US 20060234292A1

(19) United States
(12) Patent Application Publication (10) Pub. No.: US 2006/0234292 A1
Allan et al. (43) Pub. Date: Oct. 19, 2006

(54) **METHODS OF DIAGNOSING AND TREATING DIABETES AND INSULIN RESISTANCE**

(76) Inventors: **Bernard Allan**, San Francisco, CA (US); **Francine Gregoire**, Lafayette, CA (US); **Brian Lavan**, San Francisco, CA (US); **Shonna Moodie**, San Francisco, CA (US); **Steve Waters**, San Ramon, CA (US); **Chi-Wai Wong**, Hayward, CA (US)

Correspondence Address:
TOWNSEND AND TOWNSEND AND CREW, LLP
TWO EMBARCADERO CENTER
EIGHTH FLOOR
SAN FRANCISCO, CA 94111-3834 (US)

(21) Appl. No.: **10/516,780**

(22) PCT Filed: **Jun. 5, 2003**

(86) PCT No.: **PCT/US03/18046**

Related U.S. Application Data

(60) Provisional application No. 60/386,551, filed on Jun. 5, 2002. Provisional application No. 60/386,527, filed on Jun. 5, 2002. Provisional application No. 60/386,521, filed on Jun. 5, 2002. Provisional application No. 60/386,936, filed on Jun. 6, 2002. Provisional application No. 60/386,429, filed on Jun. 6, 2002. Provisional application No. 60/386,954, filed on Jun. 6, 2002. Provisional application No. 60/387,301, filed on Jun. 7, 2002.

Publication Classification

(51) **Int. Cl.**

A61K 38/22	(2006.01)
G01N 33/53	(2006.01)
C07H 21/04	(2006.01)
C07K 14/575	(2006.01)

(52) **U.S. Cl.** **435/7.1**; 435/69.4; 435/320.1; 435/325; 514/12; 530/399; 536/23.5

(57) **ABSTRACT**

The present invention provides compositions and methods for diagnosing and treating diabetes and insulin resistance. In particular, the invention provides methods of identifying modulators of the polynucleotides or polypeptides of the invention and using those modulators to treat diabetes, as well as methods of diagnosing diabetes by measuring the levels of the polynucleotides or polypeptides of the invention in a patient.

METHODS OF DIAGNOSING AND TREATING DIABETES AND INSULIN RESISTANCE

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority of U.S. provisional application No. 60/386,429, filed Jun. 6, 2002; U.S. provisional application No. 60/386,551, filed Jun. 5, 2002; U.S. provisional application No. 60/387,301, filed Jun. 7, 2002; U.S. provisional application No. 60/386,527, filed Jun. 5, 2002; U.S. provisional application No. 60/386,521, filed Jun. 5, 2002; U.S. provisional application No. 60/386,954, filed Jun. 6, 2002; and U.S. provisional application No. 60/386,936, filed Jun. 6, 2002, each of which applications is herein incorporated by referenced.

BACKGROUND OF THE INVENTION

[0002] Diabetes mellitus can be divided into two clinical syndromes, Type 1 and Type 2 diabetes mellitus. Type 1, or insulin-dependent diabetes mellitus (IDDM), is a chronic autoimmune disease characterized by the extensive loss of beta cells in the pancreatic Islets of Langerhans, which produce insulin. As these cells are progressively destroyed, the amount of secreted insulin decreases, eventually leading to hyperglycemia (abnormally high level of glucose in the blood) when the amount of secreted insulin drops below the level required for euglycemia (normal blood glucose level). Although the exact trigger for this immune response is not known, patients with IDDM have high levels of antibodies against proteins expressed in pancreatic beta cells. However, not all patients with high levels of these antibodies develop IDDM.

[0003] Type 2 diabetes (also referred to as non-insulin dependent diabetes mellitus (NIDDM)) develops when muscle, fat and liver cells fail to respond normally to insulin. This failure to respond (called insulin resistance) may be due to reduced numbers of insulin receptors on these cells, or a dysfunction of signaling pathways within the cells, or both. The beta cells initially compensate for this insulin resistance by increasing insulin output. Over time, these cells become unable to produce enough insulin to maintain normal glucose levels, indicating progression to Type 2 diabetes.

[0004] Type 2 diabetes is brought on by a combination of genetic and acquired risk factors, including a high-fat diet, lack of exercise, and aging. Worldwide, Type 2 diabetes has become an epidemic, driven by increases in obesity and a sedentary lifestyle, widespread adoption of western dietary habits, and the general aging of the population in many countries. In 1985, an estimated 30 million people worldwide had diabetes—by 2000, this figure had increased 5-fold, to an estimated 154 million people. The number of people with diabetes is expected to double between now and 2025, to about 300 million.

[0005] Type 2 diabetes is a complex disease characterized by defects in glucose and lipid metabolism. Typically there are perturbations in many metabolic parameters including increases in fasting plasma glucose levels, free fatty acid levels and triglyceride levels, as well as a decrease in the ratio of HDL/LDL. As discussed above, one of the principal underlying causes of diabetes is thought to be an increase in insulin resistance in peripheral tissues, principally muscle and fat.

[0006] Therapies aimed at reducing peripheral insulin resistance are available. The most relevant to this invention are drugs of the thiazolidinedione (TZD) class namely troglitazone, pioglitazone, and rosiglitazone. In the US these have been marketed under the names Rezulin™, Avandia™ and Actos™, respectively. The principal effect of these drugs is to improve glucose homeostasis. Notably in diabetics treated with TZDs there are increases in peripheral glucose disposal rates indicative of increased insulin sensitivity in both muscle and fat.

[0007] The molecular target of TZDs is a member of the PPAR family of ligand-activated transcription factors called PPAR gamma. This transcription factor is highly expressed in adipose tissue with much lower levels being observed in muscle. Binding of TZDs to PPAR gamma in target cells and tissues such as fat and muscle brings about a change in gene expression. The link between TZD-altered gene expression in fat and muscle and increased insulin sensitivity is unknown. The present invention addresses this and other problems. The present invention addresses this and other problems.

BRIEF SUMMARY OF THE INVENTION

[0008] The present invention provides methods for identifying an agent for treating a diabetic or pre-diabetic individual. In some embodiments, the methods comprise the steps of: (i) contacting an agent to a mixture comprising a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a nucleic acid encoding SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34; and (ii) selecting an agent that modulates the expression or activity of the polypeptide or that binds to the polypeptide, thereby identifying an agent for treating a diabetic or pre-diabetic individual. In some embodiments, the methods further comprise selecting an agent that modulates insulin sensitivity.

[0009] In some embodiments, step (ii) comprises selecting an agent that modulates expression of the polypeptide. In some embodiments, step (ii) comprises selecting an agent that modulates the activity of the polypeptide. In some embodiments, step (ii) comprises selecting an agent that specifically binds to the polypeptide. In some embodiments, the polypeptide is expressed in a cell and the cell is contacted with the agent. In some embodiments, the polypeptide is SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34. In other embodiments, the polypeptide is SEQ ID NO:4, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:18, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:32, SEQ ID NO:36, or SEQ ID NO:38.

[0010] The present invention also provides methods of treating a diabetic or pre-diabetic animal. In some embodiments, the methods comprise administering to the animal a therapeutically effective amount of an agent identified as described above. In some embodiments, the agent is an antibody. In some embodiments, the antibody is a monoclonal antibody. In some embodiments, the animal is a human.

[0011] The present invention also provides methods of introducing an expression cassette into a cell. In some

emodiments, the methods comprise introducing into the cell an expression cassette comprising a promoter operably linked to a polynucleotide encoding a polypeptide, wherein the polynucleotide hybridizes under stringent conditions to a nucleic acid encoding SEQ BD NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

[0012] In some embodiments, the polypeptide comprises SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ D NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34. In other embodiments, the polypeptide comprises SEQ ID NO:4, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:18, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:32, SEQ ID NO:36, or SEQ ID NO:38. In some embodiments, the cell is selected from the group consisting of an adipocyte and a skeletal muscle cell.

[0013] In some embodiments, the methods further comprising introducing the cell into a human. In some embodiments, the human is diabetic. In some embodiments, the human is prediabetic. In some embodiments, the cell is from the human.

[0014] The present invention also provides methods of diagnosing an individual who has Type 2 diabetes or is prediabetic. In some embodiments, the method comprises, detecting in a sample from the individual the level of a polypeptide or the level of a polynucleotide encoding the polypeptide, wherein the polynucleotide hybridizes under stringent conditions to a nucleic acid encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:34; wherein a modulated level of the polypeptide or polynucleotide in the sample compared to a level of the polypeptide or polynucleotide in either a lean individual or a previous sample from the individual indicates that the individual is diabetic or prediabetic. In some embodiments, the amino acid sequence comprises SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

[0015] In some embodiments, the detecting step comprises contacting the sample with an antibody that specifically binds to the polypeptide.

[0016] In some embodiments, the detecting step comprises quantifying mRNA encoding the polypeptide. In some embodiments, the mRNA is reverse transcribed and amplified in a polymerase chain reaction.

[0017] In some embodiments, the sample is a blood, urine or tissue sample.

[0018] The present invention also provides isolated nucleic acids that hybridize under stringent conditions to a polynucleotide encoding a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:34.

[0019] In some embodiments, the nucleic acid is SEQ ID NO:1, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:15, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33. In some embodiments, the nucleic acid encodes SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

[0020] In other embodiments, the nucleic acid is SEQ ID NO:3, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:17, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:31, SEQ ID NO:35, or SEQ ID NO:37. In some embodiments, the nucleic acid encodes SEQ ID NO:4, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:18, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:32, SEQ ID NO:36, or SEQ ID NO:38.

[0021] The present invention also provides an expression cassette comprising a heterologous promoter operably linked to a polynucleotide that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:34.

[0022] In some embodiments, the polynucleotide comprises SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35 or SEQ ID NO:37. In some embodiments, the polynucleotide encodes SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36 or SEQ ID NO:38.

[0023] The present invention also provides host cells transfected with a polynucleotide that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide having an amino acid sequence selected from the group consisting of SEQ BD NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ BD NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:34. In some embodiments, the polynucleotide encodes SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36 or SEQ ID NO:38. In some embodiments, the polynucleotide comprises SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ BD NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35 or SEQ ID NO:37. In some embodiments, the host cell is a human cell. In other embodiments, the host cell is a bacterium.

[0024] The present invention also provides isolated polypeptides comprising an amino acid sequence at least 70% identical to SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:34, SEQ ID NO:10, or SEQ ID NO:28. In some embodi-

ments, the polypeptide is SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36 or SEQ ID NO:38.

DEFINITIONS

[0025] "Insulin sensitivity" refers to the ability of a cell or tissue to respond to insulin. Responses include, e.g., glucose uptake of a cell or tissue in response to insulin stimulation. Sensitivity can be determined at an organismal, tissue or cellular level. For example, blood or urine glucose levels following a glucose tolerance test are indicative of insulin sensitivity. Other methods of measuring insulin sensitivity include, e.g., measuring glucose uptake (see, e.g., Garcia de Herreros, A., and Birnbaum, M. J. *J. Biol. Chem.* 264, 19994-19999 (1989); Klip, A., Li, G., and Logan, W. J. *Am. J. Physiol.* 247, E291-296 (1984)), measuring the glucose infusion rate (GINF) into tissue such as the skeletal muscle (see, e.g., Ludvik et al., *J. Clin. Invest.* 100:2354 (1997); Frias et al., *Diabetes Care* 23:64, (2000)) and measuring sensitivity of GLUT4 translocation (e.g., as described herein) in response to insulin.

[0026] "Activity" of a polypeptide of the invention refers to structural, regulatory, or biochemical functions of a polypeptide in its native cell or tissue. Examples of activity of a polypeptide include both direct activities and indirect activities. Exemplary direct activities are the result of firect interaction with the polypeptide, e.g., enzymatic activity, ligand binding, production or depletion of second messengers (e.g., cAMP, cGMP, IP_3, DAG, or Ca^{2+}), ion flux, phosphorylation levels, transcription levels, and the like. Exemplary indirect activities are observed as a change in phenotype or response in a cell or tissue to a polypeptide's directed activity, e.g., modulating insulin sensitivity of a cell as a result of the interaction of the polypeptide with other cellular or tissue components.

[0027] "Predisposition for diabetes" occurs in a person when the person is at high risk for developing diabetes. A number of risk factors are known to those of skill in the art and include: genetic factors (e.g., carrying alleles that result in a higher occurrence of diabetes than in the average population or having parents or siblings with diabetes); overweight (e.g., body mass index (BMI) greater or equal to 25 kg/m^2); habitual physical inactivity, race/ethnicity (e.g., African-American, Hispanic-American, Native Americans, Asian-Americans, Pacific Islanders); previously identified impaired fasting glucose or impaired glucose tolerance, hypertension (e.g., greater or equal to 140/90 mmHg in adults); HDL cholesterol less than or equal to 35 mg/dl; triglyceride levels greater or equal to 250 mg/dl; a history of gestational diabetes or delivery of a baby over nine pounds; and/or polycystic ovary syndrome. See, e.g., "Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus" and "Screening for Diabetes"*Diabetes Care* 25(1): S5-S24 (2002).

