCTGTTATTTT CTCGCCGACGATCCCTGGCCACCGGGCTGGCACTGACAGGCGTGGGCCCTCTCCTCCTTCACATTTGCC CCTTTTTCAGTGGCTGCTCAGCCACTACGCCCTGGAGGGGGTCCCTGCTGCTGGTGTCTGCCCTCTC CCTCCACCTAGTGGCCTGTGGTGTCTCCTCCGCCACCCCTCCCTGGCTGAGGACCTGCTGTGGGTG GTCCCAGGGCCCAACTCACCTCTCTCCTCCATCATGGCCCCCTTCTCCGTTACACTGTTGCCCTCACC CTGATCAACACTGGCTACTTCAATCCCTACCTCCACCTGGTGGCCCATCTCCAGGACCTGGATTGGGA CCCACCTACCTGCTGCCTTCTACTCTCAGTTGTTGCTATTTCTGACCTCGTGGGGCGTGTGGTCTCCG GATGGCTGGGAGATGCAGTCCCAGGGCCTGTGACACGACTCCTGATGCTCTGGACCACCTTGACTGGG GTGTCACTAGCCCTGTTCCCTGTAGCTCAGGCTCCACAGCCCTGGTGGCTCTGGCTGTGGCTACGG CTTCACATCAGGGGCTCTGGCCCCACTGGCCTTCTCCGTGCTGCCTGAACATAAGGACAGAGAAGGA TTTACTGTGGCCTGGGACTGTTGCAGATGGTAGAGAGCATCGGGGGGCTGCTGGGGCCTCCTCTCTCA GGCTACCTCCGGGATGTGACAGGCAACTACACGGCTTCTTTTGTGGTGGCTGGGGCCTTCTTCTTTC AGGGAGTGGCATTCTCCTCACCTGCCCCACTTCTTCTGCTCCTCAACTACTACCTCCGGGCCCCAGG ACCTTGTAACAGAAGCACTAGATACTAAAGTTCCCTACCCAAGGAGGGGCTGGAAGAGGACT GAACT CCACAGAGTC				
NOV6b,	CG56918-01	SEQ ID NO: 28	426 aa	MW at 44916.8kD
Protein Sequence				

MARRTEPPDGCWGWVVL ~~SAFFQ~~ SALVFGVLRSGVFFVEFVA ~~A~~ FEEQAARVSWIASIGIAVQQFGSP			
VGSALSTKFGPRPVMTGGILAA ~~L~~ GMLLASFATSLTHLYLSIGLLSGSGWALT ~~FAP~~ TLACLSYFSRR			
RSLATGLALTGVGLSSFT ~~FAP~~ FFQWLLSHYAWRGSLLLVSALSLHLVACGALLRPPSLAEDPAVGGPR			
AQLTSLHHGPFRLRYTVALTLINTGYFI ~~P~~ YLHLVAHLQDLWDPLPA ~~A~~ FLLSVVAISDLVGRVVS ~~G~~ WL			
GDAVPGPVTRLMLWT ~~TL~~ TGVSALALFPVAQAPTALVALAVAYGFTSGALAPLAFSVLPELIGTRRIYC			
GLGLLQMVESIGLLGPPLSGYL ~~R~~ DVTGNYTASFV ~~V~~ VAGAFLLSGSGILLTLPHFFCSSTTTSGPQDLV			
TEALDTKVPLPKEGLEED			
NOV6c,	CG56918-03	SEQ ID NO: 29	1445 bp
DNA Sequence		ORF Start: ATG at 9	ORF Stop: TGA at 1287
CCGCTTAGATGGCGCGCAGGACAGAGCCCCCGACGGGGCTGGGGATGGGTGGTGGTGCCTCTCAGCG			
TTCTTCCAGTCGGCGCTTGTGTTTGGGGTGCTCCGCTCCTTTGGGGTCTTCTTCGTGGAGTTTGTGGC			
GGCGTTTGGAGGACGAGCAGCGCGCTCTCTGGATCGCCTCCATAGGAATCGCGGTGCAGCAGTTTG			
GGAGCCCGGTAGGCAGTGCCTTGAGCACGAAGTTCCGGGCCAGGCCCGTGGTGATGACTGGAGGCATC			
TTGGCTGCGCTGGGGATGCTGCTCGCCTCTTTTGCTACTTCCTTGACCCACCTATACCTGAGTATTGG			
GTTGCTGTGTCAGGCTCTGGCTGGGCTTTGACCTTCGCTCCGAGCCTGGCCTGCCTGTCTGTATTTTCT			
CTCGCCGACGATCCTTGGCCACCGGGCTGGCACTGACAGGCGTGGGCTCTCCTCCTTCACATTTGCC			
CCCTTTTTCCAGTGGCTGCTCAGCCACTACGCTTGAGGGGGTCCCTGCTGCTGGTGTCTGCCCTCTC			
CCTCCACCTAGTGGCCTGTGGTGTCTCTCTCCGCCCACCTCCCTGGCTGAGGACCTGCTGTGGGTG			
GTCCCAGGGCCCACTCACCTCTCTCTCCATCATGGCCCCCTTCCTCCGTTACACTGTTGCCCTCAC			
CTGATCAACACTGGCTACTTCACTTCCCTACCTCCACCTGGTGGCCCATCTCCAGGACCTGGATTGGGA			
CCCACTACCTGCTGCCTTCTACTCTCAGTTGTTGCTATTCTTGACCTCGTGGGGCGTGTGGTCTCCG			
GATGGCTGGGAGATGCAGTCCCAGGGCCTGTGACACGACTCCTGATGCTCTGGACCACCTTGACTGGG			
GTGTCACTAGCCCTGTTCCCTGTAGCTCAGGCTCCCACAGCCCTGGTGGCTCTGGCTGTGGCCTACGG			
CTTCACATCAGGGGCTCTGGCCCCACTGGCCTTCTCCGTGCTGCCTGAACATAATAGGGACTAGAAGGA			
TTTACTGTGGCTGGGACTGTTGCAGATGGTAGAGAGCATCGGGGGCTGCTGGGGCCTCCTCTCTCA			
GGCTACCTCCGGGATGTGACAGGCAACTACACGGCTTCTTTTGTGGTGGCTGGGGCCTTCTTCTTTC			
AGGGAGTGGCATTCTCCTCACCTGCCCCACTTCTTCTGCTCCTCAACTACTACCTCCGGGCCCCAGG			
ACCTTGTAACAGAAGCACTAGATACTAAAGTTCCCTACCCAAGGAGGGACTGGAAGAGGACTTGA			
NOV6c,	CG56918-03	SEQ ID NO: 30	426 aa
Protein Sequence			MW at 44962.9kD
MARRTEPPDGCWGWVVL ~~SAFFQ~~ SALVFGVLRSGVFFVEFVA ~~A~~ FEEQAARVSWIASIGIAVQQFGSP			
VGSALSTKFGPRPVMTGGILAA ~~L~~ GMLLASFATSLTHLYLSIGLLSGSGWALT ~~FAP~~ SLACLSYFSRR			
RSLATGLALTGVGLSSFT ~~FAP~~ FFQWLLSHYAWRGSLLLVSALSLHLVACGALLRPPSLAEDPAVGGPR			
AQLTSLHHGPFRLRYTVALTLINTGYFI ~~P~~ YLHLVAHLQDLWDPLPA ~~A~~ FLLSVVAISDLVGRVVS ~~G~~ WL			
GDAVPGPVTRLMLWT ~~TL~~ TGVSALALFPVAQAPTALVALAVAYGFTSGALAPLAFSVLPELIGTRRIYC			
GLGLLQMVESIGLLGPPLSGYL ~~R~~ DVSGNYTASFV ~~V~~ VAGAFLLSGSGILLTLPHFFCFSTTTSGPQDLV			
TEALDTKVPLPKEGLEED			
NOV6d,	CG56918-04	SEQ ID NO: 31	1289 bp
DNA Sequence		ORF Start: ATG at 9	ORF Stop: TGA at 1287
GATCCACCATGGCGCGCAGGACAGAGCCCCCGACGGGGCTGGGGATGGGTGGTGGTGCCTCTCAGCG			
TTCTTCCAGTCGGCGCTTGTGTTTGGGGTGCTCCGCTCCTTTGGGGTCTTCTTCGTGGAGTTTGTGGC			
GGCGTTTGGAGGACGAGCAGCGCGCTCTCTGGATCGCCTCCATAGGAATCGCGGTGCAGCAGTTTG			
GGAGCCCGGTAGGCAGTGCCTTGAGCACGAAGTTCCGGGCCAGGCCCGTGGTGATGACTGGAGGCATC			
TTGGCTGCGCTGGGGATGCTGCTCGCCTCTTTTGCTACTTCCTTGACCCACCTATACCTGAGTATTGG			
GTTGCTGTGTCAGGCTCTGGCTGGGCTTTGACCTTCGCTCCGACCCTGGCCTGCCTGTCTGTATTTTT			
CTCGCCGACGATCCTTGGCCACCGGGCTGGCACTGACAGGCGTGGGCTCTCCTCCTTCACATTTGCC			
CCCTTTTTCCAGTGGCTGCTCAGCCACTACGCTTGAGGGGGTCCCTGCTGCTGGTGTCTGCCCTCTC			
CCTCCACCTAGTGGCCTGTGGTGTCTCTCTCCGCCCACCTCCCTGGCTGAGGACCTGCTGTGGGTG			
GTCCCAGGGCCCACTCACCTCTCTCTCCATCATGGCCCCCTTCCTCCGTTACACTGTTGCCCTCAC			
CTGATCAACACTGGCTACTTCACTTCCCTACCTCCACCTGGTGGCCCATCTCCAGGACCTGGATTGGGA			
CCCACTACCTGCTGCCTTCTACTCTCAGTTGTTGCTATTCTTGACCTCGTGGGGCGTGTGGTCTCCG			
GATGGCTGGGAGATGCAGTCCCAGGGCCTGTGACACGACTCCTGATGCTCTGGACCACCTTGACTGGG			
GTGTCACTAGCCCTGTTCCCTGTAGCTCAGGCTCCCACAGCCCTGGTGGCTCTGGCTGTGGCCTACGG			
CTTCACATCAGGGGCTCTGGCCCCACTGGCCTTCTCCGTGCTGCCTGAACATAATAGGGACTAGAAGGA			
TTTACTGTGGCTGGGACTGTTGCAGATGGTAGAGAGCATCGGGGGCTGCTGGGGCCTCCTCTCTCA			
GGCTACCTCCGGGATGTGACAGGCAACTACACGGCTTCTTTTGTGGTGGCTGGGGCCTTCTTCTTTC			
AGGGAGTGGCATTCTCCTCACCTGCCCCACTTCTTCTGCTCCTCAACTACTACCTCCGGGCCCCAGG			
ACCTTGTAACAGAAGCACTAGATACTAAAGTTCCCTACCCAAGGAGGGACTGGAAGAGGACTTGA			
NOV6d,	CG56918-04	SEQ ID NO: 32	426 aa
Protein Sequence			MW at 44916.8kD



```

MARRTEPPDGGWGWVVVLSAFFQSALVFGVLRSGVFFVEFVAAFEEQAARVSWIASIGIAVQQFGSP
VGSALSTKFGPRPVMTGGILAALGMLLASFATSLTHLYLSIGLLSGSGWALTFAPTLACLSCYFSRR
RSLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVSALSLHLVACGALLRPPSLAEDPAVGGPR
AQLTSLHHGPFRLRYTVALTLINTGYFIPYLHLVAHLQDLWDPLPAAFLLSVVAISDLVGRVVSGL
GDAVPGPVTRLLMLWTTLTGVSLALFPVAQAPTALVALAVAYGFTSGALAPLAFSVLPeligTRRIYC
GLGLLQMVESIGLLGPPLSGYL RDVTGNYTASFVVAGAFLLSGSGILLTLPHFFCSSTTTSGPQDLV
TEALDTKVPLPKEGLEED
  
```

- 5 A ClustalW comparison of the above protein sequences yields the following sequence alignment shown in Table F4.

