

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 March 2006 (30.03.2006)

PCT

(10) International Publication Number
WO 2006/034007 A2

(51) International Patent Classification:
A61K 38/54 (2006.01)

(21) International Application Number:
PCT/US2005/033076

(22) International Filing Date:
16 September 2005 (16.09.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/611,328 18 September 2004 (18.09.2004) US

(71) Applicant (for all designated States except US): **CURA-GEN CORPORATION** [US/US]; 322 E. Main Street, Brandford, CT 06405 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **ORT, Tatiana, A.** [RU/US]; 85 Viscount Drive Apt. 61-A, Milford, CT 06460 (US).

(74) Agent: **YAHWAK, George, M.**; c/o CuraGen Corporation, 322 E. Main Street, Brandford, CT 06405 (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, LY, 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 TREATING METABOLIC DISORDERS BY MODULATION OF SALT-INDUCIBLE SERINE/THREONINE KINASE 2

(57) Abstract: The invention relates to salt-inducible serine/threonine kinase polypeptides that are targets of small molecule drugs and that have properties relates to stimulation of biochemical or physiological responses in a cell, a tissue, an organ or an organism. Methods of use encompass screening, diagnostic and prognostic assay procedures as well as methods of treating diverse metabolic disorders.



WO 2006/034007 A2

METHODS OF TREATING METABOLIC DISORDERS BY MODULATION OF SALT-INDUCIBLE SERINE/THREONINE KINASE 2

5 RELATED APPLICATIONS

This application claims priority to provisional patent application U.S.S.N. 60/611,328, filed September 18, 2004, which is incorporated herein by reference in its entirety

10 FIELD OF THE INVENTION

The present invention relates to the field of therapeutics for the treatment of metabolic disorders. More specifically, the invention relates to salt-inducible serine/threonine kinase polypeptides that are targets of small molecule drugs and that have properties related to stimulation of biochemical or
15 physiological responses in a cell, a tissue, an organ or an organism. Methods of use encompass screening, diagnostic and prognostic assay procedures as well as methods of treating diverse metabolic disorders.

20 BACKGROUND OF THE INVENTION

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 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
25 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 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
30 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 (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
35 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 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 diabetes.

15 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.

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.

20 When such cells are components of multicellular organisms such as vertebrates, or more particularly organisms such as mammals, the regulation of the biochemical and physiological processes involves intricate signaling pathways. Frequently, such signaling pathways involve extracellular signaling proteins, cellular receptors that bind the signaling proteins and signal transducing components located within the cells.

25 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
30 example two different classes of cells in the same tissue or organ. One class of cells secretes the paracrine effector, which then reaches the second class of cells, for example by diffusion through the extracellular fluid. The second class of cells contains the receptors for the paracrine effector; binding of the effector results in induction of the signaling cascade that elicits the corresponding biochemical or physiological effect. Autocrine effectors are highly analogous to paracrine effectors, except that the same
35 cell type that secretes the autocrine effector also contains the receptor. Thus the autocrine effector binds to receptors on the same cell, or on identical neighboring cells. The binding process then elicits the characteristic biochemical or physiological effect.

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

5 Many pathological conditions involve dysregulation of expression of important effector proteins. In certain classes of pathologies the dysregulation is manifested as diminished or suppressed level of synthesis and secretion of protein effectors. In other classes of pathologies the dysregulation is manifested as increased or up-regulated level of synthesis and secretion of protein effectors. In a clinical setting a subject may be suspected of suffering from a condition brought on by altered or mis-regulated
10 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
15 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
20 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
25 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
30 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
35 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.

Salt-inducible kinase-2 (SIK2) is a serine/threonine protein kinase belonging to an AMP-activated protein kinase family (Kato et al., Mol Cell Endocrinol. 2004 Mar 31, 217(1-2):109-12; Okamoto et al., Trends Endocrinol Metab. 2004 Jan-Feb, 15(1):21-6). SIK2 was identified as an adipose specific kinase induced upon adipocyte differentiation (Horike et al., J Biol Chem. 2003 May 16, 278(20):18440-7). Moreover, Horike et al. showed that overexpression of SIK2 in adipocytes resulted in elevation of Ser789 phosphorylation of insulin receptor substrate-1 (IRS-1). Since serine phosphorylation of IRS-1 is known to inhibit insulin signaling (Schmitz-Peiffer, IUBMB Life. 2003 Jul, 55(7):367-74), several reports proposed that SIK2 may play important role(s) in modulating the insulin sensitivity (Kato et al., 2004; Okamoto et al., 2004; Horike et al., 2003).

DEFINITIONS

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.

As used herein, the terms and phrases "nucleic acid-molecule", "probe", "isolated", "oligonucleotide", "complementary", "fragment", "homologous nucleic acid sequence", "homologous amino acid sequence", "gene", "recombinant gene", "hybridizes under stringent conditions", "stringent hybridization conditions", "coding region", "noncoding region", "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. Furthermore, the terms 'Salt-inducible serine/threonine kinase 2', 'SIK2', and CG206886 are used interchangeably herein.

SUMMARY OF THE INVENTION

The invention includes nucleic acid sequences and the polypeptides they encode. In one aspect, the invention provides an isolated polypeptide comprising a mature form of a CG206886 amino acid. One example is a variant of a mature form of a CG206886 amino acid sequence, 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 CG206886 amino acid sequence or a variant of a CG206886 amino acid sequence, wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed. The invention also includes fragments of any of these. In another aspect, the invention also includes an isolated nucleic acid that encodes a CG206886 polypeptide, or a fragment, homolog, analog or derivative thereof.

Also included in the invention is a CG206886 polypeptide that is a naturally occurring allelic variant of a CG206886 sequence. In one embodiment, the allelic variant includes an amino acid sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a CG206886 nucleic acid sequence. In another embodiment, the CG206886 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 CG206886 polypeptide in a sample. The method involves the 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 CG206886 polypeptide, thereby determining the presence or amount of the CG206886 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 CG206886 polypeptide in a mammalian 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.

In a further embodiment, the invention includes a method of identifying an agent that modulates a CG206886 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.

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 CG206886 polypeptide. The method involves the steps of: providing a cell expressing the CG206886 polypeptide and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent. In another aspect, the invention describes a method for screening for a modulator of activity or of latency or predisposition to a

pathology associated with the CG206886 polypeptide. This method involves the following steps: administering a test compound to a test animal at increased risk for a pathology associated with the CG206886 polypeptide, wherein the test animal recombinantly expresses the CG206886 polypeptide. This method involves the steps of measuring the activity of the CG206886 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 CG206886 polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the CG206886 polypeptide in the 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 CG206886 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 CG206886 polypeptide, the method comprising introducing a cell sample expressing the CG206886 polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

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 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.

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.

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.

5 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 (CG206886) 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

10 (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
15 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.

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.
20

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;

25 (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.

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

Several cellular and 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
35 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 metabolic conditions such as obesity and/or diabetes.

The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with metabolic conditions such as obesity and/or diabetes. 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 CG206886 polypeptides.

The invention also includes an isolated nucleic acid that encodes a CG206886 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 CG206886 nucleic acid sequence. In one aspect, the invention provides a vector or a cell expressing a CG206886 nucleotide sequence.

In one embodiment, the invention discloses a method for modulating the activity of a CG206886 polypeptide. The method includes the steps of: introducing a cell sample expressing the CG206886 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 CG206886 nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising a CG206886 amino acid sequence or a variant of a mature form of the CG206886 amino acid sequence, wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. In another embodiment, the invention includes an amino acid sequence that is a variant of the CG206886 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.

In one embodiment, the invention discloses a CG206886 nucleic acid fragment encoding at least a portion of a CG206886 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 CG206886 nucleic acid molecules or a naturally occurring allelic nucleic acid variant. In another embodiment, the invention discloses a CG206886 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 CG206886 nucleic acid, wherein the nucleic acid molecule differs by a single nucleotide from a CG206886 nucleic acid sequence.

In another aspect, the invention includes a CG206886 nucleic acid, wherein one or more nucleotides in the CG206886 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 CG206886 nucleotide sequence and a nucleic acid fragment wherein one or more nucleotides in the CG206886 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 CG206886 nucleotide sequence or a complement of the CG206886 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 CG206886 nucleotide sequence or a fragment thereof.

In a further aspect, the invention includes a method for determining the presence or amount of the CG206886 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 CG206886 nucleic acid molecule, thereby determining the presence or amount of the CG206886 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.

In another aspect, the invention discloses a method for determining the presence of or predisposition to a disease associated with altered levels of the CG206886 nucleic acid molecule of in a first mammalian subject. The method involves the steps of: measuring the amount of CG206886 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 CG206886 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.

The materials, methods, and examples described herein 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 DRAWINGS

Figure 1 depicts extracellular and intracellular molecular pathways by whereby phosphorylation of IRS1/2 by the SIK2 protein leads to insulin resistance. Consequently, inhibition of SIK2 activity can be used to treat diabetes. Abbreviations: Ins = Insulin; IR = insulin receptor; IRS1/2 = Insulin Receptor Substrates 1 and 2; P = phosphorylation; S-P = serine phosphorylation; Glut4 = Glucose transporter 4; SIK2 = Salt-inducible serine/threonine kinase 2; S/T=serine/threonine; Jnk = c-Jun N-terminal kinase; IKK = I-kappa-B kinase; mTOR = mammalian target of rapamycin; PKCz = protein kinase C z; CK2 = casein kinase 2.

Figure 2 depicts protein-protein interactions. SIK2 amino acid sequence interacted with two protein tyrosine kinases – AXL Receptor Tyrosine Kinase (AXL) and Protein-Tyrosine Kinase 6 (PTK6) . PTK6 and the AXL amino acid sequence each interacted with polypeptides comprising portions of the VAV1 oncogene protein, or the RAB14 protein, or two regulatory subunits of PI3 kinase (PI3K; Phosphatidylinositol 3-kinase), PIK3R1, and PIK3R2.

DETAILED DESCRIPTION OF THE INVENTION

In particular the invention relates to the use of SIK2 polypeptides, and the polynucleotides encoding them, as diagnostic markers and/or targets for small molecule drugs and antibody therapeutics. The inventors have discovered that the SIK2 mRNA was upregulated 1.8-fold in the adipose tissue of a genetic model of obesity in mice. SIK2 mRNA was also observed to be significantly upregulated in adipose, skeletal muscle and hypothalamus tissues from diabetic patients compared to normal controls. Up-regulation of SIK2 in diabetic/obese animals and humans suggest that SIK2 may be involved in development and pathogenesis of type 2 diabetes and/or obesity. Using the yeast two-hybrid system, the inventors discovered that SIK2 may be involved in AXL and PTK6 signaling pathways. Specifically, the inventors showed that SIK2 directly interacted with AXL and PTK6 kinases that in turn interacted with regulatory subunits of PI3 kinase and the protein VAV1. By redistributing of PI3K away from insulin signaling and/or by activating Vav1 protein followed by induction of NF-kB/Jnk pathways PTK6 and AXL may promote insulin resistance/diabetes and/or obesity. SIK2 acting down-stream of AXL and PTK6 may contribute to propagation of metabolic diseases including insulin resistance, diabetes and/or obesity. The inventor concluded that modulation of SIK2 activity may be beneficial for the treatment diabetes and/or obesity.

In a particular embodiment of the invention, compounds that modulate the biological function or enzymatic activity of SIK2 are identified by screening. In a preferred embodiment, compounds that inhibit the enzymatic activity of SIK2 are identified by screening. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify SIK2

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 SIK2 has beneficial effects for treating obesity and/or diabetes by acting in many metabolic tissues, including adipose and skeletal muscle. Specifically, SIK2 inhibition will lead to an inhibition of Axl and PTK6 signaling that will beneficially affect insulin sensitivity in adipose and skeletal muscle. The finding of CG206886-01 upregulation in adipose from obese/diabetic animals and humans indicates a critical role for SIK2 in peripheral metabolism. Therefore, an antagonist of SIK2 is useful for the treatment of obesity and/or insulin resistance and diabetes.

SIK2 nucleic acids and proteins are useful for screening for an inhibitor/antagonist of SIK2 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 SIK2 activity, such as an inhibitor, activator, antagonist, or agonist of SIK2 may be useful for treatment of such disorders as obesity, diabetes, and insulin resistance, as well as for enhancement of insulin secretion.

Interfering RNA

In another aspect of the invention, CG206886 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 CG206886 gene are targeted by specific double stranded CG206886 derived siRNA nucleotide sequences that are complementary to at least a 19-25 nt long segment of the CG206886 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 CG206886 gene, or an upstream or downstream modulator of the CG206886 gene. Nonlimiting examples of upstream or downstream modulators of a CG206886 gene include, *e.g.*, a transcription factor that binds the CG206886 gene promoter, a kinase or phosphatase that interacts with a CG206886 polypeptide, and polypeptides involved in a CG206886 regulatory pathway.

According to the methods of the present invention, CG206886 gene expression is silenced using short interfering RNA. A CG206886 polynucleotide according to the invention includes a siRNA polynucleotide. Such a CG206886 siRNA can be obtained using a CG206886 polynucleotide sequence, for example, by processing the CG206886 ribopolynucleotide sequence in a cell-free system, such as but not limited to a *Drosophila* extract, or by transcription of recombinant double stranded CG206886 RNA or by chemical synthesis of nucleotide sequences homologous to a CG206886 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.

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. A contemplated recombinant expression vector of the invention comprises a CG206886 DNA molecule cloned into an expression vector comprising operatively-linked regulatory sequences flanking the CG206886 sequence in a manner that allows for expression (by transcription of the DNA molecule) of both strands. An RNA molecule that is antisense to CG206886 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 CG206886 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 silencing of the CG206886 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 siRNA. The regulatory sequences flanking the CG206886 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 CG206886 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 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 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 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 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 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 CG206886 mRNA region to be targeted by siRNA is generally selected from a desired CG206886 sequence beginning 50 to 100 nt downstream of the start codon. Alternatively, 5' or 3' UTRs and regions nearby the start codon can be used but are generally avoided, as these may be richer in 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.

In one embodiment, a complete CG206886 siRNA experiment includes the proper negative control. A negative control siRNA generally has the same nucleotide composition as the CG206886 siRNA but lack significant sequence homology to the genome. Typically, one would scramble the nucleotide sequence of the CG206886 siRNA and do a homology search to make sure it lacks homology to any other gene.