[0028] A "lean individual," when used to compare with a sample from a patient, refers to an adult with a fasting blood glucose level less than 110 mg/dl or a 2 hour PG reading of 140 mg/dl. "Fasting" refers to no caloric intake for at least 8 hours. A "2 hour PG" refers to the level of blood glucose after challenging a patient to a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water. The overall test is generally referred to as an oral glucose tolerance test (OGTT). See, e.g., *Diabetes Care*, Supplement 2002, American Diabetes Association: Clinical Practice Recommendations 2002. The level of a polypeptide in a lean individual can be a reading from a single individual, but is typically a statistically relevant average from a group of lean individuals. The level of a polypeptide in a lean individual can be represented by a value, for example in a computer program.

[0029] A "pre-diabetic individual," when used to compare with a sample from a patient, refers to an adult with a fasting blood glucose level greater than 110 mg/dl but less than 126 mg/dl or a 2 hour PG reading of greater than 140 mg/dl but less than 200 mg/dl. A "diabetic individual," when used to compare with a sample from a patient, refers to an adult with a fasting blood glucose level greater than 126 mg/dl or a 2 hour PG reading of greater than 200 mg/dl.

[0030] A "diabetes-related nucleic acid" or "diabetes-related polynucleotide" (also referred to as a "nucleic acid of the invention" or a "polynucleotide of the invention") of the invention is a subsequence or full-length polynucleotide sequence of a gene that encodes a polypeptide, whose activity modulates diabetes or insulin sensitivity, or whose presence or absence is indicative of diabetes or altered insulin sensitivity. Exemplary nucleic acids of the invention include those sequences substantially identical to SEQ ID NO:1, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:15, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33 or encode polypeptides substantially identical to SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

[0031] An "agonist" refers to an agent that binds to, stimulates, increases, activates, facilitates, enhances activation, sensitizes or up regulates the activity or expression of a polypeptide of the invention.

[0032] An "antagonist" refers to an agent that binds to, partially or totally blocks stimulation, decreases, prevents, delays activation, inactivates, desensitizes, or down regulates the activity or expression of a polypeptide of the invention.

[0033] "Antibody" refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

[0034] An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily

responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

[0035] Antibodies exist, e.g., as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce $F(ab)'_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The $F(ab)'_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the $F(ab)'_2$ dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (see, Paul (Ed.) *Fundamental Immunology*, Third Edition, Raven Press, NY (1993)). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized de novo either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized de novo using recombinant DNA methodologies (e.g., single chain Fv).

[0036] The terms "peptidomimetic" and "mimetic" refer to a synthetic chemical compound that has substantially the same structural and functional characteristics of the antagonists or agonists of the invention. Peptide analogs are commonly used in the pharmaceutical industry as non-peptide drugs with properties analogous to those of the template peptide. These types of non-peptide compound are termed "peptide mimetics" or "peptidomimetics" (Fauchere, J. *Adv. Drug Res.* 15:29 (1986); Veber and Freidinger *TINS p.* 392 (1985); and Evans et al. *J. Med. Chem.* 30:1229 (1987), which are incorporated herein by reference). Peptide mimetics that are structurally similar to therapeutically useful peptides maybe used to produce an equivalent or enhanced therapeutic or prophylactic effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (i.e., a polypeptide that has a biological or pharmacological activity), such as apolypeptide exemplified in this application, but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of, e.g., —CH2NH—, —CH2S—, —CH2—CH2-, —CH═CH— (cis and trans), —COCH2—, —CH(OH)CH2—, and —CH2SO—. The mimetic can be either entirely composed of synthetic, non-natural analogues of amino acids, or, is a chimeric molecule of partly natural peptide amino acids and partly non-natural analogs of amino acids. The mimetic can also incorporate any amount of natural amino acid conservative substitutions as long as such substitutions also do not substantially alter the mimetic's structure and/or activity. For example, a mimetic composition is within the scope of the invention if it is capable of carrying out the binding or other activities of an agonist or antagonist of a polypeptide of the invention.

[0037] The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

[0038] The term "isolated," when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein that is the predominant species present in a preparation is substantially purified. In particular, an isolated gene is separated from open reading frames that flank the gene and encode a protein other than the gene of interest. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

[0039] The term "nucleic acid" or "polynucleotide" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka et al., *J. Biol. Chem.* 260:2605-2608 (1985); and Cassol et al. (1992); Rossolini et al., *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

[0040] The terms "polypeptide,""peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. As used herein, the terms encompass amino acid chains of any length, including full-length proteins (i.e., antigens), wherein the amino acid residues are linked by covalent peptide bonds.

[0041] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. "Amino acid mimetics" refers to chemical compounds that have a structure that is different from the general chemical structure of

an amino acid, but which functions in a manner similar to a naturally occurring amino acid.

[0042] Amino acids may be referred to herein by either the commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[0043] "Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, "conservatively modified variants" refers to those nucleic acids that encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein that encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid that encodes a polypeptide is implicit in each described sequence.

[0044] As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

[0045] The following eight groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Glycine (G);

2) Aspartic acid (D), Glutamic acid (E);

3) Asparagine (N), Glutamine (Q);

4) Arginine (R), Lysine (K);

5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);

6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);

7) Serine (S), Threonine (T); and

8) Cysteine (C), Methionine (V)

(see, e.g., Creighton, Proteins (1984)).

[0046] "Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (e.g., a polypeptide of the invention), which does not comprise additions or deletions, for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

[0047] The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same sequences are substantially identical if two sequences have a specified percentage of amino acid residues or nucleotides that are the same (i.e., 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% identity over a specified region, or, when not specified, over the entire sequence), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. The invention provides polypeptides or polynucleotides that are substantially identical to the polypeptides or polynucleotides, respectively, exemplified herein (e.g., SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34; or SEQ ID NO:1, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:15, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33). This definition also refers to the complement of a test sequence. Optionally, the identity exists over a region that is at least about 50 nucleotides in length, or more preferably over a region that is 100 to 500 or 1000 or more nucleotides in length.

[0048] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0049] A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1970) *Adv. Appl. Math.* 2:482c, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Nat* 1. Acad. Sci. USA 85:2444, by computerized implementa-

tions of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection (see, e.g., Ausubel et al., *Current Protocols in Molecular Biology* (1995 supplement)).

[0050] Two examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nuc. Acids Res.* 25:3389-3402, and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always>0) and N (penalty score for mismatching residues; always<0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) or 10, M=5, N=−4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N−−4, and a comparison of both strands.

[0051] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[0052] An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

[0053] The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

[0054] The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10° C. lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C. for short probes (e.g., 10 to 50 nucleotides) and at least about 60° C. for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5×SSC, and 1% SDS, incubating at 42° C., or 5×SSC, 1% SDS, incubating at 65° C., with wash in 0.2×SSC, and 0.1% SDS at 55° C., 60° C., or 65° C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes.

[0055] Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides that they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37° C., and a wash in 1×SSC at 45° C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes. A positive hybridization is at least twice background. Those of ordinary skill will readily

recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

[0056] The phrase "a nucleic acid sequence encoding" refers to a nucleic acid which contains sequence information for a structural RNA such as rRNA, a tRNA, or the primary amino acid sequence of a specific protein or peptide, or a binding site for a trans-acting regulatory agent. This phrase specifically encompasses degenerate codons (i.e., different codons which encode a single amino acid) of the native sequence or sequences that may be introduced to conform with codon preference in a specific host cell.

[0057] The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (nonrecombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all.

[0058] The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, e.g., a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a fusion protein).

[0059] An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

[0060] The phrase "specifically (or selectively) binds to an antibody" or "specifically (or selectively) immunoreactive with", when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised against a protein having an amino acid sequence encoded by any of the polynucleotides of the invention can be selected to obtain antibodies specifically immunoreactive with that protein and not with other proteins, except for polymorphic variants. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays, Western blots, or immunohistochemistry are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See, Harlow and Lane Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, NY (1988) for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. Typically, a specific or selective reaction will be at least twice the background signal or noise and more typically more than 10 to 100 times background.

[0061] "Inhibitors,""activators," and "modulators" of expression or of activity are used to refer to inhibitory, activating, or modulating molecules, respectively, identified using in vitro and in vivo assays for expression or activity. Modulators encompass e.g., ligands, agonists, antagonists, and their homologs and mimetics. The term "modulator" includes inhibitors and activators. Inhibitors are agents that, e.g., inhibit expression of a polypeptide of the invention or bind to, partially or totally block stimulation, decrease, prevent, delay activation, inactivate, desensitize, or down regulate the activity of a polypeptide of the invention, e.g., antagonists. Activators are agents that, e.g., induce or activate the expression of a polypeptide of the invention or bind to, stimulate, increase, open, activate, facilitate, or enhance activation, sensitize or up regulate the activity of a polypeptide of the invention, e.g., agonists. Modulators include naturally occurring and synthetic ligands, antagonists, agonists, small chemical molecules and the like. Such assays for inhibitors and activators include, e.g., applying putative modulator compounds to cells expressing a polypeptide of the invention and then determining the functional effects on a polypeptide of the invention activity, as described above. Samples or assays comprising a polypeptide of the invention that are treated with a potential activator, inhibitor, or modulator are compared to control samples without the inhibitor, activator, or modulator to examine the extent of effect. Control samples (untreated with modulators) are assigned a relative activity value of 100%. Inhibition of a polypeptide of the invention is achieved when the polypeptide activity value relative to the control is about 80%, optionally 50% or 25, 10%, 5% or 1%. Activation of the polypeptide is achieved when the polypeptide activity value relative to the control is 110%, optionally 150%, optionally 200, 300%, 400%, 500%, or 1000-3000% or more higher.

DETAILED DESCRIPTION OF THE INVENTION

I. Introduction

[0062] The present application demonstrates that, surprisingly, modulated levels of mRNA comprising sequences of the invention occur in muscle tissue of insulin-resistant obese, non-diabetic individuals (which population is generally predisposed to become type 2 diabetics) or type 2 diabetic individuals in comparison to the levels in muscle tissue from lean, non-diabetic individuals. Furthermore, in some instances, mRNA levels of sequences described herein in muscle tissue from type 2 diabetic individuals treated with thiazolidinedione (TZD) is changed in comaprison to levels of the mRNA in type 2 diabetic individuals before TZD treatment. Therefore, the modulation of the sequences in the study described herein indicates the sequences' involvement in diabetes and pre-diabetes.

[0063] Without intending to limit the invention to a particular mechanism of action, it is believed that modulation of the expression or activity of the polypeptides of the invention is beneficial in treating diabetic, pre-diabetic or obese

insulin resistant, non-diabetic patients. Furthermore, modulated levels of the polypeptides of the invention are indicative of insulin resistance. Thus, the detection of a polypeptide of the invention is useful for diagnosis of diabetes and insulin resistance.

[0064] This invention also provides methods of using polypeptides of the invention and modulators of the polypeptides of the invention to diagnose and treat diabetes, pre-diabetes (including insulin resistant individuals) and related metabolic diseases. The present method also provides methods of identifying modulators of expression or activity of the polypeptides of the invention. Such modulators are useful for treating Type 2 diabetes as well as the pathological aspects of diabetes (e.g., insulin resistance).

II. General Recombinant Nucleic Acid Methods for use with the Invention

[0065] In numerous embodiments of the present invention, nucleic acids encoding a polypeptide of the present invention will be isolated and cloned using recombinant methods. Such embodiments are used, e.g., to isolate polynucleotides identical or substantially identical to SEQ ID NO:1, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:15, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33, for protein expression or during the generation of variants, derivatives, expression cassettes, or other sequences derived from an polypeptide or polynucleotide of the invention, to monitor gene expression, for the isolation or detection of sequences in different species, for diagnostic purposes in a patient, e.g., to detect mutations in a polypeptide or polynucleotide of the invention or to detect expression levels of nucleic acids or polypeptides. In some embodiments, the sequences encoding the polypeptides of the invention are operably linked to a heterologous promoter. In one embodiment, the nucleic acids of the invention are from any mammal, including, in particular, e.g., a human, a mouse, a rat, etc.

A. General Recombinant Nucleic Acid Methods

[0066] This invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook et al., *Molecular Cloning, A Laboratory Manual* (3rd ed. 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel et al., eds., 1994)).

[0067] For nucleic acids, sizes are given in either kilobases (kb) or base pairs (bp). These are estimates derived from agarose or acrylamide gel electrophoresis, from sequenced nucleic acids, or from published DNA sequences. For proteins, sizes are given in kilodaltons (kDa) or amino acid residue numbers. Proteins sizes are estimated from gel electrophoresis, from sequenced proteins, from derived amino acid sequences, or from published protein sequences.

[0068] Oligonucleotides that are not commercially available can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage & Caruthers, *Tetrahedron Letts.* 22:1859-1862 (1981), using an automated synthesizer, as described in Van Devanter et. al., *Nucleic Acids Res.* 12:6159-6168 (1984). Purification of oligonucleotides is by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson & Reanier, *J. Chrom.* 255:137-149 (1983).

[0069] The sequence of the cloned genes and synthetic oligonucleotides can be verified after cloning using, e.g., the chain termination method for sequencing double-stranded templates of Wallace et al., *Gene* 16:21-26 (1981).