Table F4. Comparison of the NOV6 protein sequences.	
NOV6a	MARRTEPPDGGWGWVVVLSAFFQSALVFGVLRSGVFFVEFVAAFEEQAARVSWIASIGI
NOV6b	MARRTEPPDGGWGWVVVLSAFFQSALVFGVLRSGVFFVEFVAAFEEQAARVSWIASIGI
NOV6c	MARRTEPPDGGWGWVVVLSAFFQSALVFGVLRSGVFFVEFVAAFEEQAARVSWIASIGI
NOV6d	MARRTEPPDGGWGWVVVLSAFFQSALVFGVLRSGVFFVEFVAAFEEQAARVSWIASIGI
NOV6a	AVQQFGSPVGSALSTKFGPRPVMTGGILAALGMLLASFATSLTHLYLSIGLLSGSGWAL
NOV6b	AVQQFGSPVGSALSTKFGPRPVMTGGILAALGMLLASFATSLTHLYLSIGLLSGSGWAL
NOV6c	AVQQFGSPVGSALSTKFGPRPVMTGGILAALGMLLASFATSLTHLYLSIGLLSGSGWAL
NOV6d	AVQQFGSPVGSALSTKFGPRPVMTGGILAALGMLLASFATSLTHLYLSIGLLSGSGWAL
NOV6a	TFAPTLACLSCYFSRRRSLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVSALS
NOV6b	TFAPTLACLSCYFSRRRSLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVSALS
NOV6c	TFAPSLACLSCYFSRRRSLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVSALS
NOV6d	TFAPTLACLSCYFSRRRSLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVSALS
NOV6a	HLVACGALLRPPSLAEDPAVGGPRAQLTSLHHGPFRLRYTVALTLINTGYFIPYLHLVAH
NOV6b	HLVACGALLRPPSLAEDPAVGGPRAQLTSLHHGPFRLRYTVALTLINTGYFIPYLHLVAH
NOV6c	HLVACGALLRPPSLAEDPAVGGPRAQLTSLHHGPFRLRYTVALTLINTGYFIPYLHLVAH
NOV6d	HLVACGALLRPPSLAEDPAVGGPRAQLTSLHHGPFRLRYTVALTLINTGYFIPYLHLVAH
NOV6a	LQDLWDPLPAAFLLSVVAISDLVGRVVSGLWDGAVPGPVTRLLMLWTTLTGVSLALFPV
NOV6b	LQDLWDPLPAAFLLSVVAISDLVGRVVSGLWDGAVPGPVTRLLMLWTTLTGVSLALFPV
NOV6c	LQDLWDPLPAAFLLSVVAISDLVGRVVSGLWDGAVPGPVTRLLMLWTTLTGVSLALFPV
NOV6d	LQDLWDPLPAAFLLSVVAISDLVGRVVSGLWDGAVPGPVTRLLMLWTTLTGVSLALFPV
NOV6a	AQAPTALVALAVAYGFTSGALAPLAFSVLPeligTRRIYCGLGLLQMVESIGLLGPPLS
NOV6b	AQAPTALVALAVAYGFTSGALAPLAFSVLPeligTRRIYCGLGLLQMVESIGLLGPPLS
NOV6c	AQAPTALVALAVAYGFTSGALAPLAFSVLPeligTRRIYCGLGLLQMVESIGLLGPPLS
NOV6d	AQAPTALVALAVAYGFTSGALAPLAFSVLPeligTRRIYCGLGLLQMVESIGLLGPPLS
NOV6a	GYLRDVTGNYTASFVVAGAFLLSGSGILLTLPHFFCSSTTTSGPQDLVTEALDTKVPLPK
NOV6b	GYLRDVTGNYTASFVVAGAFLLSGSGILLTLPHFFCSSTTTSGPQDLVTEALDTKVPLPK
NOV6c	GYLRDVTGNYTASFVVAGAFLLSGSGILLTLPHFFCSSTTTSGPQDLVTEALDTKVPLPK
NOV6d	GYLRDVTGNYTASFVVAGAFLLSGSGILLTLPHFFCSSTTTSGPQDLVTEALDTKVPLPK
NOV6a	EGLEED
NOV6b	EGLEED
NOV6c	EGLEED
NOV6d	EGLEED
NOV6a	(SEQ ID NO: 26)
NOV6b	(SEQ ID NO: 28)
NOV6c	(SEQ ID NO: 30)
NOV6d	(SEQ ID NO: 32)

Further analysis of the NOV6a protein yielded the following properties shown in Table F5.

Table F5. Protein Sequence Properties NOV6a

SignalP analysis: Cleavage site between residues 45 and 46

PSORT II analysis:

PSG: a new signal peptide prediction method

N-region: length 9; pos.chg 2; neg.chg 2

H-region: length 22; peak value 10.57

PSG score: 6.17

GvH: von Heijne's method for signal seq. recognition

GvH score (threshold: -2.1): -3.90

possible cleavage site: between 29 and 30

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation

Init position for calculation: 1

Tentative number of TMS(s) for the threshold 0.5: 6

INTEGRAL Likelihood = -6.37 Transmembrane 15 - 31

INTEGRAL Likelihood = -4.09 Transmembrane 82 - 98

INTEGRAL Likelihood = -5.73 Transmembrane 172 - 188

INTEGRAL Likelihood = -3.45 Transmembrane 249 - 265

INTEGRAL Likelihood = -1.75 Transmembrane 296 - 312

INTEGRAL Likelihood = -4.19 Transmembrane 375 - 391

PERIPHERAL Likelihood = 1.01 (at 143)

ALOM score: -6.37 (number of TMSs: 6)

MTOP: Prediction of membrane topology (Hartmann et al.)

Center position for calculation: 22

Charge difference: -2.0 C(-1.0) - N(1.0)

N >= C: N-terminal side will be inside

>>> membrane topology: type 3a

MITDISC: discrimination of mitochondrial targeting seq

R content: 2 Hyd Moment(75): 8.39

Hyd Moment(95): 9.95 G content: 0

D/E content: 2 S/T content: 1

Score: -4.15

Gavel: prediction of cleavage sites for mitochondrial preseq

R-2 motif at 14 RRT|EP

NUCDISC: discrimination of nuclear localization signals

pat4: none

pat7: none

bipartite: none

content of basic residues: 4.7%

NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:

XXRR-like motif in the N-terminus: ARRT

none

SKL: peroxisomal targeting signal in the C-terminus: none

PTS2: 2nd peroxisomal targeting signal: none
VAC: possible vacuolar targeting motif: none
RNA-binding motif: none
Actinin-type actin-binding motif: type 1: none type 2: none
NMYR: N-myristoylation pattern : none
Prenylation motif: none
memYQRL: transport motif from cell surface to Golgi: none
Tyrosines in the tail: none
Dileucine motif in the tail: none
checking 63 PROSITE DNA binding motifs: none
checking 71 PROSITE ribosomal protein motifs: none
checking 33 PROSITE prokaryotic DNA binding motifs: none
NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination Prediction: cytoplasmic Reliability: 94.1
COIL: Lupas's algorithm to detect coiled-coil regions total: 0 residues

Final Results (k = 9/23):
55.6 %: endoplasmic reticulum
22.2 %: mitochondrial
11.1 %: vesicles of secretory system
11.1 %: nuclear
>> prediction for CG56918-02 is end (k=9)

- 5 A search of the NOV6a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table F6.

Table F6. Geneseq Results for NOV6a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE16779	Human transporter and ion channel-16 (TRICH-16) protein - Homo sapiens, 426 aa. [WO200192304-A2,	1..426 1..426	424/426 (99%) 425/426 (99%)	0.0

	06-DEC-2001]			
ABG32785	Human monocarboxylic acid transporter protein TCH081 protein - Homo sapiens, 426 aa. [WO200262998-A1, 15-AUG-2002]	1..426 1..426	424/426 (99%) 425/426 (99%)	0.0
ABP61019	Novel human protein. SEQ ID 110 - Homo sapiens, 426 aa. [WO200250105-A1, 27-JUN-2002]	1..426 1..426	423/426 (99%) 425/426 (99%)	0.0
ABB80596	Human sbg102200MCTb protein - Homo sapiens, 458 aa. [WO200222802-A1, 21-MAR-2002]	1..424 2..425	421/424 (99%) 423/424 (99%)	0.0
ABP69801	Human polypeptide SEQ ID NO 1848 - Homo sapiens, 457 aa. [WO200270539-A2, 12-SEP-2002]	1..424 1..424	421/424 (99%) 423/424 (99%)	0.0

- 5 In a BLAST search of public sequence databases, the NOV6a protein was found to have homology to the proteins shown in the BLASTP data in Table F7.

Table F7. Public BLASTP Results for NOV6a

Protein Accession Number	Protein/Organism/Length	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD80156	Monocarboxylate transporter 13 - Homo sapiens (Human), 426 aa.	1..426 1..426	424/426 (99%) 425/426 (99%)	0.0
Q8CE94	Hypothetical protein (RIKEN cDNA 1700007D07) - Mus musculus (Mouse), 428 aa.	1..426 1..428	371/428 (86%) 396/428 (91%)	0.0
Q8CEH6	Hypothetical protein - Mus musculus (Mouse), 428 aa.	1..426 1..428	367/428 (85%) 393/428 (91%)	0.0
Q8CF46	Hypothetical protein - Mus musculus (Mouse), 335 aa.	94..426 1..335	281/335 (83%) 306/335 (90%)	e-160
Q8NCK7	Hypothetical protein FLJ90193 - Homo sapiens (Human), 471 aa.	7..404 32..439	199/409 (48%) 260/409 (62%)	e-100

PFam analysis predicts that the NOV6a protein contains the domains shown in the Table F8.

Table F8. Domain Analysis of NOV6a

Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Sugar_tr	12..407	73/494 (15%) 251/494 (51%)	0.078

Example F2. Human monocarboxylate transporter Gene Variants and SNPs

Protocol for SNP identification is disclosed in Example Q11.

- 10 The variants of the human monocarboxylate transporter were obtained from direct cloning and/or public databases. In addition to the human version of the monocarboxylate transporter and its variants have been identified by direct sequencing of cDNAs derived from many different human tissues and from sequences in public databases. No splice-form variants have been identified at CuraGen whereas several amino acid-changing and non-amino acid-changing cSNPs were identified
- 15 and are shown in Table F9 below, where UCP=uncharged polar, NP=non polar, A=acidic, and B=basic. Those cSNPs with ID "cgsp" refer to CuraGen proprietary SNPs, whereas those labeled "hsnp" are from public databases. The preferred variant of all those identified, to be used for screening purposes, is CG56918-02.

20 Table F9. CG56918-02 SNPs

SNP ID	SNP MATCH POSI	ALLELES	SNP MATCH POSI ON ORF	ORIG AA	NEW AA	AA CHANGE TYPE	STRAND
cgsp:13376950	399	C/T	133	Phe	Phe	Silent	+
cgsp:13376951	498	C/T	166	Tyr	Tyr	Silent	+
cgsp:13374509	1042	G/A	348	Val	Ile	Conservative	+
cgsp:13374510	1190	T/C	397	Ser	Phe	UCP to NP	+

25 **Example F3. Expression Profile of the Human monocarboxylate transporter Gene (CG56918-02)**

Protocol for quantitative gene expression analysis is disclosed in Example Q9. Protocol for CuraChip™ analysis is disclosed in Example Q12.

- 30 Expression of gene CG56918-01 was assessed using the primer-probe set Ag4532, described in Table F10. Results of the RTQ-PCR runs are shown in Tables F19-F21.

Expression of gene CG108451-01 (SLA16A1) was assessed using the primer-probe set Ag8055, described in Table F11. Results of the RTQ-PCR runs are shown in Tables F19-F21.

Expression of gene SLC16A2 was assessed using the primer-probe set Ag3576, described in Table F12. Results of the RTQ-PCR runs are shown in Tables F19-F21.

- 35 Expression of gene CG108890-01 (SLA16A3) was assessed using the primer-probe set Ag3428, described in Table F13. Results of the RTQ-PCR runs are shown in Tables F19-F21.

Expression of gene SLA16A4 was assessed using the primer-probe set Ag3572, described in Table F14. Results of the RTQ-PCR runs are shown in Tables F19-F21.

- 40 Expression of gene CG117266-01 (SLA16A5) was assessed using the primer-probe set Ag3573, described in Table F15. Results of the RTQ-PCR runs are shown in Tables F19-F21.

Expression of gene CG160563-01 (SLA16A6) was assessed using the primer-probe set Ag3575, described in Table F16. Results of the RTQ-PCR runs are shown in Tables F19-F21.

Expression of gene SLC16A7 was assessed using the primer-probe set Ag3570, described in Table F17. Results of the RTQ-PCR runs are shown in Tables F19-F21.

Expression of gene SLC16A8 was assessed using the primer-probe set Ag3571, described in Table F18. Results of the RTQ-PCR runs are shown in Table F19-F21.