Two independent CG206886 siRNA duplexes can be used to knock-down a target CG206886 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 CG206886 siRNA duplexes, e.g., a CG206886 siRNA and an siRNA for a regulator of a CG206886 gene or polypeptide. Availability of siRNA-associating proteins is believed to be more limiting than target mRNA accessibility.

A targeted CG206886 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 CG206886 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 CG206886 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.

Alternatively, if the CG206886 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)].

Transfection of CG206886 siRNA duplexes can be achieved using standard nucleic acid transfection methods, for example, OLIGOFECTAMINE Reagent (available from Invitrogen). An assay for CG206886 gene silencing is generally performed approximately 2 days after transfection. No CG206886 gene silencing has been observed in the absence of transfection reagent, allowing for a comparative analysis of the wild-type and silenced CG206886 phenotypes. In a specific embodiment, for one well of a 24-well plate, approximately 0.84 µg 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 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 CG206886 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, 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 µg single-stranded sense CG206886 siRNA will have no effect on CG206886 silencing, and 0.84 µg antisense siRNA has a weak silencing effect when compared to 0.84 µg of duplex siRNAs. Control experiments again allow for a comparative analysis of the wild-type and silenced CG206886 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 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.

5 Depending on the abundance and the half life (or turnover) of the targeted CG206886 polynucleotide in a cell, a knock-down phenotype may become apparent after 1 to 3 days, or even later. In cases where no CG206886 knock-down phenotype is observed, depletion of the CG206886 polynucleotide may be observed by immunofluorescence or Western blotting. If the CG206886 polynucleotide is still abundant after 3 days, cells need to be split and transferred to a fresh 24-well plate
10 for re-transfection. If no knock-down of the targeted protein is observed, it may be desirable to analyze whether the target mRNA (CG206886 or a CG206886 upstream or downstream gene) was effectively 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
15 is also needed as control. Effective depletion of the mRNA yet undetectable reduction of target protein may indicate that a large reservoir of stable CG206886 protein may exist in the 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.

20 An inventive therapeutic method of the invention contemplates administering a CG206886 siRNA construct as therapy to compensate for increased or aberrant CG206886 expression or activity. The CG206886 ribopolynucleotide is obtained and processed into siRNA fragments, or a CG206886 siRNA is synthesized, as described above. The CG206886 siRNA is administered to cells or tissues using known nucleic acid transfection techniques. A CG206886 siRNA specific for a CG206886 gene will decrease or
25 knockdown CG206886 transcription products, which will lead to reduced CG206886 polypeptide production, resulting in reduced CG206886 polypeptide activity in the cells or tissues.

 The present invention also encompasses a method of treating a disease or condition associated with the presence of a CG206886 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
30 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 produced to the extent it would be in the absence of the treatment.

 Where the CG206886 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 CG206886
35 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 mammal, preferably a human subject, suffering from a disease state. The CG206886 ribopolynucleotide is used to

produce siRNA constructs that are specific for the CG206886 gene product. These cells or tissues are treated by administering CG206886 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 CG206886 polypeptide or polynucleotide expression is observed in the subject sample relative to the control sample, using the assays described. This CG206886 gene knockdown approach provides a rapid method for determination of a CG206886 minus (CG206886⁻) phenotype in the treated subject sample. The CG206886⁻ 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 CG206886 siRNA is used in therapy. Methods for the generation and use of a CG206886 siRNA are known to those skilled in the art.

Production of RNAs

Sense RNA (ssRNA) and antisense RNA (asRNA) of CG206886 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 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

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 CG206886 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 CG206886 mRNA is radiolabeled and its stability is monitored by gel electrophoresis.

In a parallel experiment made with the same conditions, the double stranded RNA is internally radiolabeled with a ³²P-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 stranded RNA can be determined.

The band of double stranded RNA, about 21-23 bps, is eluded. The efficacy of these 21-23 mers for suppressing CG206886 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.

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
5 (Waters, Milford, A) purification (Biochemistry, 32:11658 (1993)).

These RNAs (20 μ M) 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.

10 Cell Culture

A cell culture known in the art to regularly express CG206886 is propagated using standard 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 ml/well). Transfection is performed using a commercially available lipofection kit and CG206886
15 expression is monitored using standard techniques with positive and negative control. A positive control is cells that naturally express CG206886 while a negative control is cells that do not express CG206886. 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
20 concentration. This indicates that siRNAs are effective at concentrations that are several orders of magnitude below the concentrations applied in conventional antisense or ribozyme gene targeting experiments.

The above method provides a way both for the deduction of CG206886 siRNA sequence and the use of such siRNA for in vitro suppression. In vivo suppression may be performed using the same siRNA
25 using well known in-vivo transfection or gene therapy transfection techniques.

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequences of
30 CG206886, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire CG206886 coding
35 strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives

and analogs of a CG206886 protein, or antisense nucleic acids complementary to a CG206886 nucleic acid sequence are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a CG206886 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 CG206886 protein.

Given the coding strand sequences encoding the CG206886 protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of CG206886 mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of CG206886 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of CG206886 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 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, 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- 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.

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 CG206886 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 major groove of the double

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

In yet another embodiment, the antisense nucleic acid molecule of the invention is an 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

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

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave CG206886 mRNA transcripts to thereby inhibit translation of CG206886 mRNA. A ribozyme having specificity for a CG206886-encoding nucleic acid can be designed based upon the nucleotide sequence of a CG206886 cDNA disclosed herein. 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 CG206886-encoding mRNA. See, e.g., U.S. Patents 4,987,071 and 5,116,742. CG206886 mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules.

Alternatively, CG206886 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the CG206886 nucleic acid (e.g., the CG206886 promoter and/or enhancers) to form triple helical structures that prevent transcription of the CG206886 gene in target cells [see *Anticancer Drug Des.* 6:569 (1991); *Ann. N.Y. Acad. Sci.* 660:27 (1992); and *Bioassays* 14: 807 (1992)].

In various embodiments, the CG206886 nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids [see *Bioorg Med Chem* 4: 5 (1996)].

5 PNA of CG206886 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 CG206886 can also be used, for example, in the analysis of single base pair mutations in a gene (e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S₁ nucleases, or as probes or primers for DNA sequence and hybridization.

10 In another embodiment, PNAs of CG206886 can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of CG206886 can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (e.g., RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleotide bases, and orientation. 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 [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)]

25 In other embodiments, the oligonucleotide may include other appended groups such as 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 cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

Chimeric and Fusion Proteins

35 The invention also provides CG206886 chimeric or fusion proteins. An "CG206886 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a CG206886 protein of SEQ ID NO:2n, wherein n is an integer between 1 and 4, whereas a "non-CG206886 polypeptide" refers to a

polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the CG206886 protein, e.g., a protein that is different from the CG206886 protein and that is derived from the same or a different organism. Within a CG206886 fusion protein the CG206886 polypeptide can correspond to all or a portion of a CG206886 protein. In one embodiment, a CG206886 fusion protein comprises at least one biologically-active portion of a CG206886 protein. In another embodiment, a CG206886 fusion protein comprises at least two biologically-active portions of a CG206886 protein. In yet another embodiment, a CG206886 fusion protein comprises at least three biologically-active portions of a CG206886 protein.

In one embodiment, the fusion protein is a GST-CG206886 fusion protein in which the CG206886 sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant CG206886 polypeptides.

In another embodiment, the fusion protein is a CG206886 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of CG206886 can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a CG206886-immunoglobulin fusion protein in which the CG206886 sequences are fused to sequences derived from a member of the immunoglobulin protein family. The CG206886-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a CG206886 ligand and a CG206886 protein on the surface of a cell, to thereby suppress CG206886-mediated signal transduction *in vivo*. The CG206886-immunoglobulin fusion proteins can be used to affect the bioavailability of a CG206886 cognate ligand. Inhibition of the CG206886 ligand/CG206886 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 CG206886-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-CG206886 antibodies in a subject, to purify CG206886 ligands, and in screening assays to identify molecules that inhibit the interaction of CG206886 with a CG206886 ligand.

A CG206886 chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, e.g., by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence. Moreover, many expression vectors are commercially available that already

encode a fusion moiety (e.g., a GST polypeptide). A CG206886-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the CG206886 protein.

CG206886 Agonists and Antagonists

5 The invention also pertains to variants of the CG206886 proteins that function as either CG206886 agonists (*i.e.*, mimetics) or as CG206886 antagonists. Variants of the CG206886 protein can be generated by mutagenesis (e.g., discrete point mutation or truncation of the CG206886 protein). An agonist of the CG206886 protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the CG206886 protein. An antagonist of the CG206886 protein can
10 inhibit one or more of the activities of the naturally occurring form of the CG206886 protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the CG206886 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject
15 relative to treatment with the naturally occurring form of the CG206886 proteins.

 Variants of the CG206886 proteins that function as either CG206886 agonists or as CG206886 antagonists can be identified by screening combinatorial libraries of mutants (e.g., truncation mutants) of the CG206886 proteins for CG206886 protein agonist or antagonist activity. In one embodiment, a variegated library of CG206886 variants is generated by combinatorial mutagenesis at the nucleic acid
20 level and is encoded by a variegated gene library. A variegated library of CG206886 variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential CG206886 sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of CG206886 sequences therein. There are a variety of methods which can be used to produce libraries
25 of potential CG206886 variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential CG206886 sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art.

Polypeptide Libraries

 In addition, libraries of fragments of the CG206886 protein coding sequences can be used to generate a variegated population of CG206886 fragments for screening and subsequent selection of variants of a CG206886 protein. In one embodiment, a library of coding sequence fragments can be
35 generated by treating a double stranded PCR fragment of a CG206886 coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from

different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the CG206886 proteins.

5 Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of CG206886 proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene
10 library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify CG206886 variants [see *PNAS USA*
15 89:7811 (1992); *Protein Engineering* 6:327 (1993)].

Anti-CG206886 Antibodies

Included in the invention are antibodies to CG206886 proteins, or fragments of CG206886 proteins.

20 An isolated protein of the invention intended to serve as an antigen, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid
25 sequence of the full length protein, such as an amino acid sequence of SEQ ID NO:2n, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide
30 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 CG206886 that is located on the surface of the protein, e.g., a hydrophilic region. A hydrophobicity analysis of the human CG206886 protein sequence will indicate which regions of a CG206886 polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues
35 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 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.

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 CG206886 polypeptide or a fragment thereof comprises at least one antigenic epitope. An anti-CG206886 antibody of the present invention is said to specifically bind to antigen CG206886 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.

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

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the 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 residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin..

Human Antibodies

Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies. 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: 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 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.

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 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 XenomouseTM [see PCT publications WO 96/33735 and WO 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 obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

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

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

A method for producing an antibody of interest, such as a human antibody (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.

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

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 that contain the idiotype to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F_{(ab')₂} fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an F_{(ab')₂} fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

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

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (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 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 immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, *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 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)].

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

Heteroconjugate Antibodies

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

Effector Function Engineering

It 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 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

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

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

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

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

Immunoliposomes

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art [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)].

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 CG206886 protein is facilitated by generation of hybridomas that bind to the fragment of an CG206886 protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an CG206886 protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Antibodies directed against a CG206886 protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of a CG206886 protein (e.g., for use in measuring levels of the CG206886 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 CG206886 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").

An antibody specific for a CG206886 protein of the invention (e.g., a monoclonal antibody or a polyclonal antibody) can be used to isolate a CG206886 polypeptide by standard techniques, such as immunoaffinity, chromatography or immunoprecipitation. An antibody to a CG206886 polypeptide can facilitate the purification of a natural CG206886 antigen from cells, or of a recombinantly produced CG206886 antigen expressed in host cells. Moreover, such an anti-CG206886 antibody can be used to detect the antigenic CG206886 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic CG206886 protein. Antibodies directed against a CG206886 protein can be used diagnostically to monitor protein levels in tissue as part of a clinical

testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include
5 horseradish peroxidase, alkaline phosphatase, b-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and
10 examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibody Therapeutics

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

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an
25 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
30 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
35 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.

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-(methacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

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

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

ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

CG206886 Recombinant Expression Vectors and Host Cells

5 Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a CG206886 protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral
10 vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable
15 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 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
20 defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

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

The term "regulatory sequence" is intended to include promoters, enhancers and other
30 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 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
35 design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, *etc.* The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or

peptides, encoded by nucleic acids as described herein (e.g., CG206886 proteins, mutant forms of CG206886 proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of CG206886 proteins in prokaryotic or eukaryotic cells. For example, CG206886 proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells 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.

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.

In another embodiment, the CG206886 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, CG206886 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
5 promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific;
10 *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,
15 e.g., the murine hox promoters (*Science* 249:374 (1990)) and the α -fetoprotein promoter (*Genes Dev.* 3:537 (1989)).

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
20 DNA molecule) of an RNA molecule that is antisense to CG206886 mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the
25 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.

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

A host cell can be any prokaryotic or eukaryotic cell. For example, CG206886 protein can be
35 expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated
5 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 of interest.
10 Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding CG206886 or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) CG206886 protein. Accordingly, the invention further provides methods for producing CG206886 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 CG206886 protein has been introduced) in a suitable medium such that CG206886 protein is produced.
15 In another embodiment, the method further comprises isolating CG206886 protein from the medium or the host cell.
20

Transgenic CG206886 Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For
25 example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which CG206886 protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous CG206886 sequences have been introduced into their genome or homologous recombinant animals in which endogenous CG206886 sequences have been altered. Such animals are useful for studying the function and/or activity of
30 CG206886 protein and for identifying and/or evaluating modulators of CG206886 protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, *etc.* A transgene is exogenous DNA that is integrated into the genome of a cell from which a
35 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 CG206886 gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal in accordance with the invention can be created by introducing a
5 CG206886-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 CG206886 cDNA sequences, i.e., any one of SEQ ID NO:2n-1 may be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human
10 CG206886 gene, such as a mouse CG206886 gene, can be isolated based on hybridization to the human CG206886 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 CG206886 transgene to direct expression of CG206886 protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become
15 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 CG206886 transgene in its genome and/or expression of CG206886 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a
20 transgene-encoding CG206886 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 CG206886 gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the CG206886 gene. The CG206886 gene can be a human gene (e.g.,
25 the cDNA of any one of SEQ ID NO:2n-1, but more preferably, is a non-human homologue of a human CG206886 gene. For example, a mouse homologue of human CG206886 gene of SEQ ID NO:2n-1 can be used to construct a homologous recombination vector suitable for altering an endogenous CG206886 gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous CG206886 gene is functionally disrupted (i.e., no longer encodes a
30 functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous CG206886 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 CG206886 protein). In the homologous recombination vector, the altered portion of the CG206886 gene is flanked at
35 its 5'- and 3'-termini by additional nucleic acid of the CG206886 gene to allow for homologous recombination to occur between the exogenous CG206886 gene carried by the vector and an endogenous CG206886 gene in an embryonic stem cell. The additional flanking CG206886 nucleic acid

is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector [see 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 CG206886 gene has homologously-recombined with the endogenous CG206886 gene are selected [see Cell, 69:915 (1992)].