B. Cloning Methods for the Isolation of Nucleotide Sequences Encoding Desired Proteins

[0070] In general, the nucleic acids encoding the subject proteins are cloned from DNA sequence libraries that are made to encode cDNA or genomic DNA. The particular sequences can be located by hybridizing with an oligonucleotide probe, the sequence of which can be derived from the sequences disclosed herein, which provide a reference for PCR primers and defines suitable regions for isolating probes pecific for the polypeptides or polynucleotides of the invention. Alternatively, where the sequence is cloned into an expression library, the expressed recombinant protein can be detected immunologically with antisera or purified antibodies made against a polypeptide of interest, including those disclosed herein.

[0071] Methods for making and screening genomic and cDNA libraries are well known to those of skill in the art (see, e.g., Gubler and Hoffman Gene 25:263-269 (1983); Benton and Davis *Science,* 196:180-182 (1977); and Sambrook, supra).

[0072] Briefly, to make the cDNA library, one should choose a source that is rich in mRNA. The mRNA can then be made into cDNA, ligated into a recombinant vector, and transfected into a recombinant host for propagation, screening and cloning. For a genomic library, the DNA is extracted from a suitable tissue and either mechanically sheared or enzymatically digested to yield fragments of preferably about 5-100 kb. The fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged in vitro, and the recombinant phages are analyzed by plaque hybridization. Colony hybridization is carried out as generally described in Grunstein et al., *Proc. Natl. Acad. Sci. USA.,* 72:3961-3965 (1975).

[0073] An alternative method combines the use of synthetic oligonucleotide primers with polymerase extension on an mRNA or DNA template. Suitable primers can be designed from specific sequences disclosed herein. This polymerase chain reaction (PCR) method amplifies the nucleic acids encoding the protein of interest directly from mRNA, cDNA, genomic libraries or cDNA libraries. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or other in vitro amplification methods may also be useful, for example, to clone nucleic acids encoding specific proteins and express said proteins, to synthesize nucleic acids that will be used as probes for detecting the presence of mRNA encoding a polypeptide of the invention in physiological samples, for nucleic acid sequencing, or for other purposes (see, U.S. Pat. Nos. 4,683,195 and 4,683,202). Genes amplified by a PCR reaction can be purified from agarose gels and cloned into an appropriate vector.

[0074] Appropriate primers and probes for identifying the genes encoding a polypeptide of the invention from mammalian tissues can be derived from the sequences provided herein. For a general overview of PCR, see, Innis et al. *PCR Protocols: A Guide to Methods and Applications, Academic Press*, San Diego (1990).

[0075] Synthetic oligonucleotides can be used to construct genes. This is done using a series of overlapping oligonucleotides, usually 40-120 bp in length, representing both the sense and anti-sense strands of the gene. These DNA fragments are then annealed, ligated and cloned.

[0076] A polynucleotide encoding a polypeptide of the invention can be cloned using intermediate vectors before transformation into mammalian cells for expression. These intermediate vectors are typically prokaryote vectors or shuttle vectors. The proteins can be expressed in either prokaryotes or eukaryotes, using standard methods well known to those of skill in the art.

III. Purification of Proteins of the Invention

[0077] Either naturally occurring or recombinant polypeptides of the invention can be purified for use in functional assays. Naturally occurring polypeptides of the invention can be purified from any source (e.g., tissues of an organism expressing an ortholog). Recombinant polypeptides can be purified from any suitable expression system.

[0078] The polypeptides of the invention may be purified to substantial purity by standard techniques, including selective precipitation with such substances as ammonium sulfate; column chromatography, immunopurification methods, and others (see, e.g., Scopes, *Protein Purification: Principles and Practice* (1982); U.S. Pat. No. 4,673,641; Ausubel et al., supra; and Sambrook et al., supra).

[0079] A number of procedures can be employed when recombinant polypeptides are being purified. For example, proteins having established molecular adhesion properties can be reversibly fused to a polypeptide of the invention. With the appropriate ligand, either protein can be selectively adsorbed to a purification column and then freed from the column in a relatively pure form. The fused protein may be then removed by enzymatic activity. Finally polypeptides can be purified using immunoaffinity columns.

A. Purification of Proteins from Recombinant Bacteria

[0080] When recombinant proteins are expressed by the transformed bacteria in large amounts, typically after promoter induction, although expression can be constitutive, the proteins may form insoluble aggregates. There are several protocols that are suitable for purification of protein inclusion bodies. For example, purification of aggregate proteins (hereinafter referred to as inclusion bodies) typically involves the extraction, separation and/or purification of inclusion bodies by disruption of bacterial cells typically, but not limited to, by incubation in a buffer of about 100-150 μg/ml lysozyme and 0.1% Nonidet P40, a non-ionic detergent. The cell suspension can be ground using a Polytron grinder (Brinkman Instruments, Westbury, N.Y.). Alternatively, the cells can be sonicated on ice. Alternate methods of lysing bacteria are described in Ausubel et al. and Sambrook et al., both supra, and will be apparent to those of skill in the art.

[0081] The cell suspension is generally centrifuged and the pellet containing the inclusion bodies resuspended in buffer which does not dissolve but washes the inclusion bodies, e.g., 20 mM Tris-HCl (pH 7.2), 1 mM EDTA, 150 mM NaCl and 2% Triton-X 100, a non-ionic detergent. It may be necessary to repeat the wash step to remove as much cellular debris as possible. The remaining pellet of inclusion bodies may be resuspended in an appropriate buffer (e.g., 20 mM sodium phosphate, pH 6.8, 150 mM NaCl). Other appropriate buffers will be apparent to those of skill in the art.

[0082] Following the washing step, the inclusion bodies are solubilized by the addition of a solvent that is both a strong hydrogen acceptor and a strong hydrogen donor (or a combination of solvents each having one of these properties). The proteins that formed the inclusion bodies may then be renatured by dilution or dialysis with a compatible buffer. Suitable solvents include, but are not limited to, urea (from about 4 M to about 8 M), formamide (at least about 80%, volume/volume basis), and guanidine hydrochloride (from about 4 M to about 8 M). Some solvents that are capable of solubilizing aggregate-forming proteins, such as SDS (sodium dodecyl sulfate) and 70% formic acid, are inappropriate for use in this procedure due to the possibility of irreversible denaturation of the proteins, accompanied by a lack of immunogenicity and/or activity. Although guanidine hydrochloride and similar agents are denaturants, this denaturation is not irreversible and renaturation may occur upon removal (by dialysis, for example) or dilution of the denaturant, allowing re-formation of the immunologically and/or biologically active protein of interest. After solubilization, the protein can be separated from other bacterial proteins by standard separation techniques.

[0083] Alternatively, it is possible to purify proteins from bacteria periplasm. Where the protein is exported into the periplasm of the bacteria, the periplasmic fraction of the bacteria can be isolated by cold osmotic shock in addition to other methods known to those of skill in the art (see, Ausubel et al., supra). To isolate recombinant proteins from the periplasm, the bacterial cells are centrifuged to form a pellet. The pellet is resuspended in a buffer containing 20% sucrose. To lyse the cells, the bacteria are centrifuged and the pellet is resuspended in ice-cold 5 mM $MgSO_4$ and kept in an ice bath for approximately 10 minutes. The cell suspension is centrifuged and the supernatant decanted and saved. The recombinant proteins present in the supernatant can be separated from the host proteins by standard separation techniques well known to those of skill in the art.

B. Purification of Proteins from Insect Cells

[0084] Proteins can also be purified from eukaryotic gene expression systems as described in, e.g., Fernandez and Hoeffler, *Gene Expression Systems* (1999). In some embodiments, baculovirus expression systems are used to isolate proteins of the invention. Recombinant baculoviruses are generally generated by replacing the polyhedrin coding sequence of a baculovirus with a gene to be expressed (e.g., encoding a polypeptide of the invention). Viruses lacking the polyhedrin gene have a unique plaque morphology making them easy to recognize. In some embodiments, a recombinant baculovirus is generated by first cloning a polynucleotide of interest into a transfer vector (e.g., a pUC based vector) such that the polynucleotide is operably linked to a polyhedrin promoter. The transfer vector is transfected with wildtype DNA into an insect cell (e.g., Sf9, Sf21 or BT1-TN-5B1-4 cells), resulting in homologous recombination and replacement of the polyhedrin gene in the wildtype viral DNA with the polynucleotide of interest. Virus can then be generated and plaque purified. Protein expression results upon viral infection of insect cells. Expressed proteins can

be harvested from cell supernatant if secreted, or from cell lysates if intracellular. See, e.g., Ausubel et al. and Fernandez and Hoeffler, supra.

C. Standard Protein Separation Techniques For Purifying Proteins

1. Solubility Fractionation

[0085] Often as an initial step, and if the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the recombinant protein of interest. The preferred salt is ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol is to add saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is between 20-30%. This will precipitate the most hydrophobic proteins. The precipitate is discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer and the excess salt removed if necessary, through either dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well known to those of skill in the art and can be used to fractionate complex protein mixtures.

2. Size Differential Filtration

[0086] Based on a calculated molecular weight, a protein of greater and lesser size can be isolated using ultrafiltration through membranes of different pore sizes (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

3. Column Chromatography

[0087] The proteins of interest can also be separated from other proteins on the basis of their size, net surface charge, hydrophobicity and affinity for ligands. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these methods are well known in the art.

[0088] Immunoaffinity chromatography using antibodies raised to a variety of affinity tags such as hemagglutinin (HA), FLAG, Xpress, Myc, hexahistidine (His), glutathione S transferase (GST) and the like can be used to purify polypeptides. The His tag will also act as a chelating agent for certain metals (e.g., Ni) and thus the metals can also be used to purify His-containing polypeptides. After purification, the tag is optionally removed by specific proteolytic cleavage.

[0089] It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (e.g. Pharmacia Biotech).

IV. Detection of Polynucleotides of the Invention

[0090] Those of skill in the art will recognize that detection of expression of polynucleotides and polypeptides of the invention has many uses. For example, as discussed herein, detection of levels of polynucleotides and polypeptides of the invention in a patient is useful for diagnosing diabetes or a predisposition for at least some of the pathological effects of diabetes. Moreover, detection of gene expression is useful to identify modulators of expression of polynucleotides and polypeptides of the invention.

[0091] A variety of methods of specific DNA and RNA measurement that use nucleic acid hybridization techniques are known to those of skill in the art (see, Sambrook, supra). Some methods involve an electrophoretic separation (e.g., Southern blot for detecting DNA, and Northern blot for detecting RNA), but measurement of DNA and RNA can also be carried out in the absence of electrophoretic separation (e.g., by dot blot). Southern blot of genomic DNA (e.g., from a human) can be used for screening for restriction fragment length polymorphism (RFLP) to detect the presence of a genetic disorder affecting a polypeptide of the invention.

[0092] The selection of a nucleic acid hybridization format is not critical. A variety of nucleic acid hybridization formats are known to those skilled in the art. For example, common formats include sandwich assays and competition or displacement assays. Hybridization techniques are generally described in Hames and Higgins *Nucleic Acid Hybridization, A Practical Approach*, IRL Press (1985); Gall and Pardue, *Proc. Natl. Acad. Sci. U.S.A.,* 63:378-383 (1969); and John et al. *Nature,* 223:582-587 (1969).

[0093] Detection of a hybridization complex may require the binding of a signal-generating complex to a duplex of target and probe polynucleotides or nucleic acids. Typically, such binding occurs through ligand and anti-ligand interactions as between a ligand-conjugated probe and an anti-ligand conjugated with a signal. The binding of the signal generation complex is also readily amenable to accelerations by exposure to ultrasonic energy.

[0094] The label may also allow indirect detection of the hybridization complex. For example, where the label is a hapten or antigen, the sample can be detected by using antibodies. In these systems, a signal is generated by attaching fluorescent or enzyme molecules to the antibodies or in some cases, by attachment to a radioactive label (see, e.g., Tijssen, "*Practice and Theory of Enzyme Immunoassays,*" *Laboratory Techniques in Biochemistry and Molecular Biology*, Burdon and van Knippenberg Eds., Elsevier (1985), pp. 9-20).

[0095] The probes are typically labeled either directly, as with isotopes, chromophores, lumiphores, chromogens, or indirectly, such as with biotin, to which a streptavidin complex may later bind. Thus, the detectable labels used in the assays of the present invention can be primary labels (where the label comprises an element that is detected directly or that produces a directly detectable element) or secondary labels (where the detected label binds to a primary label, e.g., as is common in immunological labeling). Typically, labeled signal nucleic acids are used to detect hybridization. Complementary nucleic acids or signal nucleic acids may be labeled by any one of several methods

typically used to detect the presence of hybridized polynucleotides. The most common method of detection is the use of autoradiography with ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P-labeled probes or the like.

[0096] Other labels include, e.g., ligands that bind to labeled antibodies, fluorophores, chemiluminescent agents, enzymes, and antibodies that can serve as specific binding pair members for a labeled ligand. An introduction to labels, labeling procedures and detection of labels is found in Polak and Van Noorden *Introduction to Immunocytochemistry,* 2nd ed., Springer Verlag, N.Y. (1997); and in Haugland *Handbook of Fluorescent Probes and Research Chemicals*, a combined handbook and catalogue Published by Molecular Probes, Inc. (1996).

[0097] In general, a detector that monitors a particular probe or probe combination is used to detect the detection reagent label. Typical detectors include spectrophotometers, phototubes and photodiodes, microscopes, scintillation counters, cameras, film and the like, as well as combinations thereof. Examples of suitable detectors are widely available from a variety of commercial sources known to persons of skill in the art. Commonly, an optical image of a substrate comprising bound labeling moieties is digitized for subsequent computer analysis.