Table F10. Probe Name Ag4532

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-actcacctctctcctccatcat-3'	22	626	103
Probe	TET-5'-cgttacactgttgccctcaccctgat-3'- TAMRA	26	660	104
Reverse	5'-aggggaatgaagtagccagtgtt-3'	22	687	105

10 Table F11. Probe Name Ag8055

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttccggttggttgcaaatgg-3'	18	1017	106
Probe	TET-5'-tgtgtcatatgctagcacctttatcca-3'- TAMRA	27	1037	107
Reverse	5'-aatccaaagaatcccgcata-3'	20	1087	108

Table F12. Probe Name Ag3576

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-acaatctccttgaggatcaaagc-3'	22	2134	109
Probe	TET-5'-tccaaactcattaactaaattctcccca-3'- TAMRA	28	2165	110
Reverse	5'-agaaaggccctggaaaattta-3'	21	2202	111

Table F13. Probe Name Ag3428

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tttcgctcatcactggcttct-3'	20	154	112
Probe	TET-5'-cttccccaaggccgtcagtgtctt-3'- TAMRA	24	181	113
Reverse	5'-aactcctgtatgagctccttga-3'	22	207	114

Table F14. Probe Name Ag3572

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-atggggcaattaaggaagaa-3'	20	95	115
Probe	TET-5'-caccgttttctggaatgttactctttt-3'-	27	58	116

	TAMRA			
Reverse	5'-cacaatgcagcaaaatgaaa-3'	20	29	117

5 Table F15. Probe Name Ag3573

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-acaacctcagccagctctactt-3'	22	322	118
Probe	TET-5'-acagcaggattcatcacaggcctg-3'- TAMRA	24	345	119
Reverse	5'-atgcttgactggaagctgaag-3'	21	377	120

Table F16. Probe Name Ag3575

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tgcttttagttttctcccaactg-3'	22	1209	121
Probe	TET-5'-ccatcctatcacaatatTTTggcaa-3'- TAMRA	26	1180	122
Reverse	5'-aactgcagtgactatggaacgt-3'	22	1156	123

Table F17. Probe Name Ag3570

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ggccttctgtaggattaattgc-3'	22	890	124
Probe	TET-5'-tcgacctcgaattcagtacttcttca-3'- TAMRA	26	927	125
Reverse	5'-gtgacacactccattgaacatg-3'	22	963	126

Table F18. Probe Name Ag3571

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gctgccgagtctgtataactca-3'	22	1607	127
Probe	TET-5'-cccagtgacttgggaacacagcttct-3'- TAMRA	26	1649	128
Reverse	5'-agcgtaccaggctctctga-3'	19	1679	129

Table F19. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag4532, Run 222735297	Rel. Exp.(%) Ag3576, Run 217343281	Rel. Exp.(%) Ag3428, Run 216839464	Rel. Exp.(%) Ag3428, Run 218535201	Rel. Exp.(%) Ag3573, Run 217342760	Rel. Exp.(%) Ag3575, Run 217343280	Rel. Exp.(%) Ag3570, Run 217343116	Rel. Exp.(%) Ag3571, Run 217343117
Adipose	2.0	6.7	0.4	0.5	0.3	0.1	44.4	1.4
Melanoma* Hs688(A).T	3.6	16.6	32.1	41.2	3.7	0.0	44.4	0.7

Melanoma* Hs688(B).T	5.0	17.3	28.7	31.2	4.3	0.0	39.8	0.3
Melanoma* M14	18.2	0.5	6.1	5.9	5.9	17.4	0.7	3.1
Melanoma* LOXIMVI	5.4	0.4	3.6	4.1	0.8	0.0	0.0	0.0
Melanoma* SK-MEL-5	2.7	0.1	0.1	0.1	0.0	100.0	46.7	2.2
Squamous cell carcinoma SCC-4	6.3	15.8	1.1	1.2	2.3	0.1	9.4	0.0
Testis Pool	1.0	1.7	1.5	1.5	0.6	0.8	5.0	1.7
Prostate ca.* (bone met) PC-3	7.3	9.9	14.8	17.7	0.1	0.0	12.2	2.6
Prostate Pool	1.5	4.5	0.3	0.3	1.7	0.0	3.4	8.5
Placenta	9.0	6.3	6.7	8.8	9.2	0.0	0.1	5.8
Uterus Pool	0.8	2.8	0.1	0.1	0.6	0.0	8.4	0.1
Ovarian ca. OVCAR-3	10.7	1.1	3.3	3.5	5.2	0.4	25.5	13.4
Ovarian ca. SK-OV-3	2.5	2.8	29.3	36.6	8.1	0.0	20.9	30.1
Ovarian ca. OVCAR-4	9.0	0.1	1.9	2.4	1.8	0.0	22.5	5.8
Ovarian ca. OVCAR-5	29.5	9.7	38.4	34.9	32.1	0.5	58.2	47.6
Ovarian ca. IGROV-1	4.0	8.0	12.1	12.9	10.9	0.0	0.7	50.0
Ovarian ca. OVCAR-8	14.6	3.9	28.5	30.6	3.1	0.0	3.1	6.8
Ovary	1.7	5.0	0.4	0.6	1.6	0.1	8.2	2.9
Breast ca. MCF-7	17.1	7.5	1.6	2.1	0.3	3.0	17.9	11.0
Breast ca. MDA-MB-231	20.0	1.2	77.9	100.0	5.6	0.0	8.4	9.9
Breast ca. BT 549	32.3	100.0	1.8	2.6	3.2	0.3	0.0	0.9
Breast ca. T47D	68.8	12.3	100.0	91.4	100.0	1.2	76.8	100.0
Breast ca. MDA-N	13.5	0.2	0.0	0.1	6.0	0.0	16.4	4.3
Breast Pool	2.3	9.3	0.3	0.3	1.4	0.1	16.3	0.8
Trachea	7.3	5.1	1.7	2.2	3.8	0.2	13.7	3.3
Lung	0.2	2.2	0.0	0.1	0.2	0.0	3.5	2.7
Fetal Lung	8.4	14.8	1.8	2.3	4.0	0.2	6.9	6.8
Lung ca. NCI- N417	1.0	1.1	0.0	0.0	0.0	0.1	7.2	0.0
Lung ca. LX-1	21.2	0.2	12.0	14.1	20.0	2.7	1.1	19.9
Lung ca. NCI- H146	0.2	0.1	0.2	0.1	0.0	0.1	16.7	5.7
Lung ca. SHP- 77	13.8	0.2	0.1	0.1	0.0	0.0	49.3	24.0
Lung ca. A549	44.4	5.4	13.8	15.9	12.6	0.0	0.2	6.7

Lung ca. NCI-H526	3.1	0.5	0.1	0.0	0.0	0.1	2.2	2.2
Lung ca. NCI-H23	19.5	19.1	1.1	1.4	5.8	0.2	0.4	16.8
Lung ca. NCI-H460	7.5	0.5	5.8	8.0	7.7	6.3	26.4	3.7
Lung ca. HOP-62	17.4	5.6	65.1	71.2	3.7	0.0	7.7	3.5
Lung ca. NCI-H522	26.6	16.6	0.1	0.2	1.0	0.6	0.0	23.0
Liver	23.3	11.1	0.1	0.0	0.0	0.0	1.4	0.8
Fetal Liver	34.2	16.3	0.2	0.3	1.7	0.1	15.9	0.7
Liver ca. HepG2	32.8	0.0	25.3	26.2	7.0	0.5	0.1	3.9
Kidney Pool	4.6	11.1	0.4	0.5	2.5	0.1	21.9	10.0
Fetal Kidney	3.8	6.0	0.2	0.3	2.1	0.1	3.5	5.8
Renal ca. 786-0	9.3	4.4	24.3	28.5	4.8	0.1	49.7	0.0
Renal ca. A498	10.2	3.7	15.4	16.0	0.1	0.0	8.4	0.1
Renal ca. ACHN	100.0	0.5	0.2	0.2	2.2	0.1	19.3	8.0
Renal ca. UO-31	22.2	1.3	21.3	23.0	0.5	0.3	18.3	8.8
Renal ca. TK-10	31.9	1.5	17.6	21.0	4.1	0.4	18.9	23.3
Bladder	3.7	4.2	3.1	3.8	2.0	0.2	50.0	9.4
Gastric ca. (liver met.) NCI-N87	26.6	0.3	15.4	16.2	27.5	1.1	0.2	94.6
Gastric ca. KATO III	33.7	0.1	6.8	8.9	8.4	0.0	0.0	9.9
Colon ca. SW-948	15.0	0.0	8.4	11.3	2.0	0.0	0.0	3.6
Colon ca. SW480	31.2	3.4	19.6	21.3	13.3	1.8	0.5	5.5
Colon ca.* (SW480 met) SW620	18.4	0.3	14.8	17.6	10.5	2.6	0.4	16.6
Colon ca. HT29	4.2	0.0	7.8	10.4	11.0	0.0	4.4	4.0
Colon ca. HCT-116	20.6	0.2	45.7	45.7	9.5	0.7	27.2	13.3
Colon ca. CaCo-2	17.1	0.9	1.6	2.1	7.8	0.1	3.4	7.2
Colon cancer tissue	20.4	2.0	7.0	8.1	3.6	0.2	5.2	5.0
Colon ca. SW1116	6.0	0.0	3.6	4.5	0.0	0.0	0.0	5.3
Colon ca. Colo-205	12.3	0.0	8.3	11.0	1.5	0.0	0.0	3.0
Colon ca. SW-48	13.1	0.0	4.7	6.0	3.2	0.0	0.0	0.6
Colon Pool	3.4	8.1	0.4	0.4	1.3	0.0	19.8	5.8

Small Intestine Pool	2.7	9.6	0.2	0.2	1.4	0.1	6.8	6.7
Stomach Pool	1.1	3.9	0.6	0.6	0.8	0.1	11.0	2.7
Bone Marrow Pool	1.3	4.9	0.2	0.1	0.4	0.1	6.6	0.3
Fetal Heart	1.7	6.1	1.6	2.0	0.8	0.0	14.8	4.5
Heart Pool	2.1	4.0	0.2	0.2	0.6	0.0	44.8	0.8
Lymph Node Pool	2.7	11.1	0.4	0.5	1.4	0.1	22.7	6.2
Fetal Skeletal Muscle	0.7	8.0	0.4	0.5	0.6	0.0	2.6	0.9
Skeletal Muscle Pool	1.2	1.1	3.9	4.4	1.2	0.2	24.7	1.5
Spleen Pool	5.2	2.0	1.0	1.4	0.6	0.3	26.2	0.9
Thymus Pool	4.0	5.8	0.6	0.7	0.9	0.2	14.2	7.8
CNS cancer (glio/astro) U87-MG	47.3	11.2	12.0	11.7	0.0	2.5	75.8	0.0
CNS cancer (glio/astro) U-118-MG	4.0	35.8	30.4	32.8	0.0	1.3	56.3	0.0
CNS cancer (neuro;met) SK-N-AS	1.4	4.8	1.7	1.3	0.0	0.0	15.4	3.6
CNS cancer (astro) SF-539	13.6	8.7	58.6	60.3	0.0	0.0	48.0	0.4
CNS cancer (astro) SNB-75	16.5	56.6	7.5	9.1	0.0	0.1	71.7	2.4
CNS cancer (glio) SNB-19	4.4	9.3	12.7	15.8	10.4	0.0	0.8	39.0
CNS cancer (glio) SF-295	9.3	67.8	1.2	1.1	0.0	0.1	100.0	21.2
Brain (Amygdala) Pool	0.5	3.6	0.4	0.6	0.0	0.1	7.5	17.9
Brain (cerebellum)	4.2	16.6	0.2	0.2	0.1	0.2	5.8	3.7
Brain (fetal)	3.8	27.4	0.3	0.4	0.2	0.2	12.0	7.1
Brain (Hippocampus) Pool	1.1	6.0	0.7	0.8	0.0	0.2	7.1	3.0
Cerebral Cortex Pool	0.5	7.3	0.6	0.7	0.0	0.3	16.2	4.9
Brain (Substantia nigra) Pool	1.1	4.6	1.6	1.5	0.0	0.2	13.0	24.7
Brain (Thalamus) Pool	1.3	8.5	0.6	0.6	0.0	0.4	15.3	10.7
Brain (whole)	1.3	10.9	0.2	0.3	0.2	0.2	11.2	1.7
Spinal Cord Pool	1.3	5.3	1.1	1.1	0.1	0.1	7.4	2.6
Adrenal Gland	2.8	31.2	0.4	0.5	0.3	0.7	15.6	1.7
Pituitary gland	0.6	0.0	0.1	0.1	0.0	0.0	1.8	6.5

Pool								
Salivary Gland	6.3	2.6	0.5	0.6	0.7	0.0	3.1	4.6
Thyroid (female)	6.0	7.7	0.5	0.6	2.8	0.1	1.7	1.6
Pancreatic ca. CAPAN2	32.1	0.4	14.0	18.2	13.2	0.1	46.7	14.4
Pancreas Pool	2.4	8.1	1.4	1.6	1.5	0.2	24.3	6.4