The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras. See, e.g., Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in *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 selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, see, *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 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) 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 will be a clone of the animal from which the cell is isolated.

Pharmaceutical Compositions

The CG206886 nucleic acid molecules, CG206886 proteins, and anti-CG206886 antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and

homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration.

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

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

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

include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent that delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a
5 CG206886 protein or anti-CG206886 antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization.

Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are
10 vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

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

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

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

Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

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

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration 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 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 CG206886 protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect CG206886 mRNA (*e.g.*, in a biological sample) or a genetic lesion in a CG206886 gene, and to modulate CG206886 activity, as described further, below. In addition, the CG206886 proteins can be used to screen drugs or compounds that modulate the CG206886 protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of CG206886 protein or production of CG206886 protein forms that have decreased or aberrant activity compared to CG206886 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-CG206886 antibodies of the invention can be used to detect and isolate CG206886 proteins and modulate CG206886 activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

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

Screening Assays

5 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 CG206886 proteins or have a stimulatory or inhibitory effect on, *e.g.*, CG206886 protein expression or CG206886 protein activity. The invention also includes compounds identified in the screening assays described herein.

10 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 CG206886 protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring
15 deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. See, *e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

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

Examples of methods for the synthesis of molecular libraries can be found in the art, for example
25 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 No. 5,223,409), spores (U.S. Patent 5,233,409), plasmids (*PNAS USA* 89:1865 (1992)) or on phage (U.S. Patent No.
30 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of CG206886 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 CG206886 protein
35 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 CG206886 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 CG206886 protein or biologically-active portion thereof can be determined by detecting the labeled

compound in a complex. For example, test compounds can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of
5 conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of CG206886 protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds CG206886 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 CG206886 protein, wherein determining the ability of the test compound to interact with a
10 CG206886 protein comprises determining the ability of the test compound to preferentially bind to CG206886 protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of CG206886 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
15 inhibit) the activity of the CG206886 protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of CG206886 or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the CG206886 protein to bind to or interact with a CG206886 target molecule. As used herein, a "target molecule" is a molecule with which a CG206886 protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a
20 CG206886 interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A CG206886 target molecule can be a non-CG206886 molecule or a CG206886 protein or polypeptide of the invention. In one embodiment, a CG206886 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
25 compound to a membrane-bound CG206886 molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with CG206886.

Determining the ability of the CG206886 protein to bind to or interact with a CG206886 target molecule can be accomplished by one of the methods described above for determining direct binding. In
30 one embodiment, determining the ability of the CG206886 protein to bind to or interact with a CG206886 target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (i.e., intracellular Ca^{2+} , diacylglycerol, IP_3 , etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (comprising a
35 CG206886-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a CG206886 protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the CG206886 protein or biologically-active portion thereof. Binding of the test compound to the CG206886 protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the CG206886 protein or biologically-active portion thereof with a known compound which binds CG206886 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 CG206886 protein, wherein determining the ability of the test compound to interact with a CG206886 protein comprises determining the ability of the test compound to preferentially bind to CG206886 or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting CG206886 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 CG206886 protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of CG206886 can be accomplished, for example, by determining the ability of the CG206886 protein to bind to a CG206886 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 CG206886 protein can be accomplished by determining the ability of the CG206886 protein further modulate a CG206886 target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the CG206886 protein or biologically-active portion thereof with a known compound which binds CG206886 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 CG206886 protein, wherein determining the ability of the test compound to interact with a CG206886 protein comprises determining the ability of the CG206886 protein to preferentially bind to or modulate the activity of a CG206886 target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of CG206886 protein. In the case of cell-free assays comprising the membrane-bound form of CG206886 protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of CG206886 protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton[®] X-100, Triton[®] X-114, Thesit[®], Isotridecypoly(ethylene glycol ether)_n, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either CG206886 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 CG206886 protein, or interaction of CG206886 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-CG206886 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 CG206886 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 CG206886 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 CG206886 protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated CG206886 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 CG206886 protein or target molecules, but which do not interfere with binding of the CG206886 protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or CG206886 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 CG206886 protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the CG206886 protein or target molecule.

In another embodiment, modulators of CG206886 protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of CG206886 mRNA or protein in the cell is determined. The level of expression of CG206886 mRNA or protein in the presence of the candidate compound is compared to the level of expression of CG206886 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of CG206886 mRNA or protein expression based upon this comparison. For example, when expression of CG206886 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 CG206886 mRNA or protein expression. Alternatively, when expression of CG206886 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of CG206886 mRNA or protein expression. The level of

CG206886 mRNA or protein expression in the cells can be determined by methods described herein for detecting CG206886 mRNA or protein.

In yet another aspect of the invention, the CG206886 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 CG206886 ("CG206886-binding proteins" or "CG206886-bp") and modulate CG206886 activity. Such CG206886-binding proteins are also involved in the propagation of signals by the CG206886 proteins as, for example, upstream or downstream elements of the CG206886 pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for CG206886 is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a CG206886-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein that interacts with CG206886.

In yet another aspect of the invention a method for identifying compounds that modulate target polypeptide (CG206886) 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 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.

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 CG206886 target polypeptide.

The method also further comprises a step of identifying the test compound that modulates 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 (CG206886) could be an isolated polypeptide.

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 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.

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.

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.

Diagnostic Assays

5 An exemplary method for detecting the presence or absence of CG206886 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 CG206886 protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes CG206886 protein such that the presence of CG206886 is detected in the biological sample. An agent for detecting CG206886 mRNA or genomic DNA is a labeled nucleic acid probe
10 capable of hybridizing to CG206886 mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length CG206886 nucleic acid, such as the nucleic acid of SEQ ID NO:2n-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 CG206886 mRNA or genomic DNA.

An agent for detecting CG206886 protein is an antibody capable of binding to CG206886 protein,
15 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.

20 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 CG206886 mRNA, protein, or genomic
25 DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of CG206886 mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of CG206886 protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of CG206886 genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of CG206886 protein
30 include introducing into a subject a labeled anti-CG206886 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
35 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 CG206886 protein, mRNA, or genomic DNA, such that the presence of CG206886 protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of CG206886 protein, mRNA or genomic DNA in the control sample with the presence of CG206886 protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of CG206886 in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting CG206886 protein or mRNA in a biological sample; means for determining the amount of CG206886 in the sample; and means for comparing the amount of CG206886 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 CG206886 protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant CG206886 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 CG206886 protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant CG206886 expression or activity in which a test sample is obtained from a subject and CG206886 protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of CG206886 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant CG206886 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant CG206886 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 CG206886 expression or activity in which a test sample is obtained and CG206886 protein or nucleic acid is detected (e.g., wherein the presence of CG206886 protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant CG206886 expression or activity).

The methods of the invention can also be used to detect genetic lesions in a CG206886 gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a CG206886-protein, or the misexpression of the CG206886 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 CG206886 gene; (ii) an addition of one or more nucleotides to a CG206886 gene; (iii) a substitution of one or more nucleotides of a CG206886 gene, (iv) a chromosomal rearrangement of a CG206886 gene; (v) an alteration in the level of a messenger RNA transcript of a CG206886 gene, (vi) aberrant modification of a CG206886 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 CG206886 gene, (viii) a non-wild-type level of a CG206886 protein, (ix) allelic loss of a CG206886 gene, and (x) inappropriate post-translational modification of a CG206886 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 CG206886 gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (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 in the CG206886-gene (*Nucl. Acids Res.* 23: 675 (1995)).

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 CG206886 gene under conditions such that hybridization and amplification of the CG206886 gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*PNAS USA* 87:1874 (1990)), transcriptional amplification system (*PNAS USA* 86:1173 (1989)); Q β Replicase (*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 CG206886 gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in CG206886 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotide probes [see *Human Mutation* 7:244 (1996); *Nat. Med.* 2:753 (1996)]. For example, genetic mutations in CG206886 can be identified in two-dimensional arrays containing 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 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 CG206886 gene and detect mutations by comparing the sequence of the sample CG206886 with the corresponding wild-type (control) sequence [e.g. those described in *PNAS 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 CG206886 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA 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 CG206886 sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S₁ nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation [see *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 CG206886 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 CG206886 sequence, e.g., a wild-type CG206886 sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like [see U.S. Patent No. 5,459,039].

In other embodiments, alterations in electrophoretic mobility may be used to identify mutations in CG206886 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 CG206886 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)].

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)].

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 oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization [see *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 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 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 CG206886 gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which CG206886 is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on CG206886 activity (e.g., CG206886 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 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 CG206886 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 the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of CG206886 protein, expression of CG206886 nucleic acid, or mutation content of CG206886 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. 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, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

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

Thus, the activity of CG206886 protein, expression of CG206886 nucleic acid, or mutation content of CG206886 genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a CG206886 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 CG206886 (e.g., the ability to modulate aberrant cell proliferation and/or differentiation) can be applied

not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase CG206886 gene expression, protein levels, or upregulate CG206886 activity, can be monitored in clinical trials of subjects exhibiting decreased CG206886 gene expression, protein levels, or downregulated CG206886 activity.

5 Alternatively, the effectiveness of an agent determined by a screening assay to decrease CG206886 gene expression, protein levels, or downregulate CG206886 activity, can be monitored in clinical trials of subjects exhibiting increased CG206886 gene expression, protein levels, or upregulated CG206886 activity. In such clinical trials, the expression or activity of CG206886 and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read
10 out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including CG206886, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) that modulates CG206886 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
15 prepared and analyzed for the levels of expression of CG206886 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 CG206886 or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the
20 physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising
25 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 CG206886 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 CG206886 protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the CG206886 protein, mRNA, or genomic DNA in the pre-administration sample with the CG206886 protein, mRNA, or genomic
30 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 CG206886 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
35 or activity of CG206886 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 CG206886 expression or activity. The disorders include but are not 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 CG206886 protein, such as those summarized in Table 1.

Diseases and Disorders

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

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

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

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant CG206886 expression or activity, by administering to the subject an agent that modulates CG206886 expression or at least one CG206886 activity. Subjects at risk for a disease that is caused or contributed to by aberrant CG206886 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the CG206886 aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of CG206886 aberrancy, for example, a CG206886 agonist or CG206886 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating CG206886 expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of CG206886 protein activity associated with the cell. An agent that modulates CG206886 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a CG206886 protein, a peptide, a CG206886 peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more CG206886 protein activity. Examples of such stimulatory agents include active CG206886 protein and a nucleic acid molecule encoding CG206886 that has been introduced into the cell. In another embodiment, the agent inhibits one or more CG206886 protein activity. Examples of such inhibitory agents include antisense CG206886 nucleic acid molecules and anti-CG206886 antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a CG206886 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) CG206886 expression or activity. In another embodiment, the method involves administering a CG206886 protein or nucleic acid molecule as therapy to compensate for reduced or aberrant CG206886 expression or activity.

Stimulation of CG206886 activity is desirable *in situations* in which CG206886 is abnormally downregulated and/or in which increased CG206886 activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation

and/or differentiation (e.g., cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (e.g., preclampsia).

Determination of the Biological Effect of the Therapeutic

5 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.

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

15 Prophylactic and Therapeutic Uses of the Compositions of the Invention

The CG206886 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 CG206886 protein, such as those summarized in Table 1.

20 As an example, a cDNA encoding the CG206886 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.

25 Both the novel nucleic acid encoding the CG206886 protein, and the CG206886 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.

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

Discovery Process

The following sections describe the study designs and the techniques used to identify the SIK2 – encoded protein, and any variants thereof, as being suitable as diagnostic markers, targets for antibody therapeutic and targets for small molecule drugs for treatment of Obesity and Diabetes. Procedures useful in the identification of small molecule modulators of SIK2 activity are also described.

Example 1. Genetically Obese Mice vs Genetically Lean Mice Study

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 mouse 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. Gene expression was measured in several tissues including whole brain, skeletal muscle, visceral adipose, and liver.

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 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

The differentially expressed gene fragment is isolated, cloned into a plasmid, and sequenced.

- 5 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

- 10 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)

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.

- 15 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.

20

Results

- A fragment of the mouse salt inducible kinase-2 (SIK2) gene was initially found to be up-regulated by 1.8 fold in adipose from genetically obese mice (NZB strain) relative to adipose from C57L/J average weight mice using CuraGen's GeneCalling® method of differential gene expression. A differentially expressed mouse gene fragment migrating at approximately 174 nucleotides in length was definitively identified as a component of the mouse SIK2 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 SIK2 were ablated when a gene-specific primer (shown in Table 1) competes with primers in the linker-adaptors during the PCR amplification. The peaks at 174 nt in length were ablated in the sample from both the genetically obese NZB mice and the average weight C57L/J strain.

Table 1. Competitive PCR primer for the mouse salt inducible kinase-2 (Sik2): The nucleotide sequence of 2985 nucleotide-long gene from nucleotide 1470 to 2603 has shown below. A gene fragment (from 1951 to 2123 nt, band size: 173nt) is shown in bold, and the sequences of the gene-specific primers used for competitive PCR are underlined on the cDNA sequence of the mouse Sik2.