[0098] The amount of, for example, an RNA is measured by quantitating the amount of label fixed to the solid support by binding of the detection reagent. Typically, the presence of a modulator during incubation will increase or decrease the amount of label fixed to the solid support relative to a control incubation that does not comprise the modulator, or as compared to a baseline established for a particular reaction type. Means of detecting and quantitating labels are well known to those of skill in the art.

[0099] In some embodiments, the target nucleic acid or the probe is immobilized on a solid support. Solid supports suitable for use in the assays of the invention are known to those of skill in the art. As used herein, a solid support is a matrix of material in a substantially fixed arrangement.

[0100] A variety of automated solid-phase assay techniques are also appropriate. For instance, very large scale immobilized polymer arrays (VLSIPS™), i.e. Gene Chips or microarrays, available from Affymetrix, Inc. in Santa Clara, Calif. can be used to detect changes in expression levels of a plurality of genes involved in the same regulatory pathways simultaneously. See, Tijssen, supra., Fodor et al. (1991) *Science,* 251: 767-777; Sheldon et al. (1993) *Clinical Chemistry* 39(4): 718-719, and Kozal et al. (1996) *Nature Medicine* 2(7): 753-759. Similarly, spotted cDNA arrays (arrays of cDNA sequences bound to nylon, glass or another solid support) can also be used to monitor expression of a plurality of genes.

[0101] Typically, the array elements are organized in an ordered fashion so that each element is present at a specified location on the substrate. Because the array elements are at specified locations on the substrate, the hybridization patterns and intensities (which together create a unique expression profile) can be interpreted in terms of expression levels of particular genes and can be correlated with a particular disease or condition or treatment. See, e.g., Schena et al., *Science* 270: 467-470 (1995)) and (Lockhart et al., Nature Biotech. 14: 1675-1680 (1996)).

[0102] Hybridization specificity can be evaluated by comparing the hybridization of specificity-control polynucleotide sequences to specificity-control polynucleotide probes that are added to a sample in a known amount. The specificity-control target polynucleotides may have one or more sequence mismatches compared with the corresponding polynucleotide sequences. In this manner, whether only complementary target polynucleotides are hybridizing to the polynucleotide sequences or whether mismatched hybrid duplexes are forming is determined.

[0103] Hybridization reactions can be performed in absolute or differential hybridization formats. In the absolute hybridization format, polynucleotide probes from one sample are hybridized to the sequences in a microarray format and signals detected after hybridization complex formation correlate to polynucleotide probe levels in a sample. In the differential hybridization format, the differential expression of a set of genes in two biological samples is analyzed. For differential hybridization, polynucleotide probes from both biological samples are prepared and labeled with different labeling moieties. A mixture of the two labeled polynucleotide probes is added to a microarray. The microarray is then examined under conditions in which the emissions from the two different labels are individually detectable. Sequences in the microarray that are hybridized to substantially equal numbers of polynucleotide probes derived from both biological samples give a distinct combined fluorescence (Shalon et al. PCT publication WO95/35505). In some embodiments, the labels are fluorescent labels with distinguishable emission spectra, such as Cy3 and Cy5 fluorophores.

[0104] After hybridization, the microarray is washed to remove nonhybridized nucleic acids and complex formation between the hybridizable array elements and the polynucleotide probes is detected. Methods for detecting complex formation are well known to those skilled in the art. In some embodiments, the polynucleotide probes are labeled with a fluorescent label and measurement of levels and patterns of fluorescence indicative of complex formation is accomplished by fluorescence microscopy, such as confocal fluorescence microscopy.

[0105] In a differential hybridization experiment, polynucleotide probes from two or more different biological samples are labeled with two or more different fluorescent labels with different emission wavelengths. Fluorescent signals are detected separately with different photomultipliers set to detect specific wavelengths. The relative abundances/expression levels of the polynucleotide probes in two or more samples are obtained.

[0106] Typically, microarray fluorescence intensities can be normalized to take into account variations in hybridization intensities when more than one microarray is used under similar test conditions. In some embodiments, individual polynucleotide probe/target complex hybridization intensities are normalized using the intensities derived from internal normalization controls contained on each microarray.

[0107] Detection of nucleic acids can also be accomplished, for example, by using a labeled detection moiety that binds specifically to duplex nucleic acids (e.g., an antibody that is specific for RNA-DNA duplexes). One example uses an antibody that recognizes DNA-RNA heteroduplexes in which the antibody is linked to an enzyme

(typically by recombinant or covalent chemical bonding). The antibody is detected when the enzyme reacts with its substrate, producing a detectable product. Coutlee et al. (1989) *Analytical Biochemistry* 181:153-162; Bogulavski (1986) et al. *J. Immunol. Methods* 89:123-130; Prooijen-Knegt (1982) *Exp. Cell Res.* 141:397-407; Rudkin (1976) *Nature* 265:472-473, Stollar (1970) *PNAS* 65:993-1000; Ballard (1982) *Mol. Immunol.* 19:793-799; Pisetsky and Caster (1982) *Mol. Immunol.* 19:645-650; Viscidi et al. (1988) *J. Clin. Microbial.* 41:199-209; and Kiney et al. (1989) *J. Clin. Microbiol.* 27:6-12 describe antibodies to RNA duplexes, including homo and heteroduplexes. Kits comprising antibodies specific for DNA:RNA hybrids are available, e.g., from Digene Diagnostics, Inc. (Beltsville, Md.).

[0108] In addition to available antibodies, one of skill in the art can easily make antibodies specific for nucleic acid duplexes using existing techniques, or modify those antibodies that are commercially or publicly available. In addition to the art referenced above, general methods for producing polyclonal and monoclonal antibodies are known to those of skill in the art (see, e.g., Paul (ed) Fundamental Immunology, Third Edition Raven Press, Ltd., NY (1993); Coligan *Current Protocols in Immunology* Wiley/Greene, NY (1991); Harlow and Lane Antibodies: *A Laboratory Manual* Cold Spring Harbor Press, NY (1989); Stites et al. (eds.) *Basic and Clinical Immunology* (4th ed.) Lange Medical Publications, Los Altos, Calif., and references cited therein; Goding *Monoclonal Antibodies: Principles and Practice* (2d ed.) Academic Press, New York, N.Y., (1986); and Kohler and Milstein *Nature* 256: 495-497 (1975)). Other suitable techniques for antibody preparation include selection of libraries of recombinant antibodies in phage or similar vectors (see, Huse et al. *Science* 246:1275-1281 (1989); and Ward et al. *Nature* 341:544-546 (1989)). Specific monoclonal and polyclonal antibodies and antisera will usually bind with a K_D of at least about 0.1 μM, preferably at least about 0.01 μM or better, and most typically and preferably, 0.001 μM or better.

[0109] The nucleic acids used in this invention can be either positive or negative probes. Positive probes bind to their targets and the presence of duplex formation is evidence of the presence of the target. Negative probes fail to bind to the suspect target and the absence of duplex formation is evidence of the presence of the target. For example, the use of a wild type specific nucleic acid probe or PCR primers may serve as a negative probe in an assay sample where only the nucleotide sequence of interest is present.

[0110] The sensitivity of the hybridization assays may be enhanced through use of a nucleic acid amplification system that multiplies the target nucleic acid being detected. Examples of such systems include the polymerase chain reaction (PCR) system and the ligase chain reaction (LCR) system. Other methods recently described in the art are the nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario) and Q Beta Replicase systems. These systems can be used to directly identify mutants where the PCR or LCR primers are designed to be extended or ligated only when a selected sequence is present. Alternatively, the selected sequences can be generally amplified using, for example, nonspecific PCR primers and the amplified target region later probed for a specific sequence indicative of a mutation. It is understood that various detection probes, including Taqman and molecular beacon probes can be used to monitor amplification reaction products, e.g., in real time.

[0111] An alternative means for determining the level of expression of the nucleic acids of the present invention is in situ hybridization. In situ hybridization assays are well known and are generally described in Angerer et al., *Methods Enzymol.* 152:649-660 (1987). In an in situ hybridization assay, cells, preferentially human cells from the cerebellum or the hippocampus, are fixed to a solid support, typically a glass slide. If DNA is to be probed, the cells are denatured with heat or alkali. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of specific probes that are labeled. The probes are preferably labeled with radioisotopes or fluorescent reporters.

[0112] Single nucleotide polymorphism (SNP) analysis is also useful for detecting differences between alleles of the polynucleotides (e.g., genes) of the invention. SNPs linked to genes encoding polypeptides of the invention are useful, for instance, for diagnosis of diseases (e.g., diabetes) whose occurrence is linked to the gene sequences of the invention. For example, if an individual carries at least one SNP linked to a disease-associated allele of the gene sequences of the invention, the individual is likely predisposed for one or more of those diseases. If the individual is homozygous for a disease-linked SNP, the individual is particularly predisposed for occurrence of that disease (e.g., diabetes). In some embodiments, the SNP associated with the gene sequences of the invention is located within 300,000; 200,000; 100, 000; 75,000; 50,000; or 10,000 base pairs from the gene sequence.

[0113] Various real-time PCR methods including, e.g., Taqman or molecular beacon-based assays (e.g., U.S. Pat. Nos. 5,210,015; 5,487,972; Tyagi et al., *Nature Biotechnology* 14:303 (1996); and PCT WO 95/13399 are useful to monitor for the presence of absence of a SNP. Additional SNP detection methods include, e.g., DNA sequencing, sequencing by hybridization, dot blotting, oligonucleotide array (DNA Chip) hybridization analysis, or are described in, e.g., U.S. Pat. No. 6,177,249; Landegren et al., *Genome Research,* 8:769-776 (1998); Botstein et al., *Am J Human Genetics* 32:314-331 (1980); Meyers et al., Methods in Enzymology 155:501-527 (1987); Keen et al., *Trends in Genetics* 7:5 (1991); Myers et al., *Science* 230:1242-1246 (1985); and Kwok et al., *Genomics* 23:138-144 (1994).

V. Immunological Detection of Polypeptides of the Invention

[0114] In addition to the detection of polynucloetides of the invention and gene expression using nucleic acid hybridization technology, one can also use immunoassays to detect polypeptides of the invention. Immunoassays can be used to qualitatively or quantitatively analyze polypeptides of the invention. A general overview of the applicable technology can be found in Harlow & Lane, Antibodies: A Laboratory Manual (1988).

A. Antibodies to Target Proteins or Other Immunogens

[0115] Methods for producing polyclonal and monoclonal antibodies that react specifically with a protein of interest or other immunogen are known to those of skill in the art (see, e.g., Coligan, supra; and Harlow and Lane, supra; Stites et

al., supra and references cited therein; Goding, supra; and Kohler and Milstein *Nature,* 256:495-497 (1975)). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors (see, Huse et al., supra; and Ward et al., supra). For example, in order to produce antisera for use in an immunoassay, the protein of interest or an antigenic fragment thereof, is isolated as described herein. For example, a recombinant protein is produced in a transformed cell line. An inbred strain of mice or rabbits is immunized with the protein using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. Alternatively, a synthetic peptide derived from the sequences disclosed herein is conjugated to a carrier protein and used as an immunogen.

[0116] Polyclonal sera are collected and titered against the immunogen in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Polyclonal antisera with a titer of 10^4 or greater are selected and tested for their crossreactivity against proteins other than the polypeptides of the invention or even other homologous proteins from other organisms, using a competitive binding immunoassay. Specific monoclonal and polyclonal antibodies and antisera will usually bind with a K_D of at least about 0.1 mM, more usually at least about 1 μM, preferably at least about 0.1 μM or better, and most preferably, 0.01 μM or better.

[0117] A number of proteins of the invention comprising immunogens may be used to produce antibodies specifically or selectively reactive with the proteins of interest. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring protein may also be used either in pure or impure form. Synthetic peptides made using the protein sequences described herein may also be used as an immunogen for the production of antibodies to the protein. Recombinant protein can be expressed in eukaryotic or prokaryotic cells and purified as generally described supra. The product is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the protein.

[0118] Methods of production of polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified protein, is mixed with an adjuvant and animals are immunized. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to polypeptides of the invention. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired (see, Harlow and Lane, supra).

[0119] Monoclonal antibodies may be obtained using various techniques familiar to those of skill in the art. Typically, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (see, Kohler and Milstein, *Eur. J. Immunol.* 6:511-519 (1976)). Alternative methods of immortalization include, e.g., transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences that encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse et al., supra.

[0120] Once target immunogen-specific antibodies are available, the immunogen can be measured by a variety of immunoassay methods with qualitative and quantitative results available to the clinician. For a review of immunological and immunoassay procedures in general see, Stites, supra. Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in Maggio *Enzyme Immunoassay,* CRC Press, Boca Raton, Fla. (1980); Tijssen, supra; and Harlow and Lane, supra.

[0121] Immunoassays to measure target proteins in a human sample may use a polyclonal antiserum that was raised to full-length polypeptides of the invention or a fragment thereof. This antiserum is selected to have low cross-reactivity against other proteins and any such cross-reactivity is removed by immunoabsorption prior to use in the immunoassay.