5 Table F20. General_screening_panel_v1.7

Tissue Name	Rel. Exp.(%) Ag8055, Run 318841807	Tissue Name	Rel. Exp.(%) Ag8055, Run 318841807
Adipose	5.1	Gastric ca. (liver met.) NCI-N87	3.8
HUVEC	33.7	Stomach	0.1
Melanoma* Hs688(A).T	0.0	Colon ca. SW-948	12.8
Melanoma* Hs688(B).T	19.9	Colon ca. SW480	5.4
Melanoma (met) SK-MEL-5	16.8	Colon ca. (SW480 met) SW620	40.9
Testis	4.3	Colon ca. HT29	72.7
Prostate ca. (bone met) PC-3	0.3	Colon ca. HCT-116	94.0
Prostate ca. DU145	7.1	Colon cancer tissue	0.3
Prostate pool	0.5	Colon ca. SW1116	4.8
Uterus pool	0.5	Colon ca. Colo-205	0.4
Ovarian ca. OVCAR-3	10.7	Colon ca. SW-48	13.7
Ovarian ca. (ascites) SK-OV-3	1.6	Colon	8.9
Ovarian ca. OVCAR-4	19.2	Small Intestine	0.2
Ovarian ca. OVCAR-5	12.2	Fetal Heart	7.9
Ovarian ca. IGROV-1	97.9	Heart	2.2
Ovarian ca. OVCAR-8	100.0	Lymph Node pool 1	0.2
Ovary	3.1	Lymph Node pool 2	6.2
Breast ca. MCF-7	3.9	Fetal Skeletal Muscle	1.9
Breast ca. MDA-MB-231	0.6	Skeletal Muscle pool	1.5
Breast ca. BT-549	9.3	Skeletal Muscle	15.0
Breast ca. T47D	0.6	Spleen	0.6
Breast pool	0.0	Thymus	0.5

Trachea	1.9	CNS cancer (glio/astro) SF-268	6.9
Lung	2.3	CNS cancer (glio/astro) T98G	4.9
Fetal Lung	4.4	CNS cancer (neuro;met) SK-N-AS	1.9
Lung ca. NCI-N417	8.7	CNS cancer (astro) SF- 539	28.1
Lung ca. LX-1	3.3	CNS cancer (astro) SNB-75	25.2
Lung ca. NCI-H146	6.0	CNS cancer (glio) SNB-19	40.1
Lung ca. SHP-77	20.0	CNS cancer (glio) SF- 295	5.3
Lung ca. NCI-H23	21.0	Brain (Amygdala)	1.6
Lung ca. NCI-H460	27.0	Brain (Cerebellum)	5.3
Lung ca. HOP-62	10.3	Brain (Fetal)	16.0
Lung ca. NCI-H522	62.0	Brain (Hippocampus)	2.2
Lung ca. DMS-114	11.2	Cerebral Cortex pool	1.5
Liver	1.8	Brain (Substantia nigra)	0.6
Fetal Liver	4.3	Brain (Thalamus)	2.7
Kidney pool	3.2	Brain (Whole)	12.1
Fetal Kidney	1.9	Spinal Cord	1.9
Renal ca. 786-0	27.5	Adrenal Gland	2.2
Renal ca. A498	6.3	Pituitary Gland	1.5
Renal ca. ACHN	16.2	Salivary Gland	0.1
Renal ca. UO-31	19.8	Thyroid	1.5
Renal ca. TK-10	14.7	Pancreatic ca. PANC-1	4.4
Bladder	4.5	Pancreas pool	0.8

5 Table F21. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag4532, Run 306350409	Rel. Exp.(%) Ag8055, Run 319958809	Rel. Exp.(%) Ag3428, Run 319961185	Rel. Exp.(%) Ag3572, Run 319961189	Rel. Exp.(%) Ag3573, Run 319960726	Rel. Exp.(%) Ag3575, Run 279370904	Rel. Exp.(%) Ag3570, Run 319961186
97457_Patient- 02go_adipose	1.6	11.1	0.6	6.8	3.7	4.4	15.5
97476_Patient- 07sk_skeletal muscle	0.0	0.0	0.0	0.0	0.0	0.0	0.0
97477_Patient- 07ut_uterus	1.0	41.5	0.4	1.2	7.0	6.6	15.2

97478_Patient-07pl_placenta	3.6	6.3	5.4	49.7	21.9	16.0	3.2
99167_Bayer Patient 1	6.4	1.7	19.6	0.5	11.3	45.1	8.3
97482_Patient-08ut_uterus	0.0	26.4	1.0	0.8	2.2	6.5	8.5
97483_Patient-08pl_placenta	2.9	2.9	3.7	23.5	13.6	6.8	0.6
97486_Patient-09sk_skeletal muscle	0.0	4.6	0.2	0.7	5.7	3.3	3.7
97487_Patient-09ut_uterus	0.0	1.2	0.3	0.3	2.1	5.9	12.8
97488_Patient-09pl_placenta	2.1	1.8	3.1	2.0	22.2	12.9	0.0
97492_Patient-10ut_uterus	0.9	13.5	1.2	0.8	3.1	5.3	5.6
97493_Patient-10pl_placenta	4.8	6.3	14.7	90.8	69.3	27.4	1.6
97495_Patient-11go_adipose	0.5	5.3	0.2	2.6	11.4	5.3	18.7
97496_Patient-11sk_skeletal muscle	1.4	5.5	0.7	0.3	3.9	3.2	23.7
97497_Patient-11ut_uterus	0.6	33.2	1.0	2.4	10.7	12.7	24.0
97498_Patient-11pl_placenta	6.2	1.2	5.8	32.1	28.9	13.2	1.0
97500_Patient-12go_adipose	5.3	3.5	1.4	3.6	6.4	6.3	18.0
97501_Patient-12sk_skeletal muscle	4.7	13.5	17.0	2.4	19.3	6.1	23.0
97502_Patient-12ut_uterus	1.4	48.3	1.1	1.2	3.4	7.5	20.7
97503_Patient-12pl_placenta	13.3	7.0	16.3	100.0	67.8	33.9	1.2
94721_Donor 2 U A_Mesenchymal Stem Cells	9.3	18.7	43.8	15.3	35.1	84.7	26.2
94722_Donor 2 U B_Mesenchymal Stem Cells	5.0	7.1	29.5	6.9	21.6	72.2	13.0
94723_Donor 2 U C_Mesenchymal Stem Cells	7.1	9.9	43.5	10.7	23.3	81.8	31.0
94709_Donor 2 AM	59.9	10.7	36.3	6.7	10.9	49.0	46.7

A_adipose							
94710_Donor 2 AM -	1.9	11.2	31.0	8.5	16.3	38.4	23.8
B_adipose							
94711_Donor 2 AM -	0.5	10.2	20.2	5.6	10.7	25.7	20.0
C_adipose							
94712_Donor 2 AD -	4.4	7.4	30.8	19.2	45.7	58.2	87.7
A_adipose							
94713_Donor 2 AD -	5.8	38.4	31.6	27.5	44.1	86.5	97.9
B_adipose							
94714_Donor 2 AD -	7.0	30.4	27.7	16.2	26.8	54.0	100.0
C_adipose							
94742_Donor 3 U -	5.1	2.5	20.4	8.2	11.0	21.9	11.6
A_Mesenchymal Stem Cells							
94743_Donor 3 U -	2.4	6.5	15.7	7.2	12.9	26.8	11.5
B_Mesenchymal Stem Cells							
94730_Donor 3 AM -	5.6	32.8	60.7	8.7	19.5	44.1	47.6
A_adipose							
94731_Donor 3 AM -	10.6	45.4	74.2	12.5	43.8	67.8	73.2
B_adipose							
94732_Donor 3 AM -	7.2	46.7	61.1	11.0	34.6	61.1	54.3
C_adipose							
94733_Donor 3 AD -	5.6	41.5	100.0	21.9	28.9	100.0	97.9
A_adipose							
94734_Donor 3 AD -	9.1	66.4	92.7	19.8	19.2	92.7	88.9
B_adipose							
94735_Donor 3 AD -	1.2	18.2	36.3	6.5	12.7	34.4	34.2
C_adipose							
77138_Liver_HepG2untreated	100.0	100.0	45.4	2.0	100.0	51.8	3.4
73556_Heart_Cardiac stromal cells (primary)	9.9	4.9	1.9	1.6	1.9	5.9	7.6
81735_Small Intestine	5.4	7.2	0.7	6.4	9.9	9.2	12.7

72409_Kidney Proximal Convolute Tubule	5.4	29.7	17.6	46.3	20.0	12.3	42.9
82685_Small intestine_Duod enum	12.0	5.1	0.7	3.0	22.2	13.6	9.1
90650_Adrenal Adrenocortica l adenoma	0.0	1.1	0.5	1.2	0.7	6.0	20.9
72410_Kidney HRCE	36.9	15.9	46.0	37.9	60.3	17.7	32.3
72411_Kidney HRE	19.6	16.3	12.1	5.2	46.3	9.9	4.7
73139_Uterus_ Uterine smooth muscle cells	6.2	28.1	12.9	13.8	8.1	33.0	13.7

- 5 **General_screening_panel_v1.4 Summary:** Ag4532 The CG56918-02 gene shows widespread expression in this panel with highest expression in a renal cancer cell line (ACHN)(CT=26.4). Among metabolic tissues, this gene has moderate to high levels of expression in adipose, liver, heart, skeletal muscle, adrenal, pituitary, thyroid, and pancreas. Expression of this gene is higher in liver tissues compared to other normal tissues in this panel. In addition, this higher expression in liver
- 10 when compared to other normal tissues is seen only for MCT encoded by CG56918-02, where as other members of the MCT family (SLA16A1 to SLA16A8; please see results from panel 1.7 for SLA16A1) show either ubiquitous expression or very low/undetectable expression in the liver. Therefore, an antagonist for the novel monocarboxylate transporter encoded by this gene may be beneficial in the treatment of diabetes.
- 15 Ag3576 The SLA16A2 gene shows moderate to low expression in normal tissues and number of cancer cell lines with highest expression in breast cancer BT 549 cell line (CT=27.9). Among tissues with metabolic or endocrine function, this gene is expressed at low to moderate levels in pancreas, adipose, adrenal gland, thyroid, skeletal muscle, heart, fetal liver and the gastrointestinal tract.
- 20 Ag3428 Two experiment with same probe-primer sets are in good agreement with highest expression of SLA16A3 gene detected in breast cancer MDA-MB-231 and T47D cell lines (CTs=23-23.3). Among tissues with metabolic or endocrine function, this gene is expressed at low to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, fetal liver and the gastrointestinal tract.
- 25 Ag3573 The SLA16A5 gene shows moderate to high expression mainly in number of cancer cell lines with highest expression in breast cancer T47D cell line (CT=26.4). Among tissues with metabolic or endocrine function, this gene is expressed at low to moderate levels in pancreas, adipose, thyroid, skeletal muscle, heart, fetal liver and the gastrointestinal tract.
- 30 Ag3575 Highest expression of the SLA16A6 gene is detected in melanoma SK-MEL-5 cell line (CT=24.3). Among tissues with metabolic or endocrine function, this gene is expressed at low levels in pancreas, adipose, adrenal gland, thyroid, skeletal muscle, and the gastrointestinal tract.

Ag3570 The SLA16A7 gene shows widespread expression with highest expression in brain cancer SF-295 cell line (CT=27.3). Among the normal tissues, this gene is highly expressed in tissues with metabolic/endocrine function including pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract.

Ag3571 The SLA16A2 gene shows moderate to low expression in number of normal tissues and cancer cell lines with highest expression in breast cancer T47D and gastric cancer NCI-N87 cell lines (CTs=29). Among tissues with metabolic or endocrine function, this gene shows low expression only in pancreas, pituitary gland and gastrointestinal tract.

General_screening_panel_v1.7 Summary: Ag8055 The SLA16A1 gene shows widespread expression with highest expression in ovarian cancer OVCAR-8 and IGROV-1 cell lines (CTs=25.9). This gene shows moderate to low expression in tissues with metabolic or endocrine function including pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract.

Panel 5 Islet Summary: Ag4532 Highest expression of the CG56918-02 gene is detected in a liver cancer cell line (CT=30.9). Moderate to low expression of this gene is also seen in midway differentiated and differentiated adipose tissue. In addition, low expression of this gene is seen in placenta of a diabetic patient on insulin. Please see panel 1.4 for further discussion on the utility of this gene.