1470 GACAGAAGGA GAGGCTGAGG AAGACCCCAG TCAGGCCTTT GAAGCTTTCC AGGCCACACG
1530 CAGTGGGCAG CGACGGCACA CTCTGTCAGA AGTGACTAAC CAATTGGTTG TGATGCCCCG
1590 GGCAGGGAAA ATGTTCTCCA TGAGTGATAA CCCCTCCCTT GAGAGTGTGG ACTCTGAGTA
1650 TGATATGGGG TCTGCCCAGA GGGACCTGAA CTTTCTGGAA GACAGCCCTT CCTTGAAGGA
1710 CATCATGTTA GCCAATCAGC CGTCACCCCG CATGACATCT CCTTTCATAA GCCTCAGACC
1770 TGCCAACCCA GCCATGCAGG CTCTGAGCTC ACAGAAGCGT GAGGCCACCA ACCGGTCCCC
1830 CGTGAGCTTC CGAGAAGGCC GCAGAGCATC AGATACGTCC CTTACACAAG GAATTGTGGC
1890 ATTTAGACAA CATCTTCAGA ATCTTGCTAG AACCAAAGGA ATTCTGGAGT TGAACAAAGT
1950 ACAATTGCTG TATGAACAAA TGGGATCAAA CGCAGACCCT ACCTTAACAT CAACTGCTCC
2010 TCAGCTCCAA GACCTTTCGA GCAGTTGCCC TCAGGAGGAA ATCTCCCAGC AGCAGGAAAG
2070 TGTCTCCAGC CTGTCTGCCA GCATGCACCC TCAGCTCTCA CCACAGCAAA GCTTGGAAC
2130 CCAGTACCTA CAGCATCGAC TCCAGAAGCC CAATCTTCTG CCAAAGGCCC AGAGTCCTTG
2190 TCCAGTGTAT TGTAAGAGC CACCTCGGAG CCTGGAACAG CAGCTACAGG AGCATAGGCT
2250 CCAACAGAAG CGACTCTTCC TCCAGAAGCA GTCTCAGCTG CAAGCATATT TTAATCAGAT
2310 GCAGATAGCA GAGAGCTCCT ACCCTGGACC AAGTCAGCAG CTGGCTCTTC CCCACCAGGA
2370 GACTCCACTG ACATCCCAGC AGCCCCATC ATTCAGCCTG ACCCAGGCCC TGAGTCCTGT
2430 CCTCGAGCCC TCTTCTGAGC AGATGCAATT TAGCTCTTTC CTCAGCCAAT ACCCAGAGAT
2490 GCAGTTGCAG CCACTGCCCT CTACTCCCGG CCCCCGGGCT CCACCTCCCT TACCCTCACA

2550 ATTGCAGCAA CATCAGCAGC CACCACCCCC ACCACCACCC CCTCCACCAC AGCA

[SEQ ID NO:13]

Example 2. CG206886 nucleic acids and polypeptides; Molecular cloning.

CG206886-01 sequences were derived by in-silico sequence prediction. CG206886-02, CG206886-03, and CG206886-04 represent physical clones that were derived by laboratory screening of cDNA libraries. CG206886-02 and CG206886-04 nucleic acid sequences encode polypeptides which comprise a poly-histidine (His-6) tag which is useful, for example, in the purification of the recombinantly expressed polypeptides. In silico prediction was based on sequences available in CuraGen Corporation's proprietary sequence databases or in the public human sequence databases, and provided either the full-length DNA sequence, or some portion thereof. The laboratory cloning was performed using one or more of the methods summarized below:

RACE: Techniques based on the polymerase chain reaction such as rapid amplification of cDNA ends (RACE), were used to isolate or complete the predicted sequence of the cDNA of the invention.

Usually multiple clones were sequenced from one or more human samples to derive the sequences for fragments. Various human tissue samples from different donors were used for the RACE reaction. The sequences derived from these procedures were included in the SeqCalling Assembly process described in preceding paragraphs.

Exon Linking: The cDNAs coding for the CG206886 sequences were cloned by the polymerase chain reaction (PCR) using the primers designed based on known cDNA sequences or in silico 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 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.

The CG206886 nucleotide and encoded polypeptide sequences are shown in Table 2

Table 1A. NOV1 Sequence Analysis		
CG206886-01	SEQ ID NO: 1	4870 bp

DNA Sequence	ORF Start: ATG at 97	ORF Stop: TAG at 2875
GAGCAAGCGGAGCGGCCGTCGCCCAAGCCAAGCCGCGCTGCCAACCCCTCCCGCCCCGCCGCGCTCCT		
G		
TCCGCCGTGTCTAGCAGCGGGGCCAGC ATGGT CATGGCGGATGGCCCGAGGCACTTGCAGCGCGGG		
C		
CGGTCCGGGTGGGGTTCTACGACATCGAGGGCACGCTGGGCAAGGGCAACTTCGCTGTGGTGAAGCT		
G		
GGGCGGCACCGGATCACCAAGACGGAGGTGGCAATAAAAAATAATCGATAAGTCTCAGCTGGATGCAG		
T		
GAACCTTGAGAAAATCTACCGAGAAGTACAAATAATGAAAATGTTAGACCACCCTCACATAATCAAA		
C		
TTTATCAGGTAATGGAGACCAAAAGTATGTTGTACCTTGTGACAGAATATGCCAAAAATGGAGAAAT		
T		
TTTGACTATCTTGCTAATCATGGCCGGTTAAATGAGTCTGAAGCCAGGCGAAAATTCTGGCAAATCC		
T		
GTCTGCTGTTGATTATTGTTCATGGTCGGAAGATTGTGCACCGTGACCTCAAAGCTGAAAATCTCCTG		
C		
TGGATAACAACATGAATATCAAAATAGCAGATTTTCGGTTTTGGAAATTTCTTTAAAAGTGGTGAAC		
G		
CTGGCAACATGGTGTGGCAGCCCCCTTATGCAGCCCCAGAAGTCTTTGAAGGGCAGCAGTATGAAG		
G		
ACCACAGCTGGACATCTGGAGTATGGGAGTTGTTCTTTATGTCCTTGTCTGTGGAGCTCTGCCCTTT		
G		
ATGGACCGACTCTTCCAATTTTGAGGCAGAGGGTCTGGAAGGAAGATTCCGGATTCCGTATTTTCAT		
G		
TCAGAAGATTGCGAGCACCTTATCCGAAGGATGTTGGTCCTAGACCCATCCAAACGGCTAACCATAG		
C		
CCAAATCAAGGAGCATAAATGGATGCTCATAGAAGTTCCTGTCCAGAGACCTGTTCTCTATCCACAA		
G		
AGCAAGAAAATGAGCCATCCATCGGGGAGTTTAATGAGCAGGTTCTGCGACTGATGCACAGCCTTGG		
A		
ATAGATCAGCAGAAAACCATTGAGTCTTTGCAGAACAAGAGCTATAACCACTTTGCTGCCATTTATT		
T		
CTTGTTGGTGGAGCGCCTGAAATCACATCGGAGCAGTTTCCAGTGGAGCAGAGACTTGATGGCCGC		
C		
AGCGTCGGCCTAGCACCATTGCTGAGCAAACAGTTGCCAAGGCACAGACTGTGGGGCTCCCAGTGAC		
C		
ATGCATTACCGAACATGAGGCTGCTGCGATCTGCCCTCCTCCCCAGGCATCCAACGTGGAGGCCT		
T		
TTCATTTCCAGCATCTGGCTGTGAGCGGAAGCTGCATTTCATGGAAGAAGAGTGTGTGGACACTCCA		
A		
AGGTCAATGGCTGTCTGCTTGACCCTGTGCCTCCTGTCTGGTGCAGGAAGGGATGCCAGTCACTGCC		
C		
AGCAACATGATGGAGACCTCCATTGACGAAGGGCTGGAGACAGAAGGAGAGGCCGAGGAAGACCCCG		
C		
TCATGCCTTTGAGGCATTTAGTCCACACGCAGCGGGCAGAGACGGCACACTCTGTGAGAAGTGACC		
A		
ATCAACTGGTCGTGATGCCTGGGGCAGGGAAAATTTTCTCCATGAATGACAGCCCCCTCCCTTGACAG		
T		
GTGGACTCTGAGTATGATATGGGGTCTGTTTCAGAGGGACCTGAACTTTCTGGAAGACAACCCTTCCC		
T		
TAAGGACATCATGTTAGCCAATCAGCCTTCACCCCGCATGACATCTCCCTTCATAAGCCTGAGACCT		
A		
CCAACCCAGCCATGCAGGCTCTGAGCTCCAGAAACGAGAGGTCCACAACAGGTCTCCAGTGAGCTT		
C		
AGAGAGGGCCGCAGAGCATCAGATACCTCCCTCACCCAGGGAATTGTAGCATTTAGACAACATCTTC		

A
GAATCTGGCTAGAACCAAAGGAATTCTAGAGTTGAACAAAGTGCAGTTGTTGTATGAACAAATAGGA
C
CGGAGGCAGACCCTAACCTGGCGCCGGCGGCTCCTCAGCTCCAGGACCTTGCTAGCAGCTGCCCTCA
G
GAAGAAGTTTCTCAGCAGCAGGAAAGCGTCTCCACTCTCCCTGCCAGCGTGCATCCCCAGCTGTCCC
C
ACGGCAGAGCCTGGAGACCCAGTACCTGCAGCACAGACTCCAGAAGCCCAGCCTTCTGTCAAAGGCC
C
AGAACACCTGTCAGCTTTATTGCAAAGAACCACCGCGGAGCCTTGAGCAGCAGCTGCAGGAACATAG
G
CTCCAGCAGAAGCGACTCTTTCTTCAGAAGCAGTCTCAACTGCAGGCCTATTTTAATCAGATGCAGA
T
AGCAGAGAGCTCCTACCCACAGCCAAGTCAGCAGCTGCCCCCTCCCCGCCAGGAGACTCCACCGCCT
T
CTCAGCAGGCCCCACCGTTCAGCCTGACCCAGCCCCCTGAGCCCCGTCTGGAGCCTTCTCCGAGCA
G
ATGCAATACAGCCCTTTCCTCAGCCAGTACCAAGAGATGCAGCTTCAGCCCCTGCCCTCCACTTCCG
G
TCCCCGGGCTGCTCCTCCTCTGCCACGCAGCTACAGCAGCAGCAGCCGCCACCGCCACCACCCCT
C
CACCACCACGACAGCCAGGAGCTGCCCCAGCCCCCTTACAGTTCTCCTATCAGACTTGTGAGCTGCC
A
AGCGCTGCTTCCCCGCGCCAGACTATCCCACTCCCTGTCAGTATCCTGTGGATGGAGCCCAGCAGA
G
CGACCTAACGGGGCCAGACTGTCCCAAGAGCCCAGGACTGCAAGAGGCCCCCTCCAGCTACGACCCA
C
TAGCCCTCTCTGAGCTACCTGGACTCTTTGATTGTGAAATGCTAGACGCTGTGGATCCACAACACAA
C
GGGTATGTCTGGTGAATTAGTCTCAGCACAGGAATTGAGGTGGGTGAGGTGAAGGAAGAGTGTATG
T
TCCTATTTTTATTCCAGCCTTTTAAATTTAAAGCTTATTTTCTTGCCCTCTCCCTAACGGGGAGAAA
T
CGAGCCACCCAACCTGGAATCAGAGGGTCTGGCTGGGGTGGATGTTGCTTCTCCTGGTTCTGCCCCA
C
CACAAAGTTTTCTGTGGCAAGTGTGGAACATAGTTGTAGGCTGAGGCTCCTGCCCTTCGGTCGAGT
G
GAGCAAGCTCTCGAGGGCAGCACTGACAAATGTGTTCTAAGAAGACATTCAGACCCAGGTCTTATG
C
AGGATTACATCCGTTTATTATCAAGGGCAACCTTGGTGAAAGCAGAAAGGGTGTGTGCTATTGCATA
T
ATATGGGGGAAAAGGCAATATATTTTTTCACTGAAGCTGAGCAACCACATATTGCTACAAGGCAAATC
A
AGAAGACATCAGGAAATCAGATGCACAGGAAATAAAGGAAAGCTGTGCTTTGTCATTGAATCCTAAG
T
TCTTAGCTGCTGATGCAAGTTGTCCCCCAAGGCCATCACAAGCAGTGGGGCATGAGCTGTGTTTCA
G
GGGCCACTAAATAACAGCTGGTACTGACCCAGAAACCGCCTTCATCTCCATTCGGAAGCAGGTGAC
A
CACCCCTTCAGAAGGTGCCCTGGGTGCGGAGTGTGAGAATATACTCAGGACTCCAGAGGTGTCA
C
GTGGAAGTACAGGAGACCCGCCACCGTGGAGGCAGGGGGCAAGAACTCAAGAACGCATCAAGAGC
A
CCAGCCCTGGGCCAGGGAAGACAGGCTCTTCTGCGAGTTTCTCGTGGACACTGCTGGCTTGCGGGCA
G
TCGGTCTCCAGGGTACCTGTTGTCTTTTCCGATGTAATAACTACTTTGACCTTACACTATATGTT

[illegible]

CG206886-01

SEO ID NO: 2

926 aa

MW at 103914.2kD

Protein Sequence

VMVADGPRHLQRGPVRVGFYDIEGTLGKGNFAVVKLGRHRTKTEVAIKIIDKSLDAVNLEKIYRE
 V
 QIMKMLDHPHIKLYQVMETKSMLYLVTEYAKNGEIFDYLANHGRLNESEARRKFWQILSAVDYCHG
 R
 KIVHRDLKAENLLLDNNMNIKIADFGFGNFFKSGELLATWCGSPPYAAPEVFEGQQYEGPQLDIWSM
 G
 VVLYVLVCGALPFDGPTLPILRQRVLEGRFRIPYFMSCEDCEHLIRRMVLVDPSKRLTIAQIKEHKWM
 L
 IEVPVQRPVLYPQEQENEPSIGEFNEQVLRMLHSLGIDQQKTIESLQNKSYNHFAAIYFLLVERLKS
 H
 RSSFPVEQRLDGRQRRPSTIAEQTVAKAQTVGLPVTMHSPNMRLLRSALLPQASNVEAFSFPASGCQ
 A
 EAAFMEEEECVDTPKVNGLCLDPVPPVLVRKGCQSLPSNMMETSIDEGLETEGEAEEDPAHAFAEAFQS
 T
 RSGQRRHTLSEVTNQLVVMPPGAGKIFSMNDSPSLDSVDSEYDMGSVQRDLNFLEDNPSLKDIMLANQ
 P
 SPRMTSPFISLRPTNPAMQALSSQKREVHNRSVPVSFREGRRASDTSLTQGIVAFRQHLQNLARTKGI
 L
 ELNKVQLLYEQIGPEADPNLAPAAPQLQDLASSCPQEEVSQQQESVSTLPASVHPQLSPRQSLETQY
 L
 QHRLQKPSLLSKAONTCOLYCKEPPRSLEOOLOEHRLOOKRFLQKOSQLOQAYFNOMQIAESSYPQP