B. Immunological Binding Assays

[0122] In some embodiments, a protein of interest is detected and/or quantified using any of a number of well-known immunological binding assays (see, e.g., U.S. Pat. Nos. 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also Asai *Methods in Cell Biology Volume* 37: Antibodies in Cell Biology, Academic Press, Inc. NY (1993); Stites, supra. Immunological binding assays (or immunoassays) typically utilize a "capture agent" to specifically bind to and often immobilize the analyte (e.g., full-length polypeptides of the present invention, or antigenic subsequences thereof). The capture agent is a moiety that specifically binds to the analyte. The antibody may be produced by any of a number of means well known to those of skill in the art and as described above.

[0123] Immunoassays also often utilize a labeling agent to bind specifically to and label the binding complex formed by the capture agent and the analyte. The labeling agent may itself be one of the moieties comprising the antibody/analyte complex. Alternatively, the labeling agent may be a third moiety, such as another antibody, that specifically binds to the antibody/protein complex.

[0124] In a preferred embodiment, the labeling agent is a second antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second antibody can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

[0125] Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G, can also be used as the label agents. These proteins are normal constituents of the cell walls of streptococcal bacteria. They exhibit a strong non-immunogenic

reactivity with immunoglobulin constant regions from a variety of species (see, generally, Kronval, et al. *J. Immunol.,* 111: 1401-1406 (1973); and Akerstrom, et al. *J. Immunol.,* 135:2589-2542 (1985)).

[0126] Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. The incubation time will depend upon the assay format, analyte, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10° C. to 40° C.

1. Non-Competitive Assay Formats

[0127] Immunoassays for detecting proteins or analytes of interest from tissue samples may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of captured protein or analyte is directly measured. In one preferred "sandwich" assay, for example, the capture agent (e.g., antibodies specific for the polypeptides of the invention) can be bound directly to a solid substrate where it is immobilized. These immobilized antibodies then capture the polypeptide present in the test sample. The polypeptide of the invention thus immobilized is then bound by a labeling agent, such as a second labelled antibody specific for the polypeptide. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

2. Competitive Assay Formats

[0128] In competitive assays, the amount of protein or analyte present in the sample is measured indirectly by measuring the amount of an added (exogenous) protein or analyte displaced (or competed away) from a specific capture agent (e.g., antibodies specific for a polypeptide of the invention) by the protein or analyte present in the sample. The amount of immunogen bound to the antibody is inversely proportional to the concentration of immunogen present in the sample. In a particularly preferred embodiment, the antibody is immobilized on a solid substrate. The amount of analyte may be detected by providing a labeled analyte molecule. It is understood that labels can include, e.g., radioactive labels as well as peptide or other tags that can be recognized by detection reagents such as antibodies.

[0129] Immunoassays in the competitive binding format can be used for cross-reactivity determinations. For example, the protein encoded by the sequences described herein can be immobilized on a solid support. Proteins are added to the assay and compete with the binding of the antisera to the immobilized antigen. The ability of the above proteins to compete with the binding of the antisera to the immobilized protein is compared to that of the protein encoded by any of the sequences described herein. The percent cross-reactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% cross-reactivity with each of the proteins listed above are selected and pooled. The cross-reacting antibodies are optionally removed from the pooled antisera by immunoabsorption with the considered proteins, e.g., distantly related homologs.

[0130] The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein, thought to be perhaps a protein of the present invention, to the immunogen protein. In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required is less than 10 times the amount of the protein partially encoded by a sequence herein that is required, then the second protein is said to specifically bind to an antibody generated to an immunogen consisting of the target protein.

3. Other Assay Formats

[0131] In some embodiments, western blot (immunoblot) analysis is used to detect and quantify the presence of a polypeptide of the invention in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support (such as, e.g., a nitrocellulose filter, a nylon filter, or a derivatized nylon filter) and incubating the sample with the antibodies that specifically bind the protein of interest. For example, antibodies are selected that specifically bind to the polypeptides of the invention on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (e.g., labeled sheep anti-mouse antibodies) that specifically bind to the antibodies against the protein of interest.

[0132] Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (e.g., antibodies) and release encapsulated reagents or markers. The released chemicals are then detected according to standard techniques (see, Monroe et al. (1986) *Amer. Clin. Prod. Rev.* 5:34-41).

4. Labels

[0133] The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere with the specific binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most labels useful in such methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (e.g., Dynabeads™), fluorescent dyes (e.g., fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (e.g. ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads.

[0134] The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on the sensitivity required, the ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

[0135] Non-radioactive labels are often attached by indirect means. The molecules can also be conjugated directly to signal generating compounds, e.g., by conjugation with an enzyme or fluorescent compound. A variety of enzymes and fluorescent compounds can be used with the methods of the present invention and are well-known to those of skill in the art (for a review of various labeling or signal producing systems which may be used, see, e.g., U.S. Pat. No. 4,391, 904).

[0136] Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by means of photographic film, by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Finally simple calorimetric labels may be detected directly by observing the color associated with the label. Thus, in various dipstick assays, conjugated gold often appears pink, while various conjugated beads appear the color of the bead.

[0137] Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need to be labeled and the presence of the target antibody is detected by simple visual inspection.

VI. Identification of Modulators of Polypeptides of the Invention

[0138] Modulators of a polypeptide of the invention, i.e. agonists or antagonists of a polypeptide's activity, or polypeptide's or polynucleotide's expression, are useful for treating a number of human diseases, including diabetes. For example, administration of modulators can be used to treat diabetic patients or prediabetic individuals to prevent progression, and therefore symptoms, associated with diabetes (including insulin resistance).

A. Agents that Modulate Polypeptides of the Invention

[0139] The agents tested as modulators of polypeptides of the invention can be any small chemical compound, or a biological entity, such as a protein, sugar, nucleic acid or lipid. Typically, test compounds will be small chemical molecules and peptides. Essentially any chemical compound can be used as a potential modulator or ligand in the assays of the invention, although most often compounds that can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. The assays are designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays). Modulators also include agents designed to reduce the level of mRNA encoding a polypeptide of the invention (e.g. antisense molecules, ribozymes, DNAzymes, small inhibitory RNAs and the like) or the level of translation from an mRNA (e.g., translation blockers such as an antisense molecules that are complementary to translation start or other sequences on an mRNA molecule). It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, Mo.), Aldrich (St. Louis, Mo.), Sigma-Aldrich (St. Louis, Mo.), Fluka Chemika-Biochemica Analytika (Buchs, Switzerland) and the like.

[0140] In some embodiments, high throughput screening methods involve providing a combinatorial chemical or peptide library containing a large number of potential therapeutic compounds (potential modulator compounds). Such "combinatorial chemical libraries" or "ligand libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

[0141] A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

[0142] Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Pat. No. 5,010,175, Furka, *Int. J. Pept. Prot. Res.* 37:487-493 (1991) and Houghton et al., Nature 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (e.g., PCT Publication No. WO 91/19735), encoded peptides (e.g., PCT Publication WO 93/20242), random bio-oligomers (e.g., PCT Publication No. WO 92/00091), benzodiazepines (e.g., U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs et al., *Proc. Nat. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara et al., *J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann et al., *J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen et al., *J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho et al., *Science* 261:1303 (1993)), and/or peptidyl phosphonates (Campbell et al., *J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (see Ausubel, Berger and Sambrook, all supra), peptide nucleic acid libraries (see, e.g., U.S. Pat. No. 5,539,083), antibody libraries (see, e.g., Vaughn et al., *Nature Biotechnology,* 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al., *Science,* 274:1520-1522 (1996) and U.S. Pat. No. 5,593,853), small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN, January 18, page 33 (1993); isoprenoids, U.S. Pat. No. 5,569,588; thiazolidinones and metathiazanones, U.S. Pat. No. 5,549,974; pyrrolidines, U.S. Pat. Nos. 5,525,735 and 5,519,134; morpholino compounds, U.S. Pat. No. 5,506, 337; benzodiazepines, 5,288,514, and the like).

[0143] Devices for the preparation of combinatorial libraries are commercially available (see, e.g., 357 MPS, 390 MPS, Advanced Chem Tech, Louisville Ky., Symphony, Rainin, Woburn, Mass., 433A Applied Biosystems, Foster City, Calif., 9050 Plus, Millipore, Bedford, Mass.). In addition, numerous combinatorial libraries are themselves commercially available (see, e.g., ComGenex, Princeton, N.J., Tripos, Inc., St. Louis, Mo., 3D Pharmaceuticals, Exton, Pa., Martek Biosciences, Columbia, Md., etc.).

B. Methods of Screening for Modulators of the Polypeptides of the Invention

[0144] A number of different screening protocols can be utilized to identify agents that modulate the level of expression or activity of a polynucleotide of a polypeptide of the invention in cells, particularly mammalian cells, and especially human cells. In general terms, the screening methods involve screening a plurality of agents to identify an agent that modulates the activity of a polypeptide of the invention by, e.g., binding to the polypeptide, preventing an inhibitor or activator from binding to the polypeptide, increasing association of an inhibitor or activator with the polypeptide, or activating or inhibiting expression of the polypeptide.

[0145] Any cell expressing a full-length polypeptide of the invention or a fragment thereof can be used to identify modulators. In some embodiments, the cells are eukaryotic cells lines (e.g., CHO or HEK293) transformed to express a heterologous polypeptide of the invention. In some embodiments, a cell expressing an endogenous polypeptide of the invention is used in screens. In other embodiments, modulators are screened for their ability to effect insulin responses.

1. Polypeptide Binding Assays

[0146] Preliminary screens can be conducted by screening for agents capable of binding to polypeptides of the invention, as at least some of the agents so identified are likely modulators of a polypeptide of the invention. Binding assays are also useful, e.g., for identifying endogenous proteins that interact with polypeptides of the invention. For example, antibodies, receptors or other molecules that bind polypeptides of the invention can be identified in binding assays.

[0147] Binding assays usually involve contacting a polypeptide of the invention with one or more test agents and allowing sufficient time for the protein and test agents to form a binding complex. Any binding complexes formed can be detected using any of a number of established analytical techniques. Protein binding assays include, but are not limited to, methods that measure co-precipitation or co-migration on non-denaturing SDS-polyacrylamide gels, and co-migration on Western blots (see, e.g., Bennet, J. P. and Yamamura, H. I. (1985) "Neurotransmitter, Hormone or Drug Receptor Binding Methods," in *Neurotransmitter Receptor Binding* (Yamamura, H. I., et al., eds.), pp. 61-89. Other binding assays involve the use of mass spectrometry or NMR techniques to identify molecules bound to a polypeptide of the invention or displacement of labeled substrates. The polypeptides of the invention utilized in such assays can be naturally expressed, cloned or synthesized.

[0148] In addition, mammalian or yeast two-hybrid approaches (see, e.g., Bartel, P. L. et. al. *Methods Enzymol,* 254:241 (1995)) can be used to identify polypeptides or other molecules that interact or bind when expressed together in a host cell.

2. Polypeptide Activity

[0149] The activity of polypeptides of the invention can be assessed using a variety of in vitro and in vivo assays to determine functional, chemical, and physical effects, e.g., measuring ligand binding (e.g., radioactive or otherwise labeled ligand binding), second messengers (e.g., cAMP, cGMP, IP_3, DAG, or Ca^{2+}), ion flux, phosphorylation levels, transcription levels, and the like. Furthermore, such assays can be used to test for inhibitors and activators of the polypeptides of the invention. Modulators can also be genetically altered versions of polypeptides of the invention.

[0150] The polypeptide of the assay will be selected from a polypeptide with substantial identity to a sequence described herein, e.g., SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34, or other conservatively modified variants thereof. Generally, the amino acid sequence identity will be at least 70%, optionally at least 85%, optionally at least 90-95% to the polypeptides exemplified herein. Optionally, the polypeptide of the assays will comprise a fragment of a polypeptide of the invention, such as an extracellular domain, transmembrane domain, cytoplasmic domain, ligand binding domain, subunit association domain, active site, and the like. Either a polypeptide of the invention or a domain thereof can be covalently linked to a heterologous protein to create a chimeric protein used in the assays described herein.

[0151] Modulators of polypeptide activity are tested using either recombinant or naturally occurring polypeptides of the invention. The protein can be isolated, expressed in a cell, expressed in a membrane derived from a cell, expressed in tissue or in an animal, either recombinant or naturally occurring. For example, tissue slices, dissociated cells, e.g., from tissues expressing polypeptides of the invention, transformed cells, or membranes can be used. Modulation is tested using one of the in vitro or in vivo assays described herein.

[0152] Modulator binding to polypeptides of the invention, a domain, or chimeric protein can be tested in solution, in a bilayer membrane, attached to a solid phase, in a lipid monolayer, or in vesicles. Binding of a modulator can be tested using, e.g., changes in spectroscopic characteristics (e.g., fluorescence, absorbance, refractive index), hydrodynamic (e.g., shape), chromatographic, or solubility properties.

[0153] Samples or assays that are treated with a potential moulator (e.g., a "test compound") are compared to control samples without the test compound, to examine the extent of modulation. Control samples (untreated with activators or inhibitors) are assigned a relative activity value of 100. Inhibition of the polypeptides of the invention is achieved when the activity value relative to the control is about 90%, optionally 50%, optionally 25-0%. Activation of the polypeptides of the invention is achieved when the activity value relative to the control is 110%, optionally 150%, 200%, 300%, 400%, 500%, or 1000-2000%.