Ag8055 Highest expression of the SLA16A1 gene is detected in liver cancer HepG2 cell line (CT=31.8). Moderate to low expression of this gene is also seen in uterus, skeletal muscle, adipose and kidney. Interestingly, expression of this gene is upregulated in midway differentiated and differentiated adipose tissues.

Ag3428 Highest expression of the SLA16A3 gene is seen in differentiated adipose tissue (CT=26.3). Moderate to high expression of this gene is seen in islet cells, placenta, skeletal muscle, kidney, undifferentiated, midway and fully differentiated adipose.

Panel 5 Islet Summary: Ag3572 The SLA16A4 gene shows a widespread expression with highest expression in placenta (CTs=29).

Ag3573 The SLA16A5 gene shows ubiquitous expression with highest expression in liver cancer HepG2 cell line (CT=29.7).

Ag3575 The SLA16A6 gene shows ubiquitous expression with highest expression in differentiated adipose tissue (CT=25.2).

Ag3570 The SLA16A7 gene shows widespread expression with highest expression in differentiated adipose tissue (CT=30.3).

Example F4. Biochemistry/Cell Line Expression/ Screening Assay Formulation

Assays for screening for antibody therapeutics or small molecule drugs targeting human monocarboxylate transporter can be formulated utilizing the non-exhaustive list of cell lines that express the monocarboxylate transporter from the RTQ-PCR results shown above. To assay the

5 enzymatic activity of monocarboxylate transporter the uptake of L-[(14)C]lactate and the measurement[s] of proton influx using 2',7'-bis(carboxy-ethyl)-5(6)-carboxyfluorescein (BCECF) (1996, J Biol Chem. 271: 861) can be utilized. To assess the selectivity of the compounds, endogenous substrates like lactate, pyruvate, ketone bodies and known inhibitors like alpha-cyano-4-hydroxycinnamate (CHC), 4,4'-diisothiocyanostilbene-2,2'-disulfonate (DIDS) may be used.

10

Enzymatic Reaction:



15 Protocols

Following Examples describe procedures, protocols and technologies described in this application.

Example Q1. Mouse Diet-Induced Obesity (DIO) Study (BP24.02)

Overview

20 The predominant cause for obesity in clinical populations is excess caloric intake. This so-called diet-induced obesity (DIO) is mimicked in animal models by feeding high fat diets of greater than 40% fat content. The DIO study was established to identify the gene expression changes contributing to the development and progression of diet-induced obesity. In addition, the study design sought to identify the factors that led to the ability of certain individuals to resist the effects of a high fat diet and thereby prevent obesity.

25

The sample groups for the study normally had body weights +1 S.D., + 4 S.D. and + 7 S.D. of the chow-fed controls. In addition, the biochemical profile of the + 7 S.D. mice normally revealed a further stratification of these animals into mice that retained a normal glycemic profile in spite of obesity and mice that demonstrated hyperglycemia. Tissues examined included hypothalamus, brainstem, liver, retroperitoneal white adipose tissue (WAT), epididymal WAT, brown adipose tissue (BAT), gastrocnemius muscle (fast twitch skeletal muscle) and soleus muscle (slow twitch skeletal muscle). The differential gene expression profiles for these tissues revealed genes and pathways that can be used as therapeutic targets for obesity and/or diabetes.

30

Protocol

35 5 groups of mice were used with 3 mice per group. Occasionally, more than 3 mice were used in a single group in order to preserve correct parameters for the study. In such case only 3 mice would be sacrificed for tissues. The groups were grouped based on the following parameters:

Group 1. Chow fed mice

40

Group 2. Mice fed a high fat diet who were 1 standard deviation in weight above the chow fed mice.

Group 3. Mice fed a high fat diet who were 4 standard deviations in weight above the chow fed mice.

Group 4. Mice fed a high fat diet who were 7 standard deviations in weight above the chow fed mice and who had normal glucose levels.

Group 5: Mice fed a high fat diet who were 7 standard deviation in weight above the chow fed mice and who were hyperglycemic.

In each group of mice, there were 3 mice that were sacrificed and tissues harvested for the study.

Example Q2. Rat Pancreatic Islet Study (BP24.03)

10 Overview

An important clinical goal in the early phases of Type II diabetes is to increase insulin secretion from the beta cells of the pancreas. Numerous agents have been identified that can modulate insulin secretion experimentally and in therapeutic situations. When applied to isolated rat pancreatic islets, the changes in gene expression can be correlated with insulin secretion. In this study, acute and chronic changes in gene expression were examined from islets treated with an agent after short (4 hour) and long-term (5 days) exposure, respectively, compared with the basal state (11 mM glucose). The agents included elevated (25 mM) glucose, glucose (11 mM) and exendin-4 (1 nM), glucose (11 mM) and glybenclamide (50 μ M) and glucose (11 mM) and oleate (2 mM).

20 Protocol

All samples were isolated rat islets. They differed only in the treatment that they received. The following samples were in the 4 hour group:

- 1) 11 mM glucose
- 2) 25 mM glucose
- 25 3) 11 mM glucose & JTT 608
- 4) 11 mM glucose & Carbacol
- 5) 11 mM glucose & Exendin-4

Isolated rat islets were treated with either 11mM glucose (basal state) or 25mM glucose (elevated glucose.). Then there were 3 additional sets of rat islets that were treated with 11mM glucose and one of the 3 agents: JTT 608, Carbacol, or Exendin-4.

The following samples were in the 5 day group:

- 1) 11 mM glucose
- 2) 25 mM glucose
- 3) 11 mM glucose & 1nM Exendin
- 35 4) 11 mM glucose & 50uM Glybenclamide
- 5) 11 mM glucose & 2 mM Oleic Acid

Isolated rat islets were treated with either 11mM glucose (basal state) or 25mM glucose (elevated glucose.). Then there were 3 additional sets of rat islets that were treated with 11mM glucose and one of the 3 agents: exendin, glybenclamide, or oleic acid.

40 From all 10 samples, each was split into 2 replicates. The 2 replicates were run for differential gene expression analysis (GeneCalling®).

Example Q3. Rat Insulin Sensitivity Study (BP24.05)

Protocol

45 ZDF rats or their lean littermates were treated with a variety of agents that are known to alter insulin sensitivity. Metformin, vanadate, and AICAR enhance tissue response to insulin, while the free

5 fatty acids generated by Liposyn (intravenous lipid infusion) treatment reduces the response. A variety of tissues were harvested, including gastrocnemius and soleus muscles, liver, retroperitoneal and epididymal WAT, and IBAT.

Only gastrocnemius and soleus muscles were processed for differential gene expression analysis (GeneCalling®).

10 There were 5 groups of samples:

- 1) Metformin vehicle (vehicle M)
- 2) Metformin treated rats
- 3) AICAR and vanadate vehicle (vehicle AV)
- 4) AICAR treated rats
- 15 5) Vanadate treated rats

Treatment was for 4 hours and glucose values before and after treatment were obtained. Each sample was done in triplicate (5 groups X 3 rats X 2 tissues).

In the second part of the study rats were given an intravenous lipid infusion which should reduce tissue response to insulin in treated rats.

20 In the intravenous lipid infusion part of the study 2 groups were used:

- 1) Rats treated with lipid infusion vehicle.
- 2) Rats treated with lipid infusion.

In each group, there were 3 rats (done in triplicate) and soleus and gastrocnemius samples were processed for differential gene expression analysis (GeneCalling®) (2 groups X 3 samples X 2
25 tissues).

Example Q4. Mouse TZD Response Study (BP24.07)

Overview and Protocol

The peroxisome proliferator-activated receptor gamma (PPAR γ) is the member of the nuclear hormone receptor subfamily of transcription factors that plays a major role in regulation of
30 metabolism. The thiazolidinedione (TZD) drugs, including rosiglitazone, are synthetic agonists of PPAR γ receptors that can normalize elevated plasma glucose levels in obese, diabetic rodents and are often quite efficacious therapeutic agents for the treatment of noninsulin-dependent diabetes mellitus in humans. Diabetic animals demonstrate differential responses to TZD treatment. To understand the basis for this differential response we compared changes in gene expression
35 between diabetic animals that responded favorably and that did not respond to TZD treatment. Female db/db mice were treated daily with 10mg per kilogram body weight rosiglitazone for 7 days. On day 8, the mice were bled for blood glucose. Treated mice were grouped into either a responder group that demonstrated a significant decrease of their hyperglycemia and a non-responder group that demonstrated no change in their blood glucose level. Gene expression in skeletal muscle and
40 adipose tissues was compared between untreated diabetic mice and the two sub-groups of TZD treated mice.

3 tissues were collected for differential gene expression analysis (GeneCalling®): liver, thigh muscle, and uterine white adipose.

3 groups of samples were used:

- 45 1) Vehicle treated

2) Rosiglitazone responders

3) Rosiglitazone non-responders

Each group had 3 mice in it. Total of 27 samples were processed for differential gene expression analysis (GeneCalling®).

Example Q5. Insulin Resistance Study (MB.01)

The spontaneously hypertensive rat (SHR) is a strain exhibiting features of the human Metabolic Syndrome X. The phenotypic features include obesity, hyperglycemia, hypertension, dyslipidemia and dysfibrinolysis. Tissues were removed from adult male rats and a control strain (Wistar _ Kyoto) to identify the gene expression differences that underlie the pathologic state in the SHR and in animals treated with various anti-hyperglycemic agents such as troglitazone. Tissues included sub-cutaneous adipose, visceral adipose, brain, muscle, and liver. Each tissue was collected in triplicate for differential gene expression analysis (GeneCalling®).

Example Q6. Genetically Obese Mice vs. Genetically Lean Mice Study (MB.04)

Overview

A number of genetic models of obesity have been studied, most prominently in mouse and rat, but only a few causative genes have been identified. In this study, a set of mouse genetic models were studied in order to identify by positional expression cloning the genes for obesity in mice, which genes may be relevant to human obesity.

Protocol

7 strains of mice were used. Some exhibited a lean phenotype, some were normal weight mice, and some of the mice were genetically obese.

The following strains of mice used in the study:

- 1) CAST/EI - lean phenotype
- 2) SM/J black - lean phenotype
- 3) SWR/J - normal phenotype
- 4) C57L/J - normal phenotype
- 5) C57BL/6J - normal phenotype
- 6) AKR/J - obese phenotype
- 7) NZB/BINJ - obese phenotype

Various tissues were processed for differential gene expression analysis (GeneCalling®): brain, liver, muscle, and adipose.

Samples were processed in triplicate (i.e. 3 brain samples from 3 CAST/EI mice). Total of 7 strains X 4 tissues X 3 samples = 84 samples were used.

Example Q7. Method of Identifying the Differentially Expressed Gene and Gene Product (GeneCalling®)

The GeneCalling® technology 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). GeneCalling® technology is also disclosed in U.S. Pat. No. 5,871,697. cDNA was derived from various samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples

5 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 three methods routinely used to confirm the identity of the gene fragment found to have altered expression in models of or patients with obesity and/or diabetes are described below.

A). Direct Sequencing

20 The differentially expressed gene fragment is isolated, cloned into a plasmid, and sequenced. Afterwards, the sequence information is used to design an oligonucleotide corresponding to either or both termini of the gene fragment. This oligonucleotide, when used in a competitive PCR reaction, will ablate the electropherographic band from which the sequence is derived.

B). Competitive PCR

25 In competitive PCR, the electropherographic peaks corresponding to the gene fragment of the gene of interest are ablated when a gene-specific primer (designed from the sequenced band or available databases) competes with primers in the linker-adaptors during the PCR amplification.

C). PCR with Perfect or Mismatched 3' Nucleotides (TraPping)

30 This method utilizes a competitive PCR approach using a degenerate set of primers that extend one or two nucleotides into the gene-specific region of the fragment beyond the flanking restriction sites. As in the competitive PCR approach, primers that lead to the ablation of the electropherographic band add additional sequence information. In conjunction with the size of the gene fragment and the 12 nucleotides of sequence derived from the restriction sites, this additional sequence data can uniquely define the gene after database analysis. TraPping is disclosed in a published PCT application Pub. No. WO 01/49886.

Example Q8. Identification of Human Sequences

The laboratory cloning was performed using one or more of the methods summarized below:
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 Corporation's SeqCalling technology that is disclosed in full in U. S. Ser. Nos. 09/417,386 filed Oct. 13, 1999 (also in PCT application Pub. No.: WO 00/40757), and 09/614,505 filed July 11, 2000, the disclosures of which are incorporated herein.

5 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 bioinformatics 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
10 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.

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in
15 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
20 acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be 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
25 regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

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
30 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.

Exon Linking: The cDNA coding for the CG101190-01 sequence was cloned by the polymerase chain reaction (PCR) using the primers designed based on known cDNA sequences or in silico
35 predictions of the full length or some portion (one or more exons) of the cDNA/protein sequence of the invention. These primers were used to amplify a cDNA from a pool containing expressed human sequences derived from the following tissues: 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
40 gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus.