S QQLPLPRQETPPPSQQAPPFSLTQPLSPVLEPSSEMQOYSPFLSQYQEMQLQPLPSTSGPRAAPPLP T QLQQQQPPPPPPPPPPRQPGAAPAPLQFSYQTCELPASAASPAPDYPTPCQYPVDGAQQSDLTGPDGP R SPGLQEAPSSYDPLALSELPGFLDCEMLDAVDPQHNGYVLVN							
CG206886-02		SEQ ID NO: 3		990 bp			
DNA Sequence		ORF Start: ATG at 5		ORF Stop: at 989			
TACCATGGTCATGGCGGATGGCCCGAGGCACTTGCAGCGCGGGCCGGTCCGGGTGGGGTTCTACGAC A TCGAGGGCAGCTGGGCAAGGGCAACTTCGCTGTGGTGAAGCTGGGGCGGCACCGGATCACCAAGAC G GAGGTGGCAATAAAAAATAATCGATAAGTCTCAGCTGGATGCAGTGAACCTTGAGAAAATCTACCGAG A AGTACAAATAATGAAAATGTTAGACCACCCTCACATAATCAAACCTTTATCAGGTAATGGAGACCAAA A GTATGTTGTACCTTGTGACAGAATATGCCAAAAATGGAGAAATTTTGGACTATCTTGCTAATCATGG C CGGTTAAATGAGTCTGAAGCCAGGCGAAAATCTGGCAAATCCTGTCTGCTGTTGATTATTGTCATG G TCGGAAGATTGTGCACCGTGACCTCAAAGCTGAAAATCTCCTGCTGGATAACAACATGAATATCAAA A TAGCAGATTTTCGGTTTTGGAAATTTCTTTAAAAGTGGTGAAGTCTGGCAACATGGTGTGGCAGCCC C CCTTATGCAGCCCCAGAAGTCTTTGAAGGGCAGCAGTATGAAGGACCACAGCTGGACATCTGGAGTA T GGGAGTTGTTCTTTATGTCTTGTCTGTGGAGCTCTGCCCTTTGATGGACCGACTCTTCCAATTTTG A GGCAGAGGGTTCTGGAAGGAAGATTCCGGATTCCGTATTTTCATGTCAGAAGATTGCGAGCACCTTAT C CGAAGGATGTTGGTCCTAGACCCATCCAAACGGCTAACCATAGCCCAAATCAAGGAGCATAAATGGA T GCTCATAGAAGTTCTGTCCAGAGACCTGTTCTCTATCCACAAGAGCAAGAAAATGAGCCATCCATC G GGGAGTTTAATGAGCAGGTTCTGCGACTGATGCACAGCCTTGAATAGATCAGCAGAAAACCATGA G TCTTTGCAGAACAAGAGCCATCATCACCACCATCACTA							
CG206886-02		SEQ ID NO: 4		328 aa		MW at 37991.7kD	
Protein Sequence							
MVMADGPRHLQRGPVRVGFYDIEGTLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDAVNLEKIYRE V QIMKMLDHPHIKLYQVMETKSMLYLVTEYAKNGEIFDYLANHGRNLNESEARRKFWQILSAVDYCHG R KIVHRDLKAENLLLDNNMNIKIADFGFGNFFKSGELLATWCGSPPYAAPEVFEGQQYEGPQLDIWSM G VVLYVLVCGALPFDGPTLPILRQRVLEGRFRIPIYFMSEDCEHLIRMLVLDPSKRLTIAQIKEHKWM L IEVPVQRPVLYPQEQENEPSIGEFNEQVLRMLHSLGIDQOKTIESLQNKSHHHHHH							
CG206886-03		SEQ ID NO: 5		2785 bp			
DNA Sequence		ORF Start: ATG at 5		ORF Stop: TAG at 2783			
TACCATGGTCATGGCGGATGGCCCGAGGCACTTGCAGCGCGGGCCGGTCCGGGTGGGGTTCTACGAC A							

TCGAGGGCACGCTGGGCAAGGGCAACTTCGCTGTGGTGAAGCTGGGGCGGCACCGGATCACCAAGAC
G
GAGGTGGCAATAAAAAATAATCGATAAGTCTCAGCTGGATGCAGTGAACCTTGAGAAAATCTACCGAG
A
AGTACAAATAATGAAAATGTTAGACCACCCTCACATAATCAAACCTTATCAGGTAATGGAGACAAA
A
GTATGTTGTACCTTGTGACAGAATATGCCAAAAATGGAGAAATTTTGTACTATCTTGCTAATCATGG
C
CGGTTAAATGAGTCTGAAGCCAGGCGAAAATTCCTGGCAAATCCTGTCTGCTGTTGATTATTGTCATG
G
TCGGAAGATTGTGCACCGTGACCTCAAAGCTGAAAATCTCCTGCTGGATAACAACATGAATATCAAA
A
TAGCAGATTTTCGGTTTTGGAAATTTCTTTAAAAGTGGTGAAGTCTGGCAACATGGTGTGGCAGCCC
C
CCTTATGCAGCCCCAGAAGTCTTTGAAGGGCAGCAGTATGAAGGACCACAGCTGGACATCTGGAGTA
T
GGGAGTTGTTCTTTATGTCTTGTCTGTGGAGCTCTGCCCTTGTATGGACCGACTCTTCCAATTTTG
A
GGCAGAGGGTTCTGGAAGGAAGATTCCGGATTCCGTATTTTCATGTCAGAAGATTGCGAGCACCTTAT
C
CGAAGGATGTTGGTCTTAGACCCATCCAAACGGCTAACCATAGCCCAAATCAAGGAGCATAAATGGA
T
GCTCATAGAAGTTCTGTCCAGAGACCTGTTCTCTATCCACAAGAGCAAGAAAATGAGCCATCCATC
G
GGGAGTTTAAATGAGCAGGTTCTGCGACTGATGCACAGCCTTGGAAATAGATCAGCAGAAAACCATTTGA
G
TCTTTGCAGAACAAGAGCTATAACCACTTTGCTGCCATTTATTTCTTGTGTTGGTGGAGCGCCTGAAAT
C
ACATCGGAGCAGTTTCCCAGTGGAGCAGAGACTTGATGGCCGCCAGCGTCGGCCTAGCACCATTTGCT
G
AGCAAACAGTTGCCAAGGCACAGACTGTGGGGCTCCAGTGACCATGCATTACCGAACATGAGGCT
G
CTGCGATCTGCCCTCCTCCCCAGGCATCCAACGTGGAGGCCTTTTCATTTCCAGCATCTGGCTGTC
A
GGCGGAAGCTGCATTTCATGGAAGAAGAGTGTGTGGACACTCCAAAGGTCAATGGCTGTCTGCTTGAC
C
CTGTGCCTCCTGTCTGCTGGTGCAGGAAGGGATGCCAGTCACTGCCAGCAACATGATGGAGACCTCCAT
T
GACGAAGGGCTGGAGACAGAAGGAGAGGCCGAGGAAGACCCCGCTCATGCCTTTGAGGCATTTTCAGT
C
CACACGCAGCGGGCAGAGACGGCACACTCTGTGAGAAGTGACCAATCAACTGGTCGTGATGCCTGGG
G
CAGGGAAAATTTTCTCCATGAATGACAGCCCCCTCCCTTGACAGTGTGGACTCTGAGTATGATATGGG
G
TCTGTTTCAGAGGGACCTGAACCTTTCTGGAAGACAACCCTTCCCTTAAGGACATCATGTTAGCCAATC
A
GCCTTACCCCGCATGACATCTCCCTTCATAAGCCTGAGACCTACCAACCCAGCCATGCAGGCTCTG
A
GCTCCCAGAAACGAGAGGTCCACAACAGGTCTCCAGTGAGCTTCAGAGAGGGCCGCAGAGCATCAGA
T
ACCTCCCTCACCCAGGGAATTGTAGCATTTAGACAACATCTTCAGAATCTGGCTAGAACCAAAGGAA
T
TCTAGAGTTGAACAAAGTGCAGTTGTTGTATGAACAAATAGGACCGGAGGCAGACCCTAACCTGGCG
C
CGGCGGCTCCTCAGCTCCAGGACCTTGCTAGCAGCTGCCCTCAGGAAGAAGTTTCTCAGCAGCAGGA
A

AGCGTCTCCACTCTCCCTGCCAGCGTGCATCCCCAGCTGTCCCCACGGCAGAGCCTGGAGACCCAGT
 A
 CCTGCAGCACAGACTCCAGAAGCCCAGCCTTCTGTCAAAGGCCCAGAACACCTGTCAGCTTTATTGC
 A
 AGAACCACCGCGGAGCCTTGAGCAGCAGCTGCAGGAACATAGGCTCCAGCAGAAGCGACTCTTTCT
 T
 CAGAAGCAGTCTCAACTGCAGGCCTATTTTAATCAGATGCAGATAGCAGAGAGCTCCTACCCACAGC
 C
 AAGTCAGCAGCTGCCCCCTCCCCGCCAGGAGACTCCACCGCCTTCTCAGCAGGCCCCACCGTTCAGC
 C
 TGACCCAGCCCCTGAGCCCCGTCCTGGAGCCTTCTCCGAGCAGATGCAATACAGCCCTTTCCTCAG
 C
 CAGTACCAAGAGATGCAGCTTCAGCCCCCTGCCCTCCACTTCCGGTCCCCGGGCTGCTCCTCCTCTGC
 C
 CACGCAGCTACAGCAGCAGCAGCCGCCACCGCCACCACCCCTCCACCACCACGACAGCCAGGAGCT
 G
 CCCCAGCCCCCTTACAGTTCTCCTATCAGACTTGTGAGCTGCCAAGCGCTGCTTCCCCTGCGCCAGA
 C
 TATCCCACTCCCTGTCAGTATCCTGTGGATGGAGCCCAGCAGAGCGACCTAACGGGGCCAGACTGTC
 C
 CAGAAGCCCAGGACTGCAAGAGGCCCCCTCCAGCTACGACCCACTAGCCCTCTCTGAGCTACCTGGA
 C
 TCTTTGATTGTGAAATGCTAGACGCTGTGGATCCACAACACAACGGGTATGTCCTGGTGAAT**TAG**

CG206886-03	SEQ ID NO: 6	926 aa	MW at 103914.2kD
-------------	--------------	--------	------------------

Protein Sequence

MVMADGPRHLQRGPRVVGFDIEGTLGKGNFAVVKLGRHRTKTEVAIKIIDKSQLDVNLKITYRE
 V
 QIMKMLDHPHIIKLYQVMETKSMYLYVTEYAKNGEIFDYLANHGRNLNESEARRKFWQILSAVDYCHG
 R
 KIVHRDLKAENLLLDNNMNIKIADFGFGNFFKSGELLATWCGSPPYAAPEVFEQQYEGPQLDIWSM
 G
 VVLYVLVCGALPFDGPTLPILRQRVLEGRFRIPYFMSEDCEHLIRMLVLDPSKRLTIAQIKEHKWM
 L
 IEVPVQRPVLYPQEQENEPSIGEFNEQVLRMLHSLGIDQQKTIESLQNKSYNHFAAIYFLLVERLKS
 H
 RSSFPVEQRLDGRQRRPSTIAEQTVAKAQTVGLPVTMHSNMRLLRSALLPQASNVEAFSFPASGCQ
 A
 EAAFMEEECVDPKVNGLCLDPVPPVLVRKGCQSLPSNMMETSIDEGLETEGEAEEDPAHAFAEFQS
 T
 RSGQRRHTLSEVTNQLVVMGPAGKIFSMNDSPLSDVDSEYDMGSVQRDLNFLEDNPSLKDIMLANQ
 P
 SPRMTSPFISLRPTNPAMQALSSQKREVHNRSVPVSFREGRRASDTSLTQGIVAFRQHLQNLARTKGI
 L
 ELNKVQLLYEQIGPEADPNLAPAAPQLQDLASSCPQEEVSQQQESVSTLPASVHPQLSPRQSLETQY
 L
 QHRLQKPSLLSKAQNTCQLYCKEPPRSLEQQLQEHRLQOKRLFLQKQSQLQAYFNQMQUIAESSYPQP
 S
 QQLPLPRQETPPPSQAPPFSLTQPLSPVLEPSSEQMQYSPFLSQYQEMQLQPLPSTSGPRAAPLP
 T
 QLQQQQPPPPPPPPPPRQPGAAPAPLQFSYQTCLEPSAASPAPDYPTPCQYPVDGAQQSDLTGPDGP
 R
 SPGLQEAPSSYDPLALSELPGLFDCEMLDVDPQHNGYVLVN

CG206886-04	SEQ ID NO: 7	2806 bp
DNA Sequence	ORF Start: ATG at 5	ORF Stop: TAG at 2804