3. Expression Assays

[0154] Screening for a compound that modulates the expression of a polynucleotide or a polypeptide of the invention is also provided. Screening methods generally

involve conducting cell-based assays in which test compounds are contacted with one or more cells expressing a polynucleotide or a polypeptide of the invention, and then detecting an increase or decrease in expression (either transcript or translation product). Assays can be performed with any cells that express a polynucleotide or a polypeptide of the invention.

[0155] Expression can be detected in a number of different ways. As described infra, the expression level of a polynucleotide of the invention in a cell can be determined by probing the mRNA expressed in a cell with a probe that specifically hybridizes with a transcript (or complementary nucleic acid derived therefrom) of a polynucleotide of the invention. Probing can be conducted by lysing the cells and conducting Northern blots or without lysing the cells using in situ-hybridization techniques. Alternatively, a polypeptide of the invention can be detected using immunological methods in which a cell lysate is probed with antibodies that specifically bind to the polypeptide.

[0156] The level of expression or activity of a polynucleotide or a polypeptide of the invention can be compared to a baseline value. The baseline value can be a value for a control sample or a statistical value that is representative of expression levels of a polynucleotide or a polypeptide of the invention for a control population (e.g., lean individuals as described herein) or cells (e.g., tissue culture cells not exposed to a modulator). Expression levels can also be determined for cells that do not express the polynucleotide or a polypeptide of the invention as a negative control. Such cells generally are otherwise substantially genetically the same as the test cells.

[0157] A variety of different types of cells can be utilized in the reporter assays. Cells that do not endogenously express a polypeptide of the invention can be prokaryotic, but are preferably eukaryotic. The eukaryotic cells can be any of the cells typically utilized in generating cells that harbor recombinant nucleic acid constructs. Exemplary eukaryotic cells include, but are not limited to, yeast, and various higher eukaryotic cells such as the HEK293, HepG2, COS, CHO and HeLa cell lines.

[0158] Various controls can be conducted to ensure that an observed activity is authentic including running parallel reactions with cells that lack the reporter construct or by not contacting a cell harboring the reporter construct with test compound. Compounds can also be further validated as described below.

4. Validation

[0159] Agents that are initially identified by any of the foregoing screening methods can be further tested to validate the apparent activity. Modulators that are selected for further study can be tested on the "classic" insulin responsive cell line, mouse 3T3-L1 adipocytes, muscle cells such as L6 cells and the like. Cells (e.g., adipocytes or muscle cells) are pre-incubated with the modulators and tested for acute (up to 4 hours) and chronic (overnight) effects on basal and insulin-stimulated GLUT4 translocation and glucose uptake.

[0160] Following such studies, validity of the modulators is tested in suitable animal models. The basic format of such methods involves administering a lead compound identified during an initial screen to an animal that serves as a model for humans and then determining if expression of activity of a polypeptide of the invention is in fact modulated.

[0161] The effect of the compound will be asssesed in either diabetic animals or in diet induced insulin resistant animals. The blood glucose and insulin levels will be determined. The animal models utilized in validation studies generally are mammals of any kind. Specific examples of suitable animals include, but are not limited to, primates, mice and rats. For example, monogenic models of diabetes (e.g., ob/ob and db/db mice, Zucker rats and Zucker Diabetic Fatty rats etc) or polygenic models of diabetes (e.g., OLETF rats, GK rats, NSY mice, and KK mice) can be useful for validating modulation of a polypeptide of the invention in a diabetic or insulin resistant animal. In addition, transgenic animals expressing human polypeptides of the invention can be used to further validate drug candidates.

C. Solid Phase and Soluble High Throughput Assays

[0162] In the high throughput assays of the invention, it is possible to screen up to several thousand different modulators or ligands in a single day. In particular, each well of a microtiter plate can be used to run a separate assay against a selected potential modulator, or, if concentration or incubation time effects are to be observed, every 5-10 wells can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (e.g., 96) modulators. If 1536 well plates are used, then a single plate can easily assay from about 100 to about 1500 different compounds. It is possible to assay several different plates per day; assay screens for up to about 6,000-20,000 or more different compounds are possible using the integrated systems of the invention. In addition, microfluidic approaches to reagent manipulation can be used.

[0163] A molecule of interest (e.g., a polypeptide or polynucleotide of the invention, or a modulator thereof) can be bound to the solid-state component, directly or indirectly, via covalent or non-covalent linkage, e.g., via a tag. The tag can be any of a variety of components. In general, a molecule that binds the tag (a tag binder) is fixed to a solid support, and the tagged molecule of interest is attached to the solid support by interaction of the tag and the tag binder.

[0164] A number of tags and tag binders can be used, based upon known molecular interactions well described in the literature. For example, where a tag has a natural binder, for example, biotin, protein A, or protein G, it can be used in conjunction with appropriate tag binders (avidin, streptavidin, neutravidin, the Fc region of an immunoglobulin, poly-His, etc.) Antibodies to molecules with natural binders such as biotin are also widely available and appropriate tag binders (see, SIGMA Immunochemicals 1998 catalogue SIGMA, St. Louis Mo.).

[0165] Similarly, any haptenic or antigenic compound can be used in combination with an appropriate antibody to form a tag/tag binder pair. Thousands of specific antibodies are commercially available and many additional antibodies are described in the literature. For example, in one common configuration, the tag is a first antibody and the tag binder is a second antibody that recognizes the first antibody. In addition to antibody-antigen interactions, receptor-ligand interactions are also appropriate as tag and tag-binder pairs, such as agonists and antagonists of cell membrane receptors (e.g., cell receptor-ligand interactions such as transferrin,

c-kit, viral receptor ligands, cytokine receptors, chemoline receptors, interleukin receptors, immunoglobulin receptors and antibodies, the cadherin family, the integrin family, the selectin family, and the like; see, e.g., Pigott & Power, *The Adhesion Molecule Facts Book I* (1993)). Similarly, toxins and venoms, viral epitopes, hormones (e.g., opiates, steroids, etc.), intracellular receptors (e.g., which mediate the effects of various small ligands, including steroids, thyroid hormone, retinoids and vitamin D; peptides), drugs, lectins, sugars, nucleic acids (both linear and cyclic polymer configurations), oligosaccharides, proteins, phospholipids and antibodies can all interact with various cell receptors.

[0166] Synthetic polymers, such as polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, and polyacetates can also form an appropriate tag or tag binder. Many other tag/tag binder pairs are also useful in assay systems described herein, as would be apparent to one of skill upon review of this disclosure.

[0167] Common linkers such as peptides, polyethers, and the like can also serve as tags, and include polypeptide sequences, such as poly-gly sequences of between about 5 and 200 amino acids. Such flexible linkers are known to those of skill in the art. For example, poly(ethelyne glycol) linkers are available from Shearwater Polymers, Inc., Huntsville, Ala. These linkers optionally have amide linkages, sulfhydryl linkages, or heterofunctional linkages.

[0168] Tag binders are fixed to solid substrates using any of a variety of methods currently available. Solid substrates are commonly derivatized or functionalized by exposing all or a portion of the substrate to a chemical reagent that fixes a chemical group to the surface that is reactive with a portion of the tag binder. For example, groups that are suitable for attachment to a longer chain portion would include amines, hydroxyl, thiol, and carboxyl groups. Aminoalkylsilanes and hydroxyalkylsilanes can be used to functionalize a variety of surfaces, such as glass surfaces. The construction of such solid phase biopolymer arrays is well described in the literature (see, e.g., Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963) (describing solid phase synthesis of, e.g., peptides); Geysen et al, *J. Immun. Meth.* 102:259-274 (1987) (describing synthesis of solid phase components on pins); Frank and Doring, *Tetrahedron* 44:60316040 (1988) (describing synthesis of various peptide sequences on cellulose disks); Fodor et al., *Science,* 251:767-777 (1991); Sheldon et al., *Clinical Chemistry* 39(4):718-719 (1993); and Kozal et al., *Nature Medicine* 2(7):753759 (1996) (all describing arrays of biopolymers fixed to solid substrates). Non-chemical approaches for fixing tag binders to substrates include other common methods, such as heat, cross-linking by UV radiation, and the like.

[0169] The invention provides in vitro assays for identifying, in a high throughput format, compounds that can modulate the expression or activity of a polypeptide of the invention. Control reactions that measure activity of a polypeptide of the invention in a cell in a reaction that does not include a potential modulator are optional, as the assays are highly uniform. Such optional control reactions are appropriate and increase the reliability of the assay. Accordingly, in some embodiments, the methods of the invention include such a control reaction. For each of the assay formats described, "no modulator" control reactions that do not include a modulator provide a background level of binding activity.

[0170] In some assays it will be desirable to have positive controls. At least two types of positive controls are appropriate. First, a known activator of a polypeptide or a polynucleotide of the invention can be incubated with one sample of the assay, and the resulting increase in signal resulting from an increased expression level or activity of a polypeptide or a polynucleotide of the invention are determined according to the methods herein. Second, a known inhibitor of a polypeptide or a polynucleotide of the invention can be added, and the resulting decrease in signal for the expression or activity of a polypeptide or a polynucleotide of the invention can be similarly detected. It will be appreciated that modulators can also be combined with activators or inhibitors to find modulators that inhibit the increase or decrease that is otherwise caused by the presence of the known modulator of a polypeptide or a polynucleotide of the invention.

VII. Compositions, Kits and Integrated Systems

[0171] The invention provides compositions, kits and integrated systems for practicing the assays described herein using nucleic acids or polypeptides of the invention, antibodies, etc.

[0172] The invention provides assay compositions for use in solid phase assays; such compositions can include, for example, one or more nucleic acids encoding a polypeptide of the invention immobilized on a solid support, and a labeling reagent. In each case, the assay compositions can also include additional reagents that are desirable for hybridization. Modulators of expression or activity of a polypeptide of the invention can also be included in the assay compositions.

[0173] The invention also provides kits for carrying out the assays of the invention. The kits typically include a probe that comprises an antibody that specifically binds to a polypeptide of the invention or a polynucleotide sequence encoding such polypeptides, and a label for detecting the presence of the probe. The kits may include at least one polynucleotide sequence encoding a polypeptide of the invention. Kits can include any of the compositions noted above, and optionally further include additional components such as instructions to practice a high-throughput method of assaying for an effect on expression of the genes encoding a polypeptide of the invention, or on activity of a polypeptide of the invention, one or more containers or compartments (e.g., to hold the probe, labels, or the like), a control modulator of the expression or activity of a polypeptide of the invention, a robotic armature for mixing kit components or the like.

[0174] The invention also provides integrated systems for high-throughput screening of potential modulators for an effect on the expression or activity of a polypeptide of the invention. The systems can include a robotic armature which transfers fluid from a source to a destination, a controller which controls the robotic armature, a label detector, a data storage unit which records label detection, and an assay component such as a microtiter dish comprising a well having a reaction mixture or a substrate comprising a fixed nucleic acid or immobilization moiety.

[0175] A number of robotic fluid transfer systems are available, or can easily be made from existing components. For example, a Zymate XP (Zymark Corporation; Hopkinton, Mass.) automated robot using a Microlab 2200 (Hamilton; Reno, Nev.) pipetting station can be used to transfer parallel samples to 96 well microtiter plates to set up several parallel simultaneous binding assays.

[0176] Optical images viewed (and, optionally, recorded) by a camera or other recording device (e.g., a photodiode and data storage device) are optionally further processed in any of the embodiments herein, e.g., by digitizing the image and storing and analyzing the image on a computer. A variety of commercially available peripheral equipment and software is available for digitizing, storing and analyzing a digitized video or digitized optical image.

[0177] One conventional system carries light from the specimen field to a cooled charge-coupled device (CCD) camera, in common use in the art. A CCD camera includes an array of picture elements (pixels). The light from the specimen is imaged on the CCD. Particular pixels corresponding to regions of the specimen (e.g., individual hybridization sites on an array of biological polymers) are sampled to obtain light intensity readings for each position. Multiple pixels are processed in parallel to increase speed. The apparatus and methods of the invention are easily used for viewing any sample, e.g., by fluorescent or dark field microscopic techniques.

VIII. Administration and Pharmaceutical Compositions

[0178] Modulators of the polypeptides of the invention (e.g., antagonists or agonists) can be administered directly to the mammalian subject for modulation of activity of a polypeptide of the invention in vivo. Administration is by any of the routes normally used for introducing a modulator compound into ultimate contact with the tissue to be treated and is well known to those of skill in the art. Although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0179] The pharmaceutical compositions of the invention may comprise a pharmaceutically acceptable carrier. Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there are a wide variety of suitable formulations of pharmaceutical compositions of the present invention (see, e.g., *Remington's Pharmaceutical Sciences,* 17th ed. 1985)).

[0180] The modulators (e.g., agonists or antagonists) of the expression or activity of the a polypeptide of the invention, alone or in combination with other suitable components, can be prepared for injection or for use in a pump device. Pump devices (also known as "insulin pumps") are commonly used to administer insulin to patients and therefore can be easily adapted to include compositions of the present invention. Manufacturers of insulin pumps include Animas, Disetronic and MiniMed.

[0181] The modulators (e.g., agonists or antagonists) of the expression or activity of a polypeptide of the invention, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

[0182] Formulations suitable for administration include aqueous and non-aqueous solutions, isotonic sterile solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, orally, nasally, topically, intravenously, intraperitoneally, or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampoules and vials. Solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. The modulators can also be administered as part of a prepared food or drug.