Physical Clone: 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.

5 **Example Q9. Quantitative expression analysis (RTQ-PCR) of clones in various cells and tissues**

The quantitative expression of various NOV genes 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) performed on an Applied Biosystems (Foster City, CA) 10 ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System.

RNA integrity of all samples was determined 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 (degradation products). Control samples to detect genomic DNA contamination included RTQ-PCR reactions run in the absence of reverse 15 transcriptase using probe and primer sets designed to amplify across the span of a single exon. RNA samples were normalized in reference to nucleic acids encoding constitutively expressed genes (i.e., β -actin and GAPDH). Alternatively, 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 20 total RNA in a volume of 20 μ l or were scaled up to contain 50 μ g of total RNA in a volume of 100 μ l and were incubated for 60 minutes at 42°C. sscDNA samples were then normalized in reference to nucleic acids as described above.

Probes and primers were designed according to Applied Biosystems *Primer Express* Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using 25 the target sequence as input. Default reaction condition settings and the following parameters were set before selecting primers: 250 nM primer concentration; 58°-60° C primer melting temperature (T_m) range; 59° C primer optimal T_m; 2° C maximum primer difference (if probe does not have 5' G, probe T_m must be 10° C greater than primer T_m; and 75 bp to 100 bp amplicon size. The selected probes and primers were synthesized by Synthesgen (Houston, TX). Probes were double purified by 30 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: 900 nM forward and reverse primers, and 200nM probe.

Normalized RNA was spotted in individual wells of a 96 or 384-well PCR plate (Applied Biosystems, Foster City, CA). PCR cocktails included a single gene-specific probe and primers set or 35 two multiplexed probe and primers sets. PCR reactions were done using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles: 95° C 10 min, then 40 cycles at 95° C for 15 seconds, followed by 60° C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) and 40 plotted using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression was the reciprocal of the RNA difference multiplied by 100. CT values below 28 indicate high expression, between 28 and 32 indicate moderate expression, between 32 and 35 indicate low expression and above 35 reflect levels of expression that were too low to be measured 45 reliably.

Normalized ssDNA was analyzed by RTQ-PCR using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification and analysis were done as described above.

Panels 1, 1.1, 1.2, and 1.3D

Panels 1, 1.1, 1.2 and 1.3D included 2 control wells (genomic DNA control and chemistry control) and 94 wells of cDNA samples from cultured cell lines and primary normal tissues. Cell lines were derived from carcinomas (ca) including: lung, small cell (s cell var), non small cell (non-s or non-sm); breast; melanoma; colon; prostate; glioma (glio), astrocytoma (astro) and neuroblastoma (neuro); squamous cell (squamous); ovarian; liver; renal; gastric and pancreatic from the American Type Culture Collection. Normal tissues were obtained from individual adults or fetuses and included: adult and fetal skeletal muscle, adult and fetal heart, adult and fetal kidney, adult and fetal liver, adult and fetal lung, brain, 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. The following abbreviations are used in reporting the results: metastasis (met); pleural effusion (pl. eff or pl effusion) and * indicates established from metastasis.

General_screening_panel_v1.4, v1.5, v1.6 and v1.7

Panels 1.4, 1.5, 1.6 and 1.7 were as described for Panels 1, 1.1, 1.2 and 1.3D, above except that normal tissue samples were pooled from 2 to 5 different adults or fetuses.

Panels 2D, 2.2, 2.3 and 2.4

Panels 2D, 2.2, 2.3 and 2.4 included 2 control wells and 94 wells containing RNA or cDNA from human surgical specimens procured through the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative, Ardaïs or Clinomics BioSciences (Frederick, MD). Tissues included human malignancies and in some cases matched adjacent normal tissue (NAT). Information regarding histopathological assessment of tumor differentiation grade as well as the clinical stage of the patient from which samples were obtained was generally available. Normal tissue RNA and cDNA samples were purchased from various commercial sources such as Clontech, Research Genetics and Invitrogen.

HASS Panel v 1.0

The HASS Panel v1.0 included 93 cDNA samples and two controls including: 81 samples of cultured human cancer cell lines subjected to serum starvation, acidosis and anoxia according to established procedures for various lengths of time; 3 human primary cells; 9 malignant brain cancers (4 medulloblastomas and 5 glioblastomas); and 2 controls. Cancer cell lines (ATCC) were cultured using recommended conditions and included: breast, prostate, bladder, pancreatic and CNS. Primary human cells were obtained from Clonetics (Walkersville, MD). Malignant brain samples were gifts from the Henry Ford Cancer Center.

ARDAIS Panel v1.0 and v1.1

The ARDAIS Panel v1.0 and v1.1 included 2 controls and 22 test samples including: human lung adenocarcinomas, lung squamous cell carcinomas, and in some cases matched adjacent normal tissues (NAT) obtained from Ardaïs. Unmatched malignant and non-malignant RNA samples from lungs with gross histopathological assessment of tumor differentiation grade and stage and clinical state of the patient were obtained from Ardaïs.

5 **ARDAIS Prostate v1.0**

ARDAIS Prostate v1.0 panel included 2 controls and 68 test samples of human prostate malignancies and in some cases matched adjacent normal tissues (NAT) obtained from Ardaís. RNA from unmatched malignant and non-malignant prostate samples with gross histopathological assessment of tumor differentiation grade and stage and clinical state of the patient were also
10 obtained from Ardaís.

ARDAIS Kidney v1.0

ARDAIS Kidney v1.0 panel included 2 control wells and 44 test samples of human renal cell carcinoma and in some cases matched adjacent normal tissue obtained from Ardaís. RNA from unmatched renal cell carcinoma and normal tissue with gross histopathological assessment of tumor
15 differentiation grade and stage and clinical state of the patient were also obtained from Ardaís.

ARDAIS Breast v1.0

ARDAIS Breast v1.0 panel included 2 control wells and 71 test samples of human breast malignancies and in some cases matched adjacent normal tissue obtained from Ardaís. RNA from unmatched malignant and non-malignant breast samples with gross histopathological assessment of
20 tumor differentiation grade and stage and clinical state of the patient were also obtained from Ardaís.

Panel 3D, 3.1 and 3.2

Panels 3D, 3.1, and 3.2 included two controls, 92 cDNA samples of cultured human cancer cell lines and 2 samples of human primary cerebellum. Cell lines (ATCC, National Cancer Institute, German tumor cell bank) were cultured as recommended and were derived from: squamous cell
25 carcinoma of the tongue, melanoma, sarcoma, leukemia, lymphoma, and epidermoid, bladder, pancreas, kidney, breast, prostate, ovary, uterus, cervix, stomach, colon, lung and CNS carcinomas.

Panels 4D, 4R, and 4.1D

Panels 4D, 4R, and 4.1D included 2 control wells and 94 test samples of RNA (Panel 4R) or cDNA (Panels 4D and 4.1D) from human cell lines or tissues related to inflammatory conditions.
30 Controls included total RNA from normal tissues such as colon, lung (Stratagene, La Jolla, CA), thymus and kidney (Clontech). Total RNA from cirrhotic and lupus kidney was obtained from BioChain Institute, Inc., (Hayward, CA). Crohn's intestinal and ulcerative colitis samples were obtained from the National Disease Research Interchange (NDRI, Philadelphia, PA). Cells purchased from Clonetics included: astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth
35 muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, and human umbilical vein endothelial. These primary cell types were activated by incubating with various cytokines (IL-1 beta ~1-5 ng/ml, TNF alpha ~5-10 ng/ml, IFN gamma ~20-50 ng/ml, IL-4 ~5-10 ng/ml, IL-9 ~5-10 ng/ml, IL-13 5-10 ng/ml) or combinations of cytokines as indicated. Starved endothelial cells were
40 cultured in the basal media (Clonetics, Walkersville, MD) with 0.1% serum.

Mononuclear cells were prepared from blood donations using Ficoll. LAK cells were cultured in culture media [DMEM, 5% FCS (Hyclone, Logan, UT), 100 mM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco)] and interleukin 2 for 4-6 days. Cells were activated with 10-20
45 ng/ml PMA and 1-2 µg/ml ionomycin, 5-10 ng/ml IL-12, 20-50 ng/ml IFN gamma or 5-10 ng/ml IL-18

5 for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in culture media with ~5 mg/ml PHA (phytohemagglutinin) or PWM (pokeweed mitogen; Sigma-Aldrich Corp., St. Louis, MO). Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing them 1:1 at a final concentration of $\sim 2 \times 10^6$ cells/ml in culture media. The MLR samples
10 were taken at various time points 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 (Miltenyi Biotec, Auburn, CA) according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culturing in culture media with 50 ng/ml GM-CSF and 5 ng/ml IL-4 for 5-7 days. Macrophages were prepared by culturing monocytes for
15 5-7 days in culture media with ~50 ng/ml 10% type AB Human Serum (Life technologies, Rockville, MD) or MCSF (Macrophage colony stimulating factor; R&D, Minneapolis, MN). Monocytes, macrophages and dendritic cells were stimulated for 6 or 12-14 hours with 100 ng/ml lipopolysaccharide (LPS). Dendritic cells were also stimulated with 10 μ g/ml anti-CD40 monoclonal antibody (Pharmingen, San Diego, CA) for 6 or 12-14 hours.

20 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. CD45+RA and CD45+RO 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 Miltenyi beads were then used to
25 separate the CD45+RO CD4+ lymphocytes from CD45+RA CD4+ lymphocytes. CD45+RA CD4+, CD45+RO CD4+ and CD8+ lymphocytes were cultured in culture media at 10^6 cells/ml in culture plates precoated overnight with 0.5 mg/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, isolated CD8+ lymphocytes were activated for 4 days on
30 anti-CD28, anti-CD3 coated plates and then harvested and expanded in culture media with IL-2 (1 ng/ml). These CD8+ cells were activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as described above. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. Isolated NK cells were cultured in culture media with 1 ng/ml IL-2 for 4-6 days before RNA was prepared.

35 B cells were prepared from minced and sieved tonsil tissue (NDRI). Tonsil cells were pelleted and resuspended at 10^6 cells/ml in culture media. Cells were activated using 5 μ g/ml PWM (Sigma-Aldrich Corp) or ~10 μ g/ml anti-CD40 (Pharmingen) and 5-10 ng/ml IL-4. Cells were harvested for RNA preparation after 24, 48 and 72 hours.

To prepare primary and secondary Th1/Th2 and Tr1 cells, umbilical cord blood CD4+
40 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10^5 - 10^6 cells/ml in culture media with IL-2 (4 ng/ml) in 6-well Falcon plates (precoated overnight with 10 μ g/ml anti-CD28 (Pharmingen) and 2 μ g/ml anti-CD3 (OKT3; ATCC) then washed twice with PBS).

To stimulate Th1 phenotype differentiation, IL-12 (5 ng/ml) and anti-IL4 (1 μ g/ml) were used; for Th2 phenotype differentiation, IL-4 (5 ng/ml) and anti-IFN gamma (1 μ g/ml) were used; and for
45 Tr1 phenotype differentiation, IL-10 (5 ng/ml) was used. After 4-5 days, the activated Th1, Th2 and

5¹¹ Tr1 lymphocytes were washed once with DMEM and expanded for 4-7 days in culture media with IL-2 (1 ng/ml). Activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/CD3 and cytokines as described above with the addition of anti-CD95L (1 µg/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and expanded in culture media with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained 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.

Leukocyte cells lines Ramos, EOL-1, KU-812 were obtained from the ATCC. EOL-1 cells were further differentiated by culturing in culture media at 5×10^5 cells/ml with 0.1 mM dbcAMP for 8 days, changing the media every 3 days and adjusting the cell concentration to 5×10^5 cells/ml. RNA was prepared from resting cells or cells activated with PMA (10 ng/ml) and ionomycin (1 µg/ml) for 6 and 14 hours. RNA was prepared from resting CCD 1106 keratinocyte cell line (ATCC) or from cells activated with ~5 ng/ml TNF alpha and 1 ng/ml IL-1 beta. RNA was prepared from resting NCI-H292, airway epithelial tumor cell line (ATCC) or from cells activated for 6 and 14 hours in culture media with 5 ng/ml IL-4, 5 ng/ml IL-9, 5 ng/ml IL-13, and 25 ng/ml IFN gamma.