TACCATGGGACATCATCACCACCATCACGTCATGGCGGATGGCCCGAGGCACTTGCAGCGCGGGCCG
G
TCCGGGTGGGGTTCTACGACATCGAGGGCACGCTGGGCAAGGGCAACTTCGCTGTGGTGAAGCTGGG
G
CGGCACCGGATCACCAAGACGGAGGTGGCAATAAAAAATAATCGATAAGTCTCAGCTGGATGCAGTGA
A
CCTTGAGAAAATCTACCGAGAAGTACAAATAATGAAAATGTTAGACCACCCTCACATAATCAAACCTT
T
ATCAGGTAATGGAGACCAAAGTATGTTGTACCTTGTGACAGAATATGCCAAAAATGGAGAAATTTT
T
GACTATCTTGCTAATCATGGCCGGTTAAATGAGTCTGAAGCCAGGCGAAAATTCTGGCAAATCCTGT
C
TGCTGTTGATTATTGTCATGGTCGGAAGATTGTGCACCGTGACCTCAAAGCTGAAAATCTCCTGCTG
G
ATAACAACATGAATATCAAATAGCAGATTCGGTTTTGGAAATTTCTTTAAAAGTGGTGAAGTGGT
G
GCAACATGGTGTGGCAGCCCCCTTATGCAGCCCCAGAAGTCTTTGAAGGGCAGCAGTATGAAGGAC
C
ACAGCTGGACATCTGGAGTATGGGAGTTGTTCTTTATGTCCTTGTCTGTGGAGCTCTGCCCTTTGAT
G
GACCGACTCTTCCAATTTTGAGGCAGAGGGTTCTGGAAGGAAGATTCCGGATTCCGTATTTTCATGTC
A
GAAGATTGCGAGCACCTTATCCGAAGGATGTTGGTCCTAGACCCATCCAAACGGCTAACCATAGCCC
A
AATCAAGGAGCATAAATGGATGCTCATAGAAGTTCCTGTCCAGAGACCTGTTCTCTATCCACAAGAG
C
AAGAAAATGAGCCATCCATCGGGGAGTTTAATGAGCAGGTTCTGCGACTGATGCACAGCCTTGGAAT
A
GATCAGCAGAAAACCATTGAGTCTTTGCAGAACAAGAGCTATAACCACTTTGCTGCCATTTATTTCT
T
GTTGGTGGAGCGCCTGAAATCACATCGGAGCAGTTTCCAGTGGAGCAGAGACTTGATGGCCGCCAG
C
GTCGGCCTAGCACCATTGCTGAGCAAACAGTTGCCAAGGCACAGACTGTGGGGCTCCAGTGACCAT
G
CATTCACCGAACATGAGGCTGCTGCGATCTGCCCTCCTCCCCCAGGCATCCAACGTGGAGGCCTTTT
C
ATTTCCAGCATCTGGCTGTCAGGCGGAAGCTGCATTCATGGAAGAAGAGTGTGTGGACACTCCAAAG
G
TCAATGGCTGTCTGCTTGACCCTGTGCCTCCTGTCTGGTGCAGGAAGGGATGCCAGTCACTGCCCAG
C
AACATGATGGAGACCTCCATTGACGAAGGGCTGGAGACAGAAGGAGAGGCCGAGGAAGACCCCGCTC
A
TGCCCTTTGAGGCATTTTCAGTCCACACGCAGCGGGCAGAGACGGCACACTCTGTCAGAAGTGACCAAT
C
AACTGGTCGTGATGCCTGGGGCAGGGAAAATTTTCTCCATGAATGACAGCCCCCTCCCTTGACAGTGT
G
GACTCTGAGTATGATATGGGGTCTGTTTCAGAGGGACCTGAACCTTTCTGGAAGACAACCTTCCCTTA
A
GGACATCATGTTAGCCAATCAGCCTTCACCCCGCATGACATCTCCCTTCATAAGCCTGAGACCTACC
A
ACCCAGCCATGCAGGCTCTGAGCTCCAGAAACGAGAGGTCCACAACAGGTCTCCAGTGAGCTTCAG
A
GAGGGCCGCAGAGCATCAGATACCTCCCTCACCCAGGGAATTGTAGCATTTAGACAACATCTTCAGA
A
TCTGGCTAGAACCAAAGGAATTCTAGAGTTGAACAAAGTGCAGTTGTTGTATGAACAAATAGGACCG
G.

AGGCAGACCCTAACCTGGCGCCGGCGGCTCCTCAGCTCCAGGACCTTGCTAGCAGCTGCCCTCAGGA
 A
 GAAGTTTCTCAGCAGCAGGAAAGCGTCTCCACTCTCCCTGCCAGCGTGCATCCCCAGCTGTCCCCAC
 G
 GCAGAGCCTGGAGACCCAGTACCTGCAGCACAGACTCCAGAAGCCCAGCCTTCTGTCAAAGGCCACG
 A
 ACACCTGTCAGCTTTATTGCAAAGAACCACCGGAGCCTTGAGCAGCAGCTGCAGGAACATAGGCT
 C
 CAGCAGAAGCGACTCTTTCTTCAGAAGCAGTCTCAACTGCAGGCCTATTTTAATCAGATGCAGATAG
 C
 AGAGAGCTCCTACCCACAGCCAAGTCAGCAGCTGCCCCCTTCCCCGCCAGGAGACTCCACCGCCTTCT
 C
 AGCAGGCCCCACCGTTTCAGCCTGACCCAGCCCCCTGAGCCCCGTCTGGAGCCTTCTCCGAGCAGAT
 G
 CAATACAGCCCTTTCCTCAGCCAGTACCAAGAGATGCAGCTTCAGCCCCCTGCCCTCCACTTCCGGTC
 C
 CCGGGCTGCTCCTCCTCTGCCACGCAGCTACAGCAGCAGCAGCGCCACCGCCACCACCCCTCCA
 C
 CACCACGACAGCCAGGAGCTGCCCCAGCCCCCTTACAGTTCTCCTATCAGACTTGTGAGCTGCCAAG
 C
 GCTGCTTCCCCTGCGCCAGACTATCCCACTCCCTGTCAGTATCCTGTGGATGGAGCCCAGCAGAGCG
 A
 CCTAACGGGGCCAGACTGTCCAGAAGCCCAGGACTGCAAGAGGCCCCCTCCAGCTACGACCCACTA
 G
 CCCTCTCTGAGCTACCTGGACTCTTTGATTGTGAAATGCTAGACGCTGTGGATCCACAACACAACGG
 G
 TATGTCCTGGTGAAT**TAG**

CG206886-04

SEQ ID NO: 8

933 aa

MW at 104794.1kD

Protein Sequence

MGHHHHHHVMADGPRHLQRGFVRVGFYDIEGTLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDVAVN
 L
 EKIYREVQIMKMLDHPHIIKLYQVMETKSMYLVTEYAKNGEIFDYLANHGRNLNESEARRKFWQILS
 A
 VDYCHGRKIVHRDLKAENLLLDNNMNKIADFGFGNFFKSGELLATWCGSPPYAAPEVFEGQQYEGP
 Q
 LDIWSMGVVLYVLVCGALPFDGPTLPILRQRVLEGRFRIPIFYMSDCEHLIRMLVLDPSKRLTIAQ
 I
 KEHKWMLIEVPVQRPVLYPQEQENEPSIGEFNEQVLRMLHSLGIDQOKTIESLQNKSYNHFAAIYFL
 L
 VERLKSHRSSFPVEQRLDGRQRRPSTIAEQTVAKAQTVGLPVTMHSPNMRLRLSALLPQASNVEAFS
 F
 PASGCQAEAAFMEEECVDTPKVNGCLLDVPPVLRKGCQSLPSNMMETSIDEGLETEGEAEEDPAH
 A
 FEAFQSTRSGQRRHTLSEVTNQLVVMGAGKIFSMNDSPSLDSVDSEYDMGSVQORDLNFLDNPSLK
 D
 IMLANQPSPRMTSPFISLRPTNPAMQALSSQKREVNHRSPVSFREGRRASDTSLTQGIVAFRQHLQN
 L
 ARTKGILELNKVQLLYEQIGPEADPNLAPAAPQLQDLASSCPQEEVSQQQESVSTLPASVHPQLSPR
 Q
 SLETQYLQHRLQKPSLLSKAQNTCQLYCKEPPRSLEQQQLQEHRLQOKRLFLOKQSQLOAYFNQMOTA
 E
 SSYPQPSQQLPLPRQETPPPSQAPPFSLTQPLSPVLEPSSEMQQYSPFLSQYQEMQLQPLPSTSGP
 R
 AAPPLPTQLQQQPPPPPPPPPPRQPGAAPAPLQFSYQTCELPSAASPAPDYPTPCQYPVDGAQQSD
 L

TGPD CPRSPGLQEAPSSYDPLALSELPGLEFDCEMLDAVDPOHNGYVLVN

Example 3. Quantitative expression analysis of SIK2 on Human Metabolic Panel

The quantitative expression of the human SIK2 gene 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) 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 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, Carlsbad, CA, Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10 μ g of 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 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 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 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 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 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.

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 and/or diabetes. Metabolic tissues including placenta (PI), 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 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.

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 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.

Expression of gene CG206886-01 was assessed using the primer-probe set Ag1887, described in Table 5. Results of the RTQ-PCR runs are shown in Table 6.

Table 5.

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gttatttagtgccctgaaac-3'	22	3462	14
Probe	TET-5'-agctcatgccccactgcttggat-3'-TAMRA	25	3435	15
Reverse	5'-gttcttagctgctgatgcaagt-3'	22	3399	16

5

10

Results

Table 6. Human Metabolic Panel

Column A - Rel. Exp.(%) Ag1887, Run 328115082			
Tissue Name	A	Tissue Name	A
137857 psoas-AA.M.Diab.-hi BMI-6	39.0	139523 pancreas-HI.M.Norm-hi BMI-31	6.6
135760 psoas-HI.M.Diab.-hi BMI-21	2.7	139520 pancreas-CC.M.Norm-hi BMI-29	9.9
134827 psoas-CC.M.Diab.-hi BMI-4	4.2	142744 pancreas-HI.M.Norm-med BMI-35	1.4
137860 psoas-AA.M.Diab.-med BMI-8	14.1	139545 pancreas-AA.M.Norm-med BMI-47	14.7
137834 psoas-CC.M.Diab.-med BMI-2	18.7	139531 pancreas-AA.M.Norm-med BMI-37	3.4
137828 psoas-CC.M.Diab.-med BMI-1	42.3	137871 pancreas-CC.M.Norm-med BMI-26	2.1
135763 psoas-HI.M.Diab.-med BMI-23	26.8	139541 pancreas-HI.M.Norm-low BMI-41	11.0
142740 psoas-AS.M.Diab.-low BMI-20	21.3	139537 pancreas-CC.M.Norm-low BMI-40	18.2
134834 psoas-AA.M.Diab.-low BMI-17	11.6	139533 pancreas-CC.M.Norm-low BMI-39	1.0
137850 psoas-AS.M.Norm-hi BMI-34	20.4	137845 pancreas-AS.M.Norm-low BMI-28	1.1
135769 psoas-HI.M.Norm-hi BMI-31	22.4	143530 small intestine-AA.M.Diab.-hi BMI-6	4.9
135766 psoas-AA.M.Norm-hi BMI-25	5.6	143529 small intestine-CC.M.Diab.-hi BMI-4	20.6
142746 psoas-AA.M.Norm-med	4.8	143538 small intestine-HI.M.Diab.-med BMI-23	9.8

BMI-37			
142745 psoas-HI.M.Norm-med BMI-35	7.0	143531 small intestine-AA.M.Diab.-med BMI-8	28.9
137855 psoas-AA.M.Norm-med BMI-47	7.4	143528 small intestine-CC.M.Diab.-med BMI-2	9.9
137844 psoas-CC.M.Norm-med BMI-26	8.2	143537 small intestine-HI.M.Diab.-low BMI-22	16.4
142742 psoas-CC.M.Norm-low BMI-40	6.9	143535 small intestine-AS.M.Diab.-low BMI-20	5.4
137873 psoas-AS.M.Norm-low BMI-28	5.6	143534 small intestine-AA.M.Diab.-low BMI-17	0.6
137853 psoas-HI.M.Norm-low BMI-41	27.5	143544 small intestine-AS.M.Norm-hi BMI-34	11.0
135775 psoas-CC.M.Norm-low BMI-39	3.3	143543 small intestine-HI.M.Norm-hi BMI-31	3.3
137858 diaphragm-AA.M.Diab.-hi BMI-6	40.6	143542 small intestine-CC.M.Norm-hi BMI-29	4.6
135772 diaphragm-AS.M.Diab-hi BMI-9	22.5	143539 small intestine-AA.M.Norm-hi BMI-25	5.5
135761 diaphragm-HI.M.Diab.-hi BMI-21	12.4	143548 small intestine-AA.M.Norm-med BMI-47	5.5
134828 diaphragm-CC.M.Diab.-hi BMI-4	4.4	143547 small intestine-AA.M.Norm-med BMI-37	7.5
137835 diaphragm-CC.M.Diab.-med BMI-2	11.0	143540 small intestine-CC.M.Norm-med BMI-26	1.4
135764 diaphragm-HI.M.Diab.-med BMI-23	11.9	143550 small intestine-CC.M.Norm-low BMI-40	0.9
134835 diaphragm-AA.M.Diab.-low BMI-17	3.8	143549 small intestine-CC.M.Norm-low BMI-39	11.3
142738 diaphragm-CC.M.Norm-hi BMI-29	6.7	143546 small intestine-HI.M.Norm-low BMI-41	3.1
139517 diaphragm-AS.M.Norm-hi BMI-34	23.5	143525 hypothalamus-HI.M.Diab.-hi BMI-21	12.1
137848 diaphragm-HI.M.Norm-hi BMI-31	18.0	143515 hypothalamus-CC.M.Diab.-hi BMI-4	1.4
137843 diaphragm-AA.M.Norm-hi BMI-25	18.3	143513 hypothalamus-AA.M.Diab.-hi BMI-6	15.0
137879 diaphragm-AA.M.Norm-med BMI-47	8.4	143507 hypothalamus-AS.M.Diab.-hi BMI-9	16.8
137872 diaphragm-CC.M.Norm-med BMI-26	6.6	143506 hypothalamus-CC.M.Diab.-med BMI-1	29.3
135773 diaphragm-HI.M.Norm-med BMI-35	34.6	143505 hypothalamus-HI.M.Diab.-med BMI-23	19.6
139542 diaphragm-HI.M.Norm-low BMI-41	20.6	143509 hypothalamus-AA.M.Diab.-low BMI-17	34.9
137877 diaphragm-CC.M.Norm-low BMI-39	4.3	143508 hypothalamus-CC.M.Diab.-low BMI-13	13.1
137874 diaphragm-AS.M.Norm-low BMI-28	2.3	143503 hypothalamus-AS.M.Diab.-low BMI-20	17.8