[0183] The dose administered to a patient, in the context of the present invention should be sufficient to induce a beneficial response in the subject over time. The optimal dose level for any patient will depend on a variety of factors including the efficacy of the specific modulator employed, the age, body weight, physical activity, and diet of the patient, on a possible combination with other drugs, and on the severity of the case of diabetes. It is recommended that the daily dosage of the modulator be determined for each individual patient by those skilled in the art in a similar way as for known insulin compositions. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular subject.

[0184] In determining the effective amount of the modulator to be administered a physician may evaluate circulating plasma levels of the modulator, modulator toxicity, and the production of anti-modulator antibodies. In general, the dose equivalent of a modulator is from about 1 ng/kg to 10 mg/kg for a typical subject.

[0185] For administration, modulators of the present invention can be administered at a rate determined by the LD-50 of the modulator, and the side-effects of the modulator at various concentrations, as applied to the mass and overall health of the subject. Administration can be accomplished via single or divided doses.

[0186] The compounds of the present invention can also be used effectively in combination with one or more additional active agents depending on the desired target therapy (see, e.g., Turner, N. et al. *Prog. Drug Res.* (1998) 51: 33-94; Haffner, S. *Diabetes Care* (1998) 21: 160-178; and DeFronzo, R. et al. (eds.), *Diabetes Reviews* (1997) Vol. 5 No. 4). A number of studies have investigated the benefits of combination therapies with oral agents (see, e.g., Mahler, R., *J. Clin. Endocrinol. Metab.* (1999) 84: 1165-71; United Kingdom Prospective Diabetes Study Group: UKPDS 28, *Diabetes Care* (1998) 21: 87-92; Bardin, C. W., (ed.), *Current Therapy In Endocrinology And Metabolism,* 6th Edition (Mosby—Year Book, Inc., St. Louis, Mo. 1997); Chiasson, J. et al., *Ann. Intern. Med.* (1994) 121: 928-935; Coniff, R. et al., *Clin. Ther.* (1997) 19: 16-26; Coniff, R. et al., *Am. J. Med.* (1995) 98: 443-451; and Iwamoto, Y. et al., *Diabet. Med.* (1996) 13 365-370; Kwiterovich, P. *Am. J. Cardiol* (1998) 82(12A): 3U-17U). These studies indicate

that modulation of diabetes, among other diseases, can be further improved by the addition of a second agent to the therapeutic regimen. Combination therapy includes administration of a single pharmaceutical dosage formulation that contains a modulator of the invention and one or more additional active agents, as well as administration of a modulator and each active agent in its own separate pharmaceutical dosage formulation. For example, a modulator and a thiazolidinedione can be administered to the human subject together in a single oral dosage composition, such as a tablet or capsule, or each agent can be administered in separate oral dosage formulations. Where separate dosage formulations are used, a modulator and one or more additional active agents can be administered at essentially the same time (i.e., concurrently), or at separately staggered times (i.e., sequentially). Combination therapy is understood to include all these regimens.

[0187] One example of combination therapy can be seen in treating pre-diabetic individuals (e.g., to prevent progression into type 2 diabetes) or diabetic individuals (or treating diabetes and its related symptoms, complications, and disorders), wherein the modulators can be effectively used in combination with, for example, sulfonylureas (such as chlorpropamide, tolbutamide, acetohexamide, tolazamide, glyburide, gliclazide, glynase, glimepiride, and glipizide); biguanides (such as metformin); a PPAR beta delta agonist; a ligand or agonist of PPAR gamma such as thiazolidinediones (such as ciglitazone, pioglitazone (see, e.g., U.S. Pat. No. 6,218,409), troglitazone, and rosiglitazone (see, e.g., U.S. Pat. No. 5,859,037)); PPAR alpha agonists such as clofibrate, gemfibrozil, fenofibrate, ciprofibrate, and bezafibrate; dehydroepiandrosterone (also referred to as DHEA or its conjugated sulphate ester, DHEA-SO4); antiglucocorticoids; TNFU inhibitors; α-glucosidase inhibitors (such as acarbose, miglitol, and voglibose); amylin and amylin derivatives (such as pramlintide, (see, also, U.S. Pat. Nos. 5,902,726; 5,124,314; 5,175,145 and 6,143,718)); insulin secretogogues (such as repaglinide, gliquidone, and nateglinide (see, also, U.S. Pat. Nos. 6,251,856; 6,251,865; 6,221,633; 6,174,856)), and insulin.

IX. Gene Therapy

[0188] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding engineered polypeptides of the invention in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding polypeptides of the invention to cells in vitro. In some embodiments, the nucleic acids encoding polypeptides of the invention are administered for in vivo or ex vivo gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, Science 256:808-813 (1992); Nabel & Felgner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada et al., in *Current Topics in Microbiology and Immunology* Doerfler and Böhm (eds) (1995); and Yu et al., *Gene Therapy* 1:13-26 (1994).

[0189] Methods of non-viral delivery of nucleic acids encoding engineered polypeptides of the invention include lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. No. 5,049,386, U.S. Pat. No. 4,946,787; and U.S. Pat. No. 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024. Delivery can be to cells (ex vivo administration) or target tissues (in vivo administration).

[0190] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, *Science* 270:404-410 (1995); Blaese et al., *Cancer Gene Ther.* 2:291-297 (1995); Behr et al., *Bioconjugate Chem.* 5:382-389 (1994); Remy et al., *Bioconjugate Chem.* 5:647-654 (1994); Gao et al., *Gene Therapy* 2:710-722 (1995); Ahmad et al., *Cancer Res.* 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[0191] The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding engineered polypeptides of the invention take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro and the modified cells are administered to patients (ex vivo). Conventional viral based systems for the delivery of polypeptides of the invention could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Viral vectors are currently the most efficient and versatile method of gene transfer in target cells and tissues. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0192] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., *J. Virol.* 66:2731-2739 (1992); Johann et al., *J. Virol.* 66:1635-1640 (1992); Sommerfelt et al., *Virol.* 176:58-59 (1990); Wilson et al., *J. Virol.* 63:2374-2378 (1989); Miller et al., *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

[0193] In applications where transient expression of the polypeptides of the invention is preferred, adenoviral based

systems are typically used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., *Virology* 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994)). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, et al., *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski et al., *J. Virol.* 63:03822-3828 (1989).

[0194] pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials (Dunbar et al., *Blood* 85:3048-305 (1995); Kohn et al., *Nat. Med.* 1:1017-102 (1995); Malech et al., *PNAS* 94:22 12133-12138 (1997)). PA317/pLASN was the first therapeutic vector used in a gene therapy trial. (Blaese et al., Science 270:475-480 (1995)). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors. (Ellem et al., *Immunol Immunother.* 44(1):10-20 (1997); Dranoff et al., *Hum. Gene Ther.* 1:111-2 (1997).

[0195] Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. All vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. (Wagner et al., *Lancet* 351:9117 1702-3 (1998), Kearns et al., *Gene Ther.* 9:748-55 (1996)).

[0196] Replication-deficient recombinant adenoviral vectors (Ad) can be engineered such that a transgene replaces the Ad E1a, E1b, and E3 genes; subsequently the replication defector vector is propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiply types of tissues in vivo, including nondividing, differentiated cells such as those found in the liver, kidney and muscle system tissues. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman et al., Hum. Gene Ther. 7:1083-9 (1998)). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker et al., *Infection* 24:1 5-10 (1996); Sterman et al., Hum. Gene Ther. 9:7 1083-1089 (1998); Welsh et al., *Hum. Gene Ther.* 2:205-18 (1995); Alvarez et al., *Hum. Gene Ther.* 5:597-613 (1997); Topf et al., *Gene Ther.* 5:507-513 (1998); Sterman et al., *Hum. Gene Ther.* 7:1083-1089 (1998).

[0197] Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

[0198] In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. A viral vector is typically modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han et al., *PNAS* 92:9747-9751 (1995), reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., FAB or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

[0199] Gene therapy vectors can be delivered in vivo by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells ex vivo, such as cells explanted from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

[0200] Ex vivo cell transfection for diagnostics, research, or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In some embodiments, cells are isolated from the subject organism, transfected with a nucleic acid (gene or cDNA) encoding a polypeptides of the invention, and re-infused back into the subject organism (e.g., patient). Various cell types suitable for ex vivo transfection are well known to those of skill in the art (see, e.g., Freshney et al., *Culture of Animal Cells, A Manual of Basic Technique* (3rd ed. 1994)) and the references cited therein for a discussion of how to isolate and culture cells from patients).

[0201] In one embodiment, stem cells are used in ex vivo procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types in vitro, or can be introduced into a mammal (such as the donor of the cells) where they will

engraft in the bone marrow. Methods for differentiating CD34+ cells in vitro into clinically important immune cell types using cytokines such a GM-CSF, IFN-γ and TNF-α are known (see Inaba et al., *J. Exp. Med.* 176:1693-1702 (1992)).

[0202] Stem cells are isolated for transduction and differentiation using known methods. For example, stem cells are isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells), GR-1 (granulocytes), and Iad (differentiated antigen presenting cells) (see Inaba et al., *J. Exp. Med.* 176:1693-1702 (1992)).

[0203] Vectors (e.g., retroviruses, adenoviruses, liposomes, etc.) containing therapeutic nucleic acids can be also administered directly to the organism for transduction of cells in vivo. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0204] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention, as described below (see, e.g., *Remington's Pharmaceutical Sciences,* 17th ed., 1989).

X. Diagnosis of Diabetes

[0205] The present invention also provides methods of diagnosing diabetes or a predisposition of at least some of the pathologies of diabetes. Diagnosis can involve determination of a genotype of an individual (e.g., with SNPs) and comparison of the genotype with alleles known to have an association with the occurrence of diabetes. Alternatively, diagnosis also involves determining the level of a polypeptide or polynucleotide of the invention in a patient and then comparing the level to a baseline or range. Typically, the baseline value is representative of a polypeptide or polynucleotide of the invention in a healthy (e.g., lean) person.

[0206] As discussed above, variation of levels (e.g., low or high levels) of a polypeptide or polynucleotide of the invention compared to the baseline range indicates that the patient is either diabetic or at risk of developing at least some of the pathologies of diabetes (e.g., pre-diabetic). The level of a polypeptide in a lean individual can be a reading from a single individual, but is typically a statistically relevant average from a group of lean individuals. The level of a polypeptide in a lean individual can be represented by a value, for example in a computer program.

[0207] In some embodiments, the level of polypeptide or polynucleotide of the invention is measured by taking a blood, urine or tissue sample from a patient and measuring the amount of a polypeptide or polynucleotide of the invention in the sample using any number of detection methods, such as those discussed herein. For instance, fasting and fed blood or urine levels can be tested.

[0208] In some embodiments, the baseline level and the level in a lean sample from an individual, or at least two samples from the same individual differ by at least about 5%, 10%, 20%, 50%, 75%, 100%, 150%, 200%, 300%, 400%, 500%, 1000% or more. In some embodiments, the sample from the individual is greater by at least one of the above-listed percentages relative to the baseline level. In some embodiments, the sample from the individual is lower by at least one of the above-listed percentages relative to the baseline level.

[0209] In some embodiments, the level of a polypeptide or polynucleotide of the invention is used to monitor the effectiveness of antidiabetic therapies such as thiazolidinediones, metformin, sulfonylureas and other standard therapies. In some embodiments the activity or expression of a polypeptide or polynucleotide of the invention will be measured prior to and after treatment of diabetic or pre-diabetic patients with antidiabetic therapies as a surrogate marker of clinical effectiveness. For example, the greater the reduction in expression or activity of a polypeptide of the invention indicates greater effectiveness.

[0210] Glucose/insulin tolerance tests can also be used to detect the effect of glucose levels on levels of a polypeptide or polynucleotide of the invention. In glucose tolerance tests, the patient's ability to tolerate a standard oral glucose load is evaluated by assessing serum and urine specimens for glucose levels. Blood samples are taken before the glucose is ingested, glucose is given by mouth, and blood or urine glucose levels are tested at set intervals after glucose ingestion. Similarly, meal tolerance tests can also be used to detect the effect of insulin or food, respectively, on levels of a polypeptide or polynucleotide of the invention.

[0211] All publications, accession numbers, and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

[0212] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

EXAMPLES

[0213] The following examples are offered to illustrate, but not to limit the claimed invention.

[0214] In obese insulin-resistant or type II diabetic individuals, peripheral tissues especially muscle and fat are known to have a moderately impaired ability to respond to insulin and hence to take up glucose. This defect in glucose metabolism is usually compensated for by increased secretion of insulin from the pancreas, thereby maintaining normal glucose levels. The majority of glucose disposal occurs in the muscle. A number of obese insulin-resistant patients will progress to overt diabetics over time. The molecular defects underlying this peripheral insulin resistance in obese individuals are not well defined. Genes in muscle or fat that exhibit altered expression in obese individuals when compared to lean individuals can be causative genes for insulin resistance and are also able to predict the transition to diabetes. Modulators of such genes can reduce or reverse insulin resistance and increase or restore insulin sensitivity to normal, thereby improving whole body glucose homeostasis, including; for example; insulin secretion. Modulators

of such genes can also be used to pre-empt the transition from obesity-induced insulin resistance to diabetes. For these reasons, gene expression profiling was performed in muscle from lean and obese individuals.