RNA was prepared by lysing approximately 10^7 cells/ml using Trizol (Gibco BRL) then adding 1/10 volume of bromochloropropane (Molecular Research Corporation, Cincinnati, OH), vortexing, incubating for 10 minutes at room temperature and then spinning at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was placed in a 15 ml Falcon Tube and 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 and washed in 70% ethanol. The pellet was redissolved in 300 µl of RNase-free water with 35 ml buffer (Promega, Madison, WI) 5 µl DTT, 7 µl RNasin and 8 µl DNase and incubated at 37° C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3 M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down, placed in RNase free water and stored at -80° C.

AI_comprehensive panel_v1.0

Autoimmunity (AI) comprehensive panel v1.0 included two controls and 89 cDNA test samples isolated from male (M) and female (F) surgical and postmortem human tissues that were obtained from the Backus Hospital and Clinomics. Tissue samples included : normal, adjacent (Adj); matched normal adjacent (match control); joint tissues (synovial (Syn) fluid, synovium, bone and cartilage, osteoarthritis (OA), rheumatoid arthritis (RA)); psoriatic; ulcerative colitis colon; Crohns disease colon; and emphysematic, asthmatic, allergic and chronic obstructive pulmonary disease (COPD) lung.

Pulmonary and General inflammation (PGI) panel v1.0

Pulmonary and General inflammation (PGI) panel v1.0 included two controls and 39 test samples isolated as surgical or postmortem samples. Tissue samples include: five normal lung samples obtained from Maryland Brain and Tissue Bank, University of Maryland (Baltimore, MD), International Bioresource systems, IBS (Tuscon, AZ), and Asterand (Detroit, MI), five normal adjacent intestine tissues from Ardais, ulcerative colitis samples (UC) from Ardais; Crohns disease colon from NDRI, National Disease Research Interchange (Philadelphia, PA); emphysematous tissue samples

5 from Ardais and Genomic Collaborative Inc. (Cambridge, MA), asthmatic tissue from Maryland Brain and Tissue Bank, University of Maryland (Baltimore, MD) and Genomic Collaborative Inc and fibrotic tissue from Ardais and Genomic Collaborative.

Cellular OA/RA Panel

Cellular OA/RA panel includes 2 control wells and 35 test samples comprised of cDNA
 10 generated from total RNA isolated from human cell lines or primary cells representative of the human joint and its inflammatory condition. Cell types included normal human osteoblasts (Nhost) from Clonetics (Cambrex, East Rutherford, NJ), human chondrosarcoma SW1353 cells from ATCC (Manassas, VA), human fibroblast-like synoviocytes from Cell Applications, Inc. (San Diego, CA) and MH7A cell line (a rheumatoid fibroblast-like synoviocytes transformed with SV40 T antigen) from
 15 Riken Cell bank (Tsukuba Science City, Japan). These cell types were activated by incubating with various cytokines (IL-1 beta ~1-10 ng/ml, TNF alpha ~5-50 ng/ml, or prostaglandin E2 for Nhost cells) for 1, 6, 18 or 24 h. All these cells were starved for at least 5 h and cultured in their corresponding basal medium with ~ 0.1 to 1 % FBS.

Minitissue OA/RA Panel

20 The OA/RA mini panel includes two control wells and 31 test samples comprised of cDNA generated from total RNA isolated from surgical and postmortem human tissues obtained from the University of Calgary (Alberta, Canada), NDRI (Philadelphia, PA), and Ardais Corporation. Joint tissue samples include synovium, bone and cartilage from osteoarthritic and rheumatoid arthritis patients undergoing reconstructive knee surgery, as well as, normal synovium samples (RNA and
 25 tissue). Visceral normal tissues were pooled from 2-5 different adults and included adrenal gland, heart, kidney, brain, colon, lung, stomach, small intestine, skeletal muscle, and ovary.

AI.05 chondrosarcoma

AI.05 chondrosarcoma plates included SW1353 cells (ATCC) subjected to serum starvation and treated for 6 and 18 h with cytokines that are known to induce MMP (1, 3 and 13) synthesis (e.g.
 30 IL1beta). These treatments included: IL-1beta (10 ng/ml), IL-1beta + TNF-alpha (50 ng/ml), IL-1beta + Oncostatin (50 ng/ml) and PMA (100 ng/ml). Supernatants were collected and analyzed for MMP 1, 3 and 13 production. RNA was prepared from these samples using standard procedures.

Panels 5D and 5I

Panel 5D and 5I included two controls and cDNAs isolated from human tissues, human
 35 pancreatic islets cells, cell lines, metabolic tissues obtained from patients enrolled in the Gestational Diabetes study (described below), and cells from different stages of adipocyte differentiation, including differentiated (AD), midway differentiated (AM), and undifferentiated (U; human mesenchymal stem cells).

Gestational Diabetes study subjects were young (18 - 40 years), otherwise healthy women
 40 with and without gestational diabetes undergoing routine (elective) Caesarean section. Uterine wall smooth muscle (UT), visceral (Vis) adipose, skeletal muscle (SK), placenta (PI) greater omentum adipose (GO Adipose) and subcutaneous (SubQ) adipose samples (less than 1 cc) were collected, rinsed in sterile saline, blotted and flash frozen in liquid nitrogen. Patients included: Patient 2, an overweight diabetic Hispanic not on insulin; Patient 7-9, obese non-diabetic Caucasians with body

5 mass index (BMI) greater than 30; Patient 10, an overweight diabetic Hispanic, on insulin; Patient 11, an overweight nondiabetic African American; and Patient 12, a diabetic Hispanic on insulin.

Differentiated adipocytes were obtained from induced donor progenitor cells (Clonetics).

Differentiated human mesenchymal stem cells (HuMSCs) were prepared as described in *Science* Apr 2 1999:143-147. mRNA was isolated and sscDNA was produced from Trizol lysates or frozen
10 pellets. Human cell lines (ATCC, NCI or German tumor cell bank) included: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells and adrenal cortical adenoma cells. Cells were cultured, RNA extracted and sscDNA was produced using standard procedures.

Panel 5I also contains pancreatic islets (Diabetes Research Institute at the University of
15 Miami School of Medicine).

Human Metabolic RTQ-PCR Panel

Human Metabolic RTQ-PCR Panel included two controls (genomic DNA control and chemistry control) and 211 cDNAs isolated from human tissues and cell lines relevant to metabolic diseases. This panel identifies genes that play a role in the etiology and pathogenesis of obesity
20 and/or diabetes. Metabolic tissues including placenta (Pl), uterine wall smooth muscle (Ut), visceral adipose, skeletal muscle (Sk) and subcutaneous (SubQ) adipose were obtained from the Gestational Diabetes study (described above). Included in the panel are: Patients 7 and 8, obese non-diabetic Caucasians; Patient 12 a diabetic Caucasian with unknown BMI, on insulin (treated); Patient 13, an overweight diabetic Caucasian, not on insulin (untreated); Patient 15, an obese, untreated, diabetic
25 Caucasian; Patient 17 and 25, untreated diabetic Caucasians of normal weight; Patient 18, an obese, untreated, diabetic Hispanic; Patient 19, a non-diabetic Caucasian of normal weight; Patient 20, an overweight, treated diabetic Caucasian; Patient 21 and 23, overweight non-diabetic Caucasians; Patient 22, a treated diabetic Caucasian of normal weight; Patient 23, an overweight non-diabetic Caucasian; and Patients 26 and 27, obese, treated, diabetic Caucasians.

30 Total RNA was isolated from metabolic tissues including: hypothalamus, liver, pancreas, pancreatic islets, small intestine, psoas muscle, diaphragm muscle, visceral (Vis) adipose, subcutaneous (SubQ) adipose and greater omentum (Go) from 12 Type II diabetic (Diab) patients and 12 non diabetic (Norm) at autopsy. Control diabetic and non-diabetic subjects were matched where possible for: age; sex, male (M); female (F); ethnicity, Caucasian (CC); Hispanic (HI); African
35 American (AA); Asian (AS); and BMI, 20-25 (Low BM), 26-30 (Med BM) or overweight (Overwt), BMI greater than 30 (Hi BMI) (obese).

RNA was extracted and ss cDNA was produced from cell lines (ATCC) by standard methods.

CNS Panels

CNS Panels CNSD.01, CNS Neurodegeneration V1.0 and CNS Neurodegeneration V2.0
40 included two controls and 46 to 94 test cDNA samples isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains were removed from calvaria of donors between 4 and 24 hours after death, and frozen at -80° C in liquid nitrogen vapor.

Panel CNSD.01

Panel CNSD.01 included two specimens each from: Alzheimer's disease, Parkinson's
45 disease, Huntington's disease, Progressive Supranuclear Palsy (PSP), Depression, and normal

5 controls. Collected tissues included: cingulate gyrus (Cing Gyr), temporal pole (Temp Pole), globus palladus (Glob palladus), substantia nigra (Sub Nigra), primary motor strip (Brodmann Area 4), parietal cortex (Brodmann Area 7), prefrontal cortex (Brodmann Area 9), and occipital cortex (Brodmann area 17). Not all brain regions are represented in all cases.

Panel CNS Neurodegeneration V1.0

10 The CNS Neurodegeneration V1.0 panel included: six Alzheimer's disease (AD) brains and eight normals which included no dementia and no Alzheimer's like pathology (control) or no dementia but evidence of severe Alzheimer's like pathology (Control Path), 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. Tissues collected included: hippocampus, temporal cortex (Brodmann Area 21), parietal cortex (Brodmann area 15 7), occipital cortex (Brodmann area 17) superior temporal cortex (Sup Temporal Ctx) and inferior temporal cortex (Inf Temporal Ctx).

Gene expression was analyzed after normalization using a scaling factor calculated by subtracting the Well mean (CT average for the specific tissue) from the Grand mean (average CT value for all wells across all runs). The scaled CT value is the result of the raw CT value plus the scaling factor.

Panel CNS Neurodegeneration V2.0

The CNS Neurodegeneration V2.0 panel included sixteen cases of Alzheimer's disease (AD) and twenty-nine normal controls (no evidence of dementia prior to death) including fourteen controls (Control) with no dementia and no Alzheimer's like pathology and fifteen controls with no dementia but evidence of severe Alzheimer's like pathology (AH3), specifically senile plaque load rated as level 25 3 on a scale of 0-3; 0 no evidence of plaques, 3 severe AD senile plaque load. Tissues from the temporal cortex (Brodmann Area 21) included the inferior and superior temporal cortex that was pooled from a given individual (Inf & Sup Temp Ctx Pool).

Example Q10. PathCalling® Technology

30 The sequence of NOVX was derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were 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.

35 The laboratory screening was performed using the methods that follow. 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, 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 40 6,083,693, incorporated herein by reference in their entireties).

5 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 reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated
10 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
15 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
20 diploid generated during the screening procedure by the mating of both CuraGen Corporation

Example Q11. Determination of Single Nucleotide Polymorphisms (SNPs)

Variant sequences are included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several
25 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
30 acid encoded by the gene at the position of the SNP. Intragenic-SNPs may also be silent, however, in the case that 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 for example, alteration in temporal expression, physiological response regulation,
35 cell type expression regulation, intensity of expression, stability of transcribed message.

Method of novel SNP Identification: SNPs are identified by analyzing sequence assemblies using CuraGen's proprietary SNPTool algorithm. SNPTool identifies variation in assemblies with the following criteria: SNPs are not analyzed within 10 base pairs on both ends of an alignment; Window size (number of bases in a view) is 10; The allowed number of mismatches in a window is 2;
40 Minimum SNP base quality (PHRED score) is 23; Minimum number of changes to score an SNP is 2/assembly position. SNPTool analyzes the assembly and displays SNP positions, associated individual variant sequences in the assembly, the depth of the assembly at that given position, the putative assembly allele frequency, and the SNP sequence variation. Sequence traces are then selected and brought into view for manual validation. The consensus assembly sequence is imported
45 into CuraTools along with variant sequence changes to identify potential amino acid changes

5 resulting from the SNP sequence variation. Comprehensive SNP data analysis is then exported into the SNPCalling database.

Method of novel SNP Confirmation: SNPs are confirmed employing a validated method known as Pyrosequencing. Detailed protocols for Pyrosequencing can be found in Genome Research 10:1249 (2000).

10 In brief, Pyrosequencing is a real time primer extension process of genotyping. This protocol takes double-stranded, biotinylated PCR products from genomic DNA samples and binds them to streptavidin beads. These beads are then denatured producing single stranded bound DNA. SNPs are characterized utilizing a technique based on an indirect bioluminometric assay of pyrophosphate (PPi) that is released from each dNTP upon DNA chain elongation. Following Klenow polymerase-
15 mediated base incorporation, PPi is released and used as a substrate, together with adenosine 5'-phosphosulfate (APS), for ATP sulfurylase, which results in the formation of ATP. Subsequently, the ATP accomplishes the conversion of luciferin to its oxi-derivative by the action of luciferase. The ensuing light output becomes proportional to the number of added bases, up to about four bases. To allow processivity of the method dNTP excess is degraded by apyrase, which is also present in the
20 starting reaction mixture, so that only dNTPs are added to the template during the sequencing. The process has been fully automated and adapted to a 96-well format, which allows rapid screening of large SNP panels.