141340 subQadipose-AA.M.Diab.-hi BMI-6	7.6	143522 hypothalamus-HI.M.Norm-hi BMI-31	1.5
137836 subQadipose-HI.M.Diab.-hi BMI-21	2.5	143516 hypothalamus-AS.M.Norm-hi BMI-34	1.1
135771 subQadipose-AS.M.Diab.-hi BMI-9	6.8	143511 hypothalamus-CC.M.Norm-hi BMI-29	23.0
141329 subQadipose-AA.M.Diab-medbmi-8	4.0	143504 hypothalamus-AA.M.Norm-hi BMI-25	11.0
137862 subQadipose-CC.M.Diab.-med BMI-1	9.2	143517 hypothalamus-AA.M.Norm-med BMI-47	15.6
135762 subQadipose-HI.M.Diab.-med BMI-23	3.4	143514 hypothalamus-HI.M.Norm-med BMI-35	4.5
141338 subQadipose-AS.M.Diab.-low BMI-20	2.3	143521 hypothalamus-AS.M.Norm-low BMI-28	9.8
139547 subQadipose-HI.M.Diab.-low BMI-22	0.9	143512 hypothalamus-CC.M.Norm-low BMI-40	4.5
135757 subQadipose-CC.M.Diab.-low BMI-13	4.7	145454 Patient-25pl (CC.Diab.low BMI.no insulin)	10.9
134832 subQadipose-AA.M.Diab.-low BMI-17	3.6	110916 Patient-18pl (HI.Diab.obese.no insulin)	6.0
141332 subQadipose-HI.M.Norm-hi BMI-31	1.6	110913 Patient-18go (HI.Diab.obese.no insulin)	12.5
135767 subQadipose-CC.M.Norm-hi BMI-29	3.6	110911 Patient-17pl (CC.Diab.low BMI.no insulin)	9.7
135765 subQadipose-AS.M.Norm-hi BMI-34	2.6	110908 Patient-17go (CC.Diab.low BMI.no insulin)	40.9
141339 subQadipose-HI.M.Norm-med BMI-35	1.9	100752 Patient-15sk (CC.Diab.obese.no insulin)	19.1
141334 subQadipose-CC.M.Norm-med BMI-26	1.3	97828 Patient-13pl (CC.Diab.overwt.no insulin)	7.1
139544 subQadipose-AA.M.Norm-med BMI-47	30.4	160114 Patient 27-ut (CC.Diab.obese.insulin)	14.7
137875 subQadipose-AA.M.Norm-med BMI-37	7.1	160113 Patient 27-pl (CC.Diab.obese.insulin)	19.9
141331 subQadipose-AS.M.Norm-low BMI-28	4.2	160112 Patient 27-sk (CC.Diab.obese.insulin)	6.6
137878 subQadipose-HI.M.Norm-low BMI-41	5.2	160111 Patient 27-go (CC.Diab.obese.insulin)	55.5
137876 subQadipose-CC.M.Norm-low BMI-39	3.6	145461 Patient-26sk (CC.Diab.obese.insulin)	5.9
137859 vis.adipose-AA.M.Diab.-hi BMI-6	0.4	145441 Patient-22sk (CC.Diab.low BMI.insulin)	13.6
135770 vis.adipose-AS.M.Diab.-hi BMI-9	13.0	145438 Patient-22pl (CC.Diab.low BMI.insulin)	9.9
135759 vis.adipose-HI.M.Diab.-hi BMI-21	2.7	145427 Patient-20pl (CC.Diab.overwt.insulin)	27.2
143502 vis.adipose-CC.M.Diab.-med BMI-2	23.7	97503 Patient-12pl (CC.Diab.unknown BMI.insulin)	1.4
139510 vis.adipose-AA.M.Diab.-	0.6	145443 Patient-23pl (CC.Non-diab.overwt)	15.4

med BMI-8			
137861 vis.adipose-CC.M.Diab.-med-1	4.1	145435 Patient-21pl (CC.Non-diab.overwt)	19.5
137839 vis.adipose-HI.M.Diab.-med BMI-23	3.1	110921 Patient-19pl (CC.Non-diab.low BMI)	24.1
139546 vis.adipose-HI.M.Diab.-low BMI-22	5.2	110918 Patient-19go (CC.Non-diab.low BMI)	38.4
137831 vis.adipose-CC.M.Diab.-low BMI-13	2.3	97481 Patient-08sk (CC.Non-diab.obese)	15.2
139522 vis.adipose-HI.M.Norm-hi BMI-31	10.0	97478 Patient-07pl (CC.Non-diab.obese)	7.0
139516 vis.adipose-AS.M.Norm-hi BMI-34	4.3	160117 Human Islets-male, obese	17.9
137846 vis.adipose-CC.M.Norm-hi BMI-29	3.1	145474 PANC1 (pancreas carcinoma) 1	37.1
137841 vis.adipose-AA.M.Norm-hi BMI-25	0.8	154911 Capan2 (pancreas adenocarcinoma)	74.2
139543 vis.adipose-AA.M.Norm-med BMI-47	19.1	141190 SW579 (thyroid carcinoma)	23.0
139532 vis.adipose-AA.M.Norm-med BMI-37	1.2	145489 SK-N-MC (neuroblastoma) 1	31.2
139530 vis.adipose-HI.M.Norm-med BMI-35	0.3	145495 SK-N-SH (neuroblastoma) 1	40.1
139539 vis.adipose-HI.M.Norm-low BMI-41	4.0	145498 U87 MG (glioblastoma) 2	13.7
139535 vis.adipose-CC.M.Norm-low BMI-40	6.0	145484 HEp-2 (larynx carcinoma) 1	81.8
137852 vis.adipose-CC.M.Norm-low BMI-39	3.3	145479 A549 (lung carcinoma)	16.4
135768 vis.adipose-AS.M.Norm-low BMI-28	1.5	145488 A427 (lung carcinoma) 2	7.4
141327 liver-CC.M.Diab.-hi BMI-4	19.5	145472 FHs 738Lu (normal lung) 1	44.4
139514 liver-HI.M.Diab.-hi BMI-21	75.8	141187 SKW6.4 (B lymphocytes)	13.3
139526 liver-CC.M.Diab.-med BMI-2	69.3	154644 IM-9 (immunoglobulin secreting lymphoblast)	28.3
139511 liver-AA.M.Diab.-med BMI-8	100.0	154645 MOLT-4 (acute lymphoblastic leukemia derived from peripheral blood)	42.6
137840 liver-HI.M.Diab.-med BMI-23	2.3	154648 U-937 (histiocytic lymphoma)	36.3
137827 liver-CC.M.Diab.-med BMI-1	5.4	154647 Daudi (Burkitt's lymphoma)	8.8
137838 liver-HI.M.Diab.-low BMI-22	6.3	145494 SK-MEL-2 (melanoma) 2	39.2
135758 liver-CC.M.Diab.-low BMI-13	2.9	141176 A375 (melanoma)	15.1
139519 liver-CC.M.Norm-hi BMI-29	79.0	154642 SW 1353 (humerus chondrosarcoma)	47.6
139518 liver-AA.M.Norm-hi BMI-25	42.9	141179 HT-1080 (fibrosarcoma)	35.4

137849 liver-AS.M.Norm-hi BMI-34	4.0	145491 MG-63 (osteosarcoma) 1	34.2
137847 liver-HI.M.Norm-hi BMI-31	3.9	141186 MCF7 (breast carcinoma)	42.6
142741 liver-AA.M.Norm-med BMI-37	15.4	141193 T47D (breast carcinoma)	31.9
141341 liver-HI.M.Norm-med BMI-35	21.5	154641 BT-20 (breast carcinoma)	37.9
141335 liver-CC.M.Norm-med BMI-26	29.5	141175 293 (kidney transformed with adenovirus 5 DNA)	4.3
139540 liver-HI.M.Norm-low BMI-41	13.3	141182 HUH hepatoma 1	8.0
139534 liver-CC.M.Norm-low BMI-39	32.3	141184 HUH7 hepatoma 1	19.5
139521 liver-AS.M.Norm-low BMI-28	63.7	145478 HT1376 (bladder carcinoma)	20.2
141328 pancreas-CC.M.Diab.-hi BMI-4	10.8	145481 SCaBER (bladder carcinoma)	18.4
139525 pancreas-AS.M.Diab.-hi BMI-9	2.5	141192 SW620 (lymph node metastasis, colon carcinoma) 2	15.6
137856 pancreas-AA.M.Diab.-hi BMI-6	13.6	141180 HT29 (colon carcinoma) 1	50.3
137837 pancreas-HI.M.Diab.-hi BMI-21	1.5	141188 SW480 (colon carcinoma) 1	41.5
141337 pancreas-CC.M.Diab.-med BMI-2	2.5	154646 CAO-3 (ovary adenocarcinoma)	15.5
139527 pancreas-CC.M.Diab.-med BMI-1	5.9	141194 HeLa (cervix carcinoma)- 2	43.8
139515 pancreas-HI.M.Diab.-med BMI-23	13.9	145482 HeLa S3 (cervix carcinoma) 1	16.0
139512 pancreas-AA.M.Diab.-med BMI-8	10.8	145486 DU145 (prostate carcinoma)	11.3
142739 pancreas-AS.M.Diab.-low BMI-20	6.1	154643 PC-3 (prostate adenocarcinoma)	43.8
139513 pancreas-CC.M.Diab.-low BMI-13	5.1	154649 HCT-8 (ileocecal adenocarcinoma)	23.0
142743 pancreas-AA.M.Norm-hi BMI-25	5.7		

RTQ-PCR analysis of human SIK2 gene expression on panel of metabolically active tissues from patients with and without diabetes showed that SIK2 mRNA level is up-regulated in adipose (SubQ), skeletal muscle (psoas) and hypothalamus in diabetic patients compared to normal controls by 4 fold ($p=0.026$; T-test), 2.5 fold ($p=0.001$; T-test) and 3 fold ($p=0.044$; T-test), respectively. This finding further strengthened the hypothesis that an increased level of SIK2 contributes to insulin resistance/diabetes phenotype.

5

Example 4. Protein-protein interactions**PathCalling® Technology**

Polypeptides comprising portions of the SIK2 amino acid sequence were found to interact with other polypeptides in the two-hybrid system (Fields and Song, 1989, Nature 340:245-6). The laboratory screening was performed using the methods that follow. cDNA libraries were derived from adult brain tissue samples, and from the HeLa cell line, commercially available from Clontech (Palo Alto, CA). The libraries were then transferred from E.coli into a CuraGen Corporation proprietary yeast strain (disclosed in U.S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corporation proprietary library of human sequences was used to screen 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 manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

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

Results

Table 7 summarizes the protein domains involved in the protein-protein interaction between SIK2 and Axl and PTK6 kinases

Table 7. SIK2 protein –protein interactions detected by yeast two-hybrid system						
cDNA Library	Bait			Prey		
	gene name	clone insert codes for	domain	gene name	clone insert codes for	domain
adult human brain	AXL	amino acids 273-695 of CG59325-03	Protein kinase	SIK2	amino acids 585-743 of CG206886-01	-
	AXL	amino acids 33-251 of CG59325-03	Immuno-globulin	SIK2	amino acids 585-743 of CG206886-01	-
HeLa	PTK6	amino acids 1-451 of	Full	SIK2	amino acids 568-743 of	-

cells		CG108678-01	length		CG206886-01	
	PTK6	amino acids 1-176 of CG108678-01	SH3 + SH2	SIK2	amino acids 568-743 of CG206886-01	-

PathCalling® analysis showed that polypeptides comprising portions of the SIK2 amino acid sequence interacted with two protein tyrosine kinases – AXL Receptor Tyrosine Kinase (AXL) and Protein-Tyrosine Kinase 6 (PTK6) (Figure 2 and Table 7). PathCalling® further showed that polypeptides comprising portions of the PTK6 or the AXL amino acid sequence each interacted with polypeptides comprising portions of the VAV1 oncogene protein (GenBank Accession No. NM_005428), or the RAB14 protein (GenBank Accession No. NM_016322), or two regulatory subunits of PI3 kinase (PI3K; Phosphatidylinositol 3-kinase), PIK3R1 (GenBank Accession No. NM_181504) and PIK3R2 (GenBank Accession No. NM_005027) (Figure 2). The AXL (CG59325) and PTK6 (CG108678) nucleotide and amino acid sequences are shown in Table 8.

It is known that transgenic mice ectopically overexpressing Axl exhibit phenotypic characteristics associated with insulin resistance and diabetes including hyperglycemia and hyperinsulinemia (Augustine et al., *J Cell Physiol.* 1999 Dec;181(3):433-47). SIK2 by direct interaction with Axl may contribute to its negative effect on insulin signaling. PTK6 and AXL may attenuate insulin signaling by redistributing of PI3K away from insulin signaling and/or by activation of VAV1 followed by induction of NF-κB and Jnk signaling pathway. The redistribution of PI3 kinase as well as activation of Jnk/NF-κB pathways are implicated in development of insulin resistance, diabetes and/or obesity (Shao et al., *Diabetes* 2002 Jan;51(1):19-29; JR Burke, *Curr Opin Drug Discov Devel.* 2003 Sep;6(5):720-8).