Example Q12. Gene Expression analysis using CuraChip.

25 CuraGen has developed a gene microarray (CuraChip 1.2) for target identification. It provides a high-throughput means of global mRNA expression analyses of CuraGen's collection of cDNA sequences representing the Pharmaceutically Tractable Genome (PTG). This sequence set includes genes which can be developed into protein therapeutics, or used to develop antibody or small molecule therapeutics. CuraChip 1.2 contains almost 11,000 oligos representing approximately
30 8,500 gene loci, including (but not restricted to) kinases, ion channels, G-protein coupled receptors (GPCRs), nuclear hormone receptors, proteases, transporters, metabolic enzymes, hormones, growth factors, chemokines, cytokines, complement and coagulation factors, and cell surface receptors.

The CuraChip cDNAs were represented as 30-mer oligodeoxyribonucleotides (oligos) on a
35 glass microchip. Hybridization methods using the longer CuraChip oligos are more specific compared to methods using 25-mer oligos. CuraChip oligos were synthesized with a linker, purified to remove truncated oligos (which can influence hybridization strength and specificity), and spotted on a glass slide. Oligo-dT primers were used to generate cRNA probes for hybridization from samples of interest. A biotin-avidin conjugation system was used to detect hybridized probes with a fluorophore-
40 labeled secondary antibody. Gene expression was analyzed using clustering and correlation bioinformatics tools such as Spotfire® (Spotfire, Inc., 212 Elm Street, Somerville, MA 02144) and statistical tools such as multivariate analysis (MVA).

5 Normalization method used in CuraChip software

The median fluorescence intensity of each spot and a background for each spot is read on a scale from 0 to 65,000. CuraGen's CuraChip software, developed in-house, has the capability to present the user with either the raw data (median intensities) or normalized data. If normalized data is chosen, the CuraChip software uses the following method to do mean normalization. The

10 normalization is based on each slide/experiment. Suppose we have:

- *fg_median* is the signal/foreground median for each slide/experiment;
- *bg_median* is the background median for each slide/experiment;
- *original_value* is the difference between *fg_median* and *bg_median*;
- 15 • *flag* is an indicator of a spot's success or failure, where 0 means success and 1 means failure;
- *raw_fg_mean* is the raw foreground mean for each slide/experiment;
- *raw_bg_mean* is the raw background mean for each slide/experiment;
- *trim_percentage* is the trim percentage for each slide/experiment; this could be defined
- 20 by the user; currently we are using 2% as the trim percentage for each slide/experiment;
- *nSpots* is the number of spots on each slide;
- *nSlides* is the number of slides in each experiment;
- *fg_mean* is the trimmed foreground mean for each slide/experiment;
- *bg_mean* is the trimmed background mean for each slide/experiment;
- 25 • *max_fg_mean* is a constant among all slides/experiments, currently 2200.0;
- *normalized_value* is the final normalized value;
- *coeff* is the normalization co-efficient;
- *MAX_VALUE* is a constant representing the highest possible fluorescence reading, currently 65,000.

30

Step 1. Calculate trimmed foreground and background means

For each slide/experiment, we first calculate the trimmed foreground mean and the trimmed background mean of all spots, suppose *nSpots*, on each slide. For each spot, if the data is acceptable (*flag*=0), we calculate the raw foreground mean and background mean by subtracting the

35 background median from the foreground median for each spot. This is designated as a spot's "original value". (Note: If *flag*=1, all values are set to 0.)

original_value = *fg_median* - *bg_median*;

```
40 if ( flag == 0 )      // experiment is successful
{
    raw_fg_mean = original_value;
    raw_bg_mean = bg_median;
```

```

5   }
    else          // experiment is failed
    {
        raw_fg_mean = 0.0;
        raw_bg_mean = 0.0;

```

```

10  }

```

After that, we remove (trim) the top and bottom 2% of data points from the data set. After the above calculation, we have $nSpot$ number of foreground means and background means for each slide/experiment, and both lists are sorted. Suppose we have the following sorted lists:

```

15  raw_fg_mean[1], raw_fg_mean[2], ..., raw_fg_mean[N];    N = 1, nSpots;
    raw_bg_mean[1], raw_bg_mean[2], ..., raw_bg_mean[N];    N = 1, nSpots;

```

then we calculate the trimmed data points for each slide/experiment. Suppose a is the trimmed start data point and b is the trimmed end data point, we have:

```

20

```

```

    a = ceil(nSpots * trim_percentage);
    b = floor(nSpots * (1 - trim_percentage));

```

The "background mean" is calculated from the background medians for the trimmed data set. For the background mean, we simply calculate the average background mean in interval $[a,b]$ then assign to bg_mean :

```

25

```

```

    bg_mean = (raw_bg_mean[a] + raw_bg_mean[a+1] + ... + raw_bg_mean[b]) / (b-a+1);

```

The "foreground mean" is calculated from the "original values" (i.e. background-subtracted spot signal medians); only "original values" greater than 500 are used for this calculation (excluding the trimmed top and bottom 2% of the data). Suppose the sum of those foreground means is $sum_raw_fg_mean$ and the amount of those foreground means is k .

```

35  fg_mean = sum_raw_fg_mean / k;

```

For clarity, a snippet code in Java looks like the following,

```

    int    k = 0;
40  double sum_raw_fg_mean = 0.0;

    for (int j = a; j < b; j++) {
        if ( raw_fg_mean[j] > 500 ) {
            sum_raw_fg_mean = sum_raw_fg_mean + raw_fg_mean[j];
45            k++;

```

5

}

$$fg_mean = sum_raw_fg_mean / k;$$

- 10 After the calculation of trimmed foreground means and background means for all slides is complete, we start our normalization procedure.

Step 2. Normalize data

For each slide a normalization coefficient is calculated which compares the foreground mean of the slide to a fixed maximum foreground mean (2200). This coefficient is:

15

$$coeff = max_fg_mean / fg_mean;$$

- The normalized value of each spot is then calculated by multiplying the spot's "original value" by the normalization coefficient. Note that if this value is greater than the maximum reading of 65,000, then the value of 65,000 is used as the normalized value. Also note that if a spot's "original value" is less than the background value, the background value is used.

20

Recall that $original_value = fg_median - bg_median$

if ($original_value > bg_mean$) {

$$normalized_value = \min(coeff * original_value, MAX_VALUE);$$

25

} else {

$$normalized_value = coeff * bg_mean;$$

}

The $normalized_value$ for each spot is the final (normalized) value used in the analysis

Threshold for CuraChip data analysis

30

A number of control spots are present on CuraChip 1.2 for efficiency calculations and to provide alternative normalization methods. For example, CuraChip 1.2 contains a number of empty or negative control spots, as well as positive control spots containing a dilution series of oligos that detect the highly-expressed genes Ubiquitin and glyceraldehyde-3-phosphate dehydrogenase (GAPD). An analysis of spot signal level was performed using raw data from 67 hybridizations using all oligos. The maximum signal intensity for each oligo across all 67 hybridizations was determined, and the fold-over-background for this maximum signal was calculated (i.e. if the background reading is 20 and the raw spot intensity is 100, then the fold-over-background for that spot is 5x). The negative control or empty spots do occasionally "fire" or give a signal over the background level; however, they do not fire very strongly, with 77.1% of empty spots firing <3x over background and 91.7% <5x. The positive control spots (Ubiquitin and GAPD, the light blue and dark blue bars, respectively) always fired at >100x background. The experimental oligos (CuraOligos, in yellow below) fired over the entire range of intensities, with some at low fold-over-background intensities.

35

40

- 5 Since the negative control spots do fire occasionally at low levels, we have set a suggested threshold for data analysis at >5x background.

Thus while we have illustrated and described the preferred embodiment of our invention, it is to be understood that this invention is capable of variation and modification, and we therefore do not wish to be limited to the precise terms set forth, but desire to avail ourselves of such changes and
10 alterations which may be made for adapting the invention to various usages and conditions. Thus, such variations and modifications are properly intended to be within the full range of equivalents, and therefore within the purview of the following claims.

Having thus described our invention and the manner and a process of making and using it in such full, clear, concise and exact terms so as to enable any person skilled in the art to which it
15 pertains, or with which it is most nearly connected, to make and use the same, we claim:

5 **CLAIMS**

1. A method for identifying compounds that modulate target polypeptide activity comprising:
 - (a) combining a test compound with a target polypeptide and a substrate of the target polypeptide; and
 - 10 (b) determining whether the test compound modulates the activity of the target polypeptide; wherein the target polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 16, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain
 - 15 of SEQ ID NO:2n.
2. The method of claim 1, further comprising a step of identifying the test compound that modulates the target polypeptide activity by inhibiting the target polypeptide activity as an inhibitor of the target polypeptide activity.
- 20 3. The method of claim 1, further comprising a step of identifying the test compound that modulates the target polypeptide activity by inhibiting the target polypeptide activity as an antagonist of the target polypeptide.
- 25 4. The method of claim 1, further comprising a step of identifying the test compound that modulates the target polypeptide activity by activating the target polypeptide activity as an activator of the target polypeptide activity.
- 30 5. The method of claim 1, further comprising a step of identifying the test compound that modulates the target polypeptide activity by activating the target polypeptide activity as an agonist of the target polypeptide.
6. The method of claim 1, further comprising a step of identifying the test compound that modulates the target polypeptide activity as an enhancer of insulin secretion.
- 35 7. The method of claim 1, further comprising a step of identifying the test compound that modulates the target polypeptide activity as a therapeutic for treatment of insulin resistance.
8. The method of claim 1, further comprising a step of identifying the test compound that modulates
- 40 the target polypeptide activity as a therapeutic for treatment of obesity.
9. The method of claim 1, further comprising a step of identifying the test compound that modulates the target polypeptide activity as a therapeutic for treatment of diabetes.
- 45 10. The method of claim 1, wherein the target polypeptide is an isolated polypeptide.

- 5 11. The method of claim 1, wherein the target polypeptide is produced by expression of an endogenous nucleic acid, the endogenous nucleic acid encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 16, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain of SEQ ID NO:2n.
- 10 12. The method of claim 11, wherein the test compound is combined with the target polypeptide *in vitro*, or in a mammalian cell grown in culture.
- 15 13. The method of claim 11, wherein the endogenous nucleic acid comprises a nucleotide sequence selected from the group consisting of:
- (a) SEQ ID NO:2n-1, wherein n is an integer between 1 and 16;
 - (b) nucleotides encoding an amino acid sequence of the at least one domain of SEQ ID NO:2n; and
 - 20 (c) a nucleotide sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2n, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain of SEQ ID NO:2n.
- 25 14. An antibody that immunospecifically binds to the target polypeptide, wherein the target polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 16, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain of SEQ ID NO:2n.
- 30 15. The antibody of claim 14, wherein the antibody is selected from the group consisting of a monoclonal antibody, a humanized antibody, and a human antibody.
- 35 16. 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 target polypeptide, the method comprising:
- (a) providing a cell expressing the target polypeptide and having a property or function ascribable to the target polypeptide;
 - 40 (b) contacting the cell with a composition comprising a candidate test compound; and
 - (c) determining whether the test compound alters the property or function ascribable to the target polypeptide;
- whereby, if an alteration observed in the presence of the test compound is not observed when the cell is contacted with the composition in the absence of the test compound, the test compound is
- 45 identified as a potential therapeutic agent; and

5 wherein the target polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 16, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain of SEQ ID NO:2n.

10

17. A method for screening for a modulator of activity of or of latency or predisposition to a pathology associated with a target polypeptide, the method comprising:

- 15 (a) administering a test compound to a test animal at an increased risk for a pathology associated with the target polypeptide, wherein the test animal recombinantly expresses the target polypeptide;
- (b) measuring the activity of the target polypeptide in the test animal after administering the test compound of step (a); and
- 20 (c) comparing the activity of the target polypeptide in the test animal with the activity of the target polypeptide in a control animal not administered the test compound, wherein a change in the activity of the target polypeptide in the test animal relative to the control animal indicates that the test compound is a modulator of activity of or of latency or predisposition to, a pathology associated with the target polypeptide;

wherein the target polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 16, the amino acid sequence that is at least 95% identical to SEQ ID NO:2n, the amino acid sequence of at least one domain of SEQ ID NO:2n, and the amino acid sequence that is at least 95% identical to the at least one domain of SEQ ID NO:2n.