Table 8. PTK6 and AXL sequences

CG59325-03	SEQ ID NO: 9	2196 bp
DNA Sequence	ORF Start: ATG at 28	ORF Stop: TGA at 2113
AGCAACTTCTGAGGAAAGTTTGCACCCATGGCGTGGCGGTGCCCCAGGATGGGCAGGGTCCCGCTGG C CTGGTGCTTGGCGCTGTGCGGCTGGGCGTGCATGGCCCCAGGGGCACGCAGGCTGAAGAAAGTCCC T TCGTGGGCAACCCAGGGAATATCACAGGTGCCCCGTGAGTCCCCGGGCACCCTTCGGTGTCAGCTCCA G GTTTCAGGGAGAGCCCCCGAGGTACATTGGCTTCGGGATGGACAGATCCTGGAGCTCGCGGACAGCA C CCAGACCCAGGTGCCCTGGGTGAGGATGAACAGGATGACTGGATAGTGGTCAGCCAGCTCAGAATC A CCTCCCTGCAGCTTTCGACACGGGACAGTACCAGTGTTTGGTGTCTTCTGGGACATCAGACCTTCGT G TCCAGCCTGGCTATGTGGGCTGGAGGGCTTGCCCTACTTCTGGAGGAGCCCGAAGACAGGACTG T GGCCGCCAACACCCCTTCAACCTGAGCTGCCAAGCTCAGGGACCCCCAGAGCCCGTGGACCTACTC T		

GGCTCCAGGATGCTGTCCCCCTGGCCACGGCTCCAGGTACGGCCCCCAGCGCAGCCTGCATGTTCC
 A
 GGGCTGAACAAGACATCCTCTTTCTCCTGCGAAGCCCATAACGCCAAGGGGGTCACCACATCCCGCA
 C
 AGCCACCATCACAGTGCTCCCCAGCAGCCCCGTAACCTCCACCTGGTCTCCCACCAGCTGGTGAAG
 G
 AATCTTCAACTCCTGCCTTCTCGTGGCCCTGGTGGTATGTACTGCTAGGAGCAGTCGTGGCCGCTGC
 C
 TGTGTCTCATCTTGGCTCTCTTCTTGTCCACCGGCGAAAGAAGGAGACCCGTTATGGAGAAGTGT
 T
 TGAACCAACAGTGGATAGAGGTGAAGTGGTAGTCAGGTACCGCGTGCACAAGTCCTACAGTCGTGCG
 A
 CCACTGAAGCTACCTTGAACAGCCTGGGCATCAGTGAAGAGCTGAAGGAGAAGCTGCGGGATGTGAT
 G
 GTGGACCGGCACAAGGTGGCCCTGGGGAAGACTCTGGGAGAGGGAGAGTTTGGAGCTGTGATGGAAG
 G
 CCAGCTCAACCAGGACGACTCCATCCTCAAGGTGGCTGTGAAGACGATGAAGATTGCCATCTGCACG
 A
 GGTCAGAGCTGGAGGATTTCTGAGTGAAGCGGTCTGCATGAAGGAATTTGACCATCCCAACGTCAT
 G
 AGGCTCATCGGTGTCTGTTTCCAGGGTTCTGAACGAGAGAGCTTCCCAGCACCTGTGGTCATCTTAC
 C
 TTTCATGAAACATGGAGACCTACACAGCTTCTCCTCTATTCCCGGCTCGGGGACCAGCCAGTGTAC
 C
 TGCCCACTCAGATGCTAGTGAAGTTCATGGCAGACATCGCCAGTGGCATGGAGTATCTGAGTACCAA
 G
 AGATTACATACACCGGGACCTGGCGGCCAGGAACTGCATGCTGAATGAGAACATGTCCGTGTGTGTGG
 C
 GGACTTCGGGCTCTCCAAGAAGATCTACAATGGGGACTACTACCGCCAGGGACGTATCGCCAAGATG
 C
 CAGTCAAGTGGATTGCCATTGAGAGTCTAGCTGACCGTGTCTACACCAGCAAGAGCGATGTGTGGTC
 C
 TTCGGGGTGACAATGTGGGAGATTGCCACAAGAGGCCAAACCCCATATCCGGGCGTGGAGAACAGCG
 A
 GATTTATGACTATCTGCGCCAGGGAAATCGCCTGAAGCAGCCTGCGGACTGTCTGGATGGACTGTAT
 G
 CCTTGATGTCGCGGTGCTGGGAGCTAAATCCCCAGGACCGGCCAAGTTTTACAGAGCTGCGGGAAGA
 T
 TTGGAGAACACACTGAAGGCCTTGCTCCTGCCCAGGAGCCTGACGAAATCCTCTATGTCAACATGG
 A
 TGAGGGTGGAGGTTATCCTGAACCCCTGGAGCTGCAGGAGGAGCTGACCCCCCAACCCAGCCAGAC
 C
 CTAAGGATTCCCTGTAGCTGCCTCACTGCGGCTGAGGTCCATCCTGCTGGACGCTATGTCCTCTGCCC
 T
 TCCACAACCCCTAGCCCCGCTCAGCCTGCTGATAGGGGCTCCCCAGCAGCCCCAGGGCAGGAGGATG
 G
 TGCCCTGAGACAACCTCCACCTGGTACTCCCTCTCAGGATCCAAGCTAAGCACTGCCACTGGGGAAA
 A
 CTCCACCTTCCCACCTTTCCC

CG59325-03	SEQ ID NO: 10	695 aa	MW at 76986.9kD
Protein Sequence			
MAWRCPMRGRVPLAWCLALCGWACMAPRGTAQAEESPFVGNPGNITGARESPGTLRCQLQVQGEPPPEVH WLRDQIILELADSTQTQVPLGEDEQDDWIVVSQLRITSLQLSDTGQYQCLVFLGHQTFVSQPGYVGLE GLPYFLEEPEDRTVAANTPFNLSCQAQGPPEPVDLLWLQDAVPLATAPGHGPQRS LHVPGLNKTSSFS CEAHNAKGVTTSRTATITVLPQQPRNLHLVSHQLVKESSTPAFSWPWWYVLLGAVVAAACVLIILALFL			

VHRRKKETRYGEVFEP TVDRGELVVRYRVRKSYSRRTEATLNSLGI SEELKEKL RDVMVDRHKVALG KTLGEGEF GAVMEGQLNQDD SILKVAVKTMKIAICTRSELEDFLSEAVCMKEFDHPNVMRLIGVCFQG SERESFPAPVVILP FMKHGDLHSFLLYSRLGDQPVYLP TQMLVKFMADIASGMEYLSTKRFIHRDLAA RNCMLNENMSVCVADFGLSKKIYNGDYRQGRIAKMPVKWIAIESLADRVYTSKSDVVSFGVTMW EIA TRGQTPYPGVENSEIYDYLRQGNRLKQPADCLDGLYALMSRCWELNPQDRPSFTELREDLENTL KALP PAQEPDEILYVNMDEGGGYPEPPGAAGGADPPTQDPKDCSCLTAAEVHPAGRYVLC PSTTPSPAQP ADRGSPAAPGQEDGA			
CG108678-01	SEQ ID NO: 11	1459 bp	
DNA Sequence	ORF Start: ATG at 2	ORF Stop: TGA at 1355	
<u>CATGGTGTCCCGGGACCAGGCTCACCTGGGCCCCAAGTATGTGGGCCTCTGGGACTTCAAGTCCCGGA</u> CGGACGAGGAGCTGAGCTTCCGCGCGGGGGACGTCTTCCACGTGGCCAGGAAGGAGGAGCAGTGGTGG TGGGCCACGCTGCTGGACGAGGCGGGTGGGGCCGTGGCCAGGGCTATGTGCCCCACAAC TACCTGGC CGAGAGGGAGACGGTGGAGTCGGAACCGTGGTTCTTTGGCTGCATCTCCCGCTCGGAAGCTGTGCGTC GGCTGCAGGCCGAGGGCAACGCCACGGGCGCCTTCCTGATCAGGGTCAGCGAGAAGCCGAGTGCCGAC TACGTCCTGTGCGTGCGGGACACGCAGGCTGTGCGGCAC TACAAGATCTGGCGGCGTGCCGGGGGCCG GCTGCACCTGAACGAGGCGGTGTCTTCTCAGCCTGCCCGAGCTTGTGAACTACCACAGGGCCCCAGA GCCTGTCCCACGGCCTGCGGCTGGCCGCGCCCTGCCGGAAGCACGAGCCTGAGCCCCCTGCCCATTTGG GATGACTGGGAGAGGCCGAGGGAGGAGTTACGCTCTGCAGGAAGCTGGGGTCCGGCTACTTTGGGGA GGTCTTCGAGGGGCTCTGGAAGACCGGGTCCAGGTGGCCATTAAGGTGATTTCTCGAGACAACCTCC TGCACCAGCAGATGCTGCAGTCGGAGATCCAGGCCATGAAGAAGCTGCGGCACAAACACATCCTGGCG CTGTACGCCGTGGTGTCCGTGGGGGACCCCGTGTACATCATCAGGAGCTCATGGCCAAGGGCAGCCT GCTGGAGCTGCTCCGCGACTCTGATGAGAAAGTCTGCCGTTTCGGAGCTGCTGGACATCGCCTGGC AGGTGGCTGAGGGCATGTGTTACCTGGAGTCGAGAATTACATCCACCGGGACCTGGCCGCCAGGAAC ATCCTCGTCGGGGAAAACACCCTCTGCAAAGTTGGGGACTTCGGGTTAGCCAGGCTTATCAAGGAGGA CGTCTACCTCTCCCATGACCACAATATCCCCTACAAGTGGACGGCCCCCTGAAGCGCTCTCCCGAGGCC ATTACTCCACCAAATCCGACGTCTGGTCTTTGGGATTCTCCTGCATGAGATGTTTCAGCAGGGGTCAG GTGCCCTACCCAGGCATGTCCAACCATGAGGCCTTCTGAGGGTGGACGCCGGCTACCGCATGCCCTG CCCTCTGGAGTGCCCGCCAGCGTGCACAAGCTGATGCTGACATGCTGGTGCAGGGACCCCGAGCAGA GACCTGCTTCAAGGCCCTGCGGGAGAGGCTCTCCAGCTTACCAGCTACGAGAACCCGACCTGAGCT GCTGTGGAGCGGGCATGGCCGGGCCCTGCTGAGGAGGGGCTGGGCAGAGGGCCTGGACCTGGGATCA AGGCCACGCGCTTCCCTGGGGTTTACTGAG			
CG108678-01	SEQ ID NO: 12	451 aa	MW at 51833.6kD
Protein Sequence			
MVSRDQAH LGPKYVGLWDFKSR TDEELSFRAGDV FHVARK EEQWWATLLDEAGGAVAQGYVPHNYL A ERETVESE PWFPGCISRSEAVRRLQAE GNATGAFLIRVSEKPSADYVLSVRDTQAVRHYKIWRRAGG R LHLNEAVS FLSLP ELVNYHRAQSLSHGLRLAAPCRKHEPEPLPHWDDWERPREEFTLCRKLGSYFG E VFEGWLKDRVQVAIKVISRDNLLHQQMLQSEIQAMKKLRHKHILALYAVVSVGDPVYIITELMAKGS L LELLRDSDEKVL PVSELLDI AWQVAEGMCYLESQNYIHRDLAARNILVGENTLCKVGD FGLARLIKE D VYLSHDHNI PYKWTAPEALSRGHYSTKSDVVSFGILLHEMF SRGQVPYPGMSNHEAFLRVDAGYRMP C PLECPPSVHKLMLT CWC RDPEQRPCFKALRERLSSFTSYENPT			

Example 5. Screening Assay Formulation

- 5 Assays for screening for antibody therapeutics or small molecule drugs targeting human SIK2 are formulated utilizing the non-exhaustive list of cell lines that express the SIK2 gene from the RTQ-PCR

results shown above. To assay the activity of SIK2 the measurement of [γ -³²P] incorporation into peptide substrate for instance GST-Syntide2 substrate is utilized (Takemori et al., J Biol Chem. 2002 Nov 1;277(44):42334-43). To assess the specific SIK2 activity the peptide substrate constructed from human IRS-1 protein fragment containing Ser794 is utilized (Horike et al., J Biol Chem. 2003 May 16;278(20):18440-7).

5

CLAIMS

We claim:

- 5 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
 (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
10 SEQ ID NO:2n, wherein n is an integer between 1 and 4, 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.
- 15 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.
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.
- 30 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.
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.
- 35 8. 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 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.

5 10. The method of claim 1, wherein the target polypeptide is an isolated polypeptide.

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 4, the amino acid sequence that is at
10 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.

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 4;
- (b) nucleotides encoding an amino acid sequence of the at least one domain of
20 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
25 ID NO:2n.

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 4, the amino acid sequence that is at least 95% identical to SEQ
30 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.

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:

- 5 (a) providing a cell expressing the target polypeptide and having a property or function ascribable to the target polypeptide;
- (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 identified as a potential therapeutic agent; and

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 4, 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.

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:

- 20 (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
- 25 (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 4, 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.

Figure 1.

Serine phosphorylation of IRS1/2 as a mechanism of insulin resistance

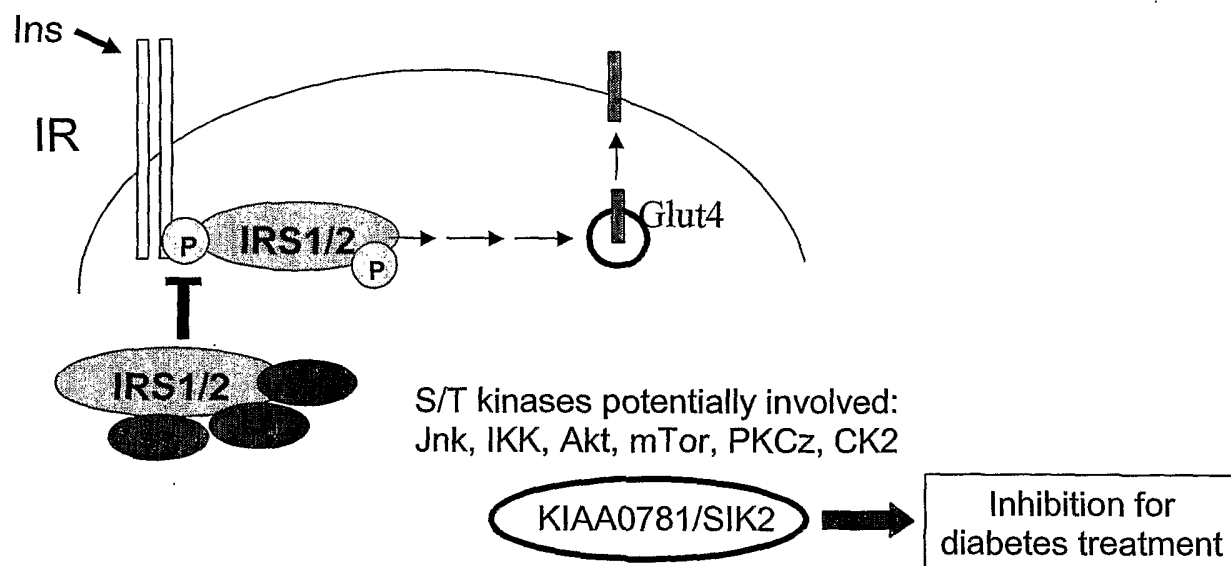


Figure 2.