[0215] The molecular mechanism by which thiazolidinediones (TZDs) cause an increase in peripheral insulin sensitivity was studied. Genes in muscle or fat whose expression is altered by TZDs may lie on a pathway leading from TZD treatment to increased insulin sensitivity. Modulators of such genes can elicit the same effect as TZD treatment. Moreover, such modulators can lack some of the side effects of TZD. The majority of glucose disposal occurs in muscle. For this reason, gene expression profiling in human muscle from diabetics treated with troglitazone was used to identify genes important for TZD action and therefore treatment of diabetes and insulin resistance.

[0216] Gene expression profiling was performed on tissue samples (muscle) obtained from lean obese and diabetic individuals. Two studies were performed. In the first study, basal samples were isolated from all individuals at the beginning of a 5 hour hyperinsulinemic euglycemic clamp. Clamp samples were isolated at the end of this procedure. Similar basal and clamp samples were taken 3 months later after all patients had taken the insulin sensitizing drug troglitazone (tro).

[0217] In the second study, samples were obtained from lean, obese, and diabetic individuals before and after a hyperinsulinemic euglycemic clamp. No troglitazone treatment was used. For all tissue samples mRNA was isolated from these muscle samples and converted to cRNA by standard procedures. The gene expression profile for each individual was determined by hybridization of cRNA to custom synthesized Affymetrix chips.

[0218] Gene expression profile differences were calculated as follows. The expression level of a particular gene is indicated by its 'average difference score'. The raw data was analyzed by a statistical test to remove 'outliers'. The mean 'average difference score' was then calculated from the average difference scores for all individuals in a particular treatment group. Genes were determined to be changed in three different comparisons by calculating the Students t test statistic between two conditions and selecting those with t less than or equal to 0.05. Fold change was determined as the ratio of mean average difference score in condition 2 to the mean average difference score in condition 1. The first comparison is basal leans (condition 1) versus basal obese (condition 2). The second comparison is basal leans (condition 1) versus basal diabetics (condition). The third comprison is basal diabetics before troglitazone treatment (condition 1) versus basal diabetics after troglitazone treatment (condition 2).

[0219] The gene expression changes observed in the human muscle samples derived from the various comparison populations analyzed in the profiling analysis described above were further confirmed by the use of real-time quantitative PCR. A combination of specific PCR primers and a Taqman probe were designed to detect and quantify expression levels for each gene. Relative gene expression levels were determined in this manner in each patient and the mean expression levels in the comparison populations were then calculated.

Fritz

[0220] Probe set MBXHUMMUS12164 detects Fritz nucleic acid sequences. Expression of Fritz transcripts was decreased in obese compared to lean patients in the gene profiling analysis.

	Lean Pre Trog			Obese Pre Trog					
B/C	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene name
B	725	186	7	168	45	7	0.23	0.024	Fritz

Legend:
B/C indicates sample is from Basal or Clamp;
"Pre-Trog" indicates sample was taken before 3 months of Troglitazone treatment;
"Mean Expr" indicates mean expression;
"SEM" indicates standard error of mean;
"n" indicates number of patient samples;
"Fold Change" indicates fold change of obese in comparison to lean patients.

[0221] Expression of Fritz was also evaluated using real time PCR. The results further show that Fritz expression is significantly reduced in muscle from obese individuals when compared to muscle from lean individuals.

Comparison	Expression Fold change	t test
Obese (7)/Lean (7)	0.18	0.019

Legend
"Fold change" indicates fold change in Fritz expression calculated as the ratio of mean obese expression/mean lean expression.
Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

[0222] Fritz contains the following protein domains (designated with reference to SEQ ID NO:2):

Fz domain encoded by amino acids 25-148,

NTR/C345C module encoded by amino acids 186-295.

PAK1B

[0223] Probe set MBXHUUS04063 detects p21 activated kinase 1B (PAK1B) nucleic acid sequences. Expression of PAK1B transcripts was increased in obese compared to lean patients in the gene profiling analysis.

	Lean Pre-Trog			Obese Pre-Trog					
B/C	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene name
B	430	46	17	659	87	16	1.53	0.030	p21 activated kinase 1B (PAK1B)

Legend:
B/C indicates sample is from Basal or Clamp;
"Pre-Trog" indicates sample was taken before 3 months of Troglitazone treatment;
"Mean Expr" indicates mean expression;
"SEM" indicates standard error of mean;
"n" indicates number of patient samples;
"Fold Change" indicates fold change of obese in comparison to lean patients.

[0224] PAK1B expression was also evaluated by real-time PCR. The results show that PAK1B is significantly over-

expressed in muscle from obese individuals when compared to muscle from lean individuals.

Comparison	Expression Fold change	t test
Obese (16)/Lean (17)	1.49	0.009

Legend
"Fold change" indicates fold change in PAK1B expression calculated as the ratio of mean obese expression/mean lean expression.
Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

[0225] PAK1B and PAK1B splice variant contain the following protein domains (designated with reference to SEQ ID NO:6 and SEQ ID NO:8, respectively):

[0226] PBD domain encoded by amino acids 75 to 135,

[0227] Serine/threonine protein kinase domain encoded by amino acids 270 to 521.

[0228] PAK1B new splice variant contains the following protein domains (designated with reference to SEQ ID NO:10):

[0229] PBD domain encoded by amino acids 74 to 132.

[0230] Serine/threonine protein kinase domain encoded by amino acids 189 to 425.

Serine Protease, Umbilical Vein Endothelium (SPUVE)

[0231] Probe set MBXHUMMUS14656 detects SPUVE nucleic acid sequences. Expression of SPUVE transcripts was increased in obese compared to lean patients in the gene profiling analysis.

	Lean Pre-Trog			Obese Pre-Trog					
B/C	Mean Expr	SEM	n	Mean Exp	SEM	n	Fold Change	Students t test	Gene name
B	502	35	7	905	149	7	1.80	0.036	SPUVE

Legend:
B/C indicates sample is from Basal or Clamp;
"Pre-Trog" indicates sample was taken before 3 months of Troglitazone treatment;
"Mean Expr" indicates mean expression;
"SEM" indicates standard error of mean;
"n" indicates number of patient samples;
"Fold Change" indicates fold change of obese in comparison to lean patients.

[0232] SPUVE expression was also evaluated using real time PCR. The results further show that SPUVE is significantly over-expressed in muscle from obese individuals when compared to muscle from lean individuals.

Comparison	Expression Fold change	t test
Obese (7)/Lean (7)	1.55	0.145

Legend
"Fold change" indicates fold change in SPUVE expression calculated as the ratio of mean obese expression/mean lean expression.
Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

[0233] SPUVE contains the following protein domain (designated with reference to SEQ ID NO:16): trypsin domain encoded by amino acids 147-179.

NK4

[0234] Probe set MBXU S32474 detects NK4 nucleic acid sequences. Expression of transcripts encoding NK4 was increased in obese compared to lean patients in the gene profiling analysis.

	Lean Pre-Trog			Obese Pre-Tro					
B/C	Mean Exp	SEM	n	Mean Exp	SEM	n	Fold Change	Students t test	Gene name
B	619	232	7	2020	419	7	3.26	0.017	NK4

Legend: B/C indicates sample is from Basal or Clamp;
"Pre-Trog" indicates sample was taken before 3 months of Troglitazone treatment;
"Mean Expr" indicates mean expression;
"SEM" indicates standard error of mean;
"n" indicates number of patient samples;
"Fold Change" indicates fold change of obese in comparison to lean patients.

[0235] NK4 expression was also evaluated using real time PCR. These data further show that NK4 is significantly over-expressed in muscle from obese individuals when compared to muscle from lean individuals.

Comparison	Expression Fold change	t test
Obese (16)/Lean (17)	3.22	0.004

Legend
"Fold change" indicates fold change in NK4 expression calculated as the ratio of mean obese expression/mean lean expression.
Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

Protein C Inhibitor (PCI)

[0236] Probe set Mbxhummus24898 detects expression of PCI nucleic acid sequences. Expression of transcripts encoding PCI was increased in diabetic patients compared to lean, non-diabetic patients in the gene profiling analysis described above.

	Lean Pre-Trog			Diabetic Pre-Trog					
B/C	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene name
B	294	84	17	930	210	19	3.17	0.013	Protein C inhibitor

[0237] In addition, expression of transcripts encoding PCI was reduced in tro-treated patients compared to untreated patients.

	Diabetic Pre-Trog			Diabetic Post-Trog				Stu-	
B/C	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	dents t test	Gene name
B	1422	340	9	722	247	10	0.51	0.005	Protein C inhibitor

Legend:
B/C indicates sample is from Basal or Clamp;
"Pre-Trog" and "Post-Trog" indicates sample was taken before or after 3 months of Troglitazone treatment;
"Mean Expr" indicates mean expression;
"SEM" indicates standard error of mean;
"n" indicates number of patient samples;
"Fold Change" indicates fold change of diabetic post tro/diabetic pre tro.

[0238] PCI expression was also evaluated by real time PCR in lean vs. obese individuals and in pre-troglitazone-treated vs post-troglitazone-treated atients.

Comparison	Expression Fold change	t test
Diabetes (19)/Lean (17)	2.22	0.027
Diabetics + Tro (9)/Diabetics − Tro (9)	0.33	0.039

Legend
"Tro" indicates treatment with Troglitazone.
"Fold change" indicates fold change in TIEG expression calculated as either the ratio of mean diabetic pre-Tro expression/mean lean pre-Tro expression or mean diabetic post-Tro expression/mean diabetic pre-Tro expression.
Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

[0239] These data further demonstrate that PCI is significantly over-expressed in muscle from diabetic individuals when compared to muscle from lean individuals. Furthermore, the data also show that treatment with Troglitazone reduces the expression of PCI in muscle from diabetic individuals.

[0240] Protein C inhibitor contains the following protein domains (designated with reference to SEQ ID NO:22): Signal Peptide at amino acids 1 to 16; and Serpin Domain at amino acids 50 to 406.

MAST205

[0241] Probe set MBXHUMMUS12940 detects both MAST205 and MAST205b transcripts.

Example 1

[0242] This example shows that MAST205/MAST205b sequences are up-regulated in muscle of diabetics when compared to muscle of lean non-diabetic individuals in the gene profiling experiment described above.

	Lean Pre-Trog			Diabetic Pre-Trog				Stu-	
B/C	Mean Exp	SEM	n	Mean Exp	SEM	n	Fold Change	dents t test	Gene name
B	505	127	17	1086	131	19	2.15	0.003	MAST205

Legend:
B/C indicates sample is from Basal or Clamp;
"Pre-Trog" and "Post-Trog" indicates sample was taken before or after 3 months of Troglitazone treatment;
"Mean Expr" indicates mean expression;
"SEM" indicates standard error of mean;
"n" indicates number of patient samples;
"Fold Change" indicates fold change of diabetic post tro/diabetic pre tro.

Example 2

[0243] This example shows that MAST205/MAST205b are down-regulated in muscle of diabetics treated for 3 months with troglitazone in the gene profiling experiment described above.

	Diabetic Pre-Trog			Diabetic Post-Trog				Stu-	
B/C	Mean Expr	SEM	N	Mean Expr	SEM	N	Fold Change	dents t test	Gene name
B	1234	411	9	919	325	8	0.74	0.01	MAST205

Example 3

[0244] Real-time PCR analysis further shows that MAST205 is significantly over-expressed in muscle from diabetic individuals when compared to muscle from lean individuals.

Comparison	Expression Fold change	t test
Diabetes (19)/Lean (17)	1.45	0.001

Legend
"Fold change" indicates fold change in MAT205 expression calculated as the ratio of mean obese expression/mean lean expression.
Numbers in parentheses indicate the number of patient samples analyzed by real time PCR

Example 4

[0245] This example shows that MAST205b is up regulated in muscle of diabetics when compared to muscle of lean non-diabetic individuals. It also demonstrates that MAST205b is down-regulated in muscle of diabetics after 3 months of troglitazone treatment compared to before treatment.

[0246] PCR primers and Taqman Probe were designed to detect specifically the expression of MAST205b. The sequences of the primers was as follows:

Forward primer:	110F - ACAGCAGTCCTGGCACTCCTT
Reverse primer:	174R - GCGGTTACTTGTCCGACAACTC
Probe 133:	TCCAGCCGCCCACTGCCG

Lean Pre-Trog Relative Exp (%)	Lean Post-Trog Relative Exp (%)	Diabetic Pre-Trog Relative Exp (%)	Diabetic Post-Tro Relative Exp (%)	Fold Change (D-/L-)	Fold Change (D+/D-)	Gene name
100	88	197	111	1.97	0.56	MAST205b

Legend:"Pre-Trog" and "Post-Trog" refer to samples taken before and after 3 months of troglitazone treatment respectively. "Relative Exp" refers to the expression of the gene relative to the Lean Pre-Trog sample, which is set to 100%. D-/L- refers to the ratio of relative expression in Diabetic Pre-Trog to relative expression in Lean Pre-Trog. D+/L+ refers to the ratio of relative expression in Diabetic Post-Trog compared to relative expression in Diabetic Pre-Trog.

Example 5

[0247] This example shows that MAST205 is up regulated in muscle of diabetics when compared to muscle of lean non-diabetic individuals. It also demonstrates that MAST205 is down-regulated in muscle of diabetics after 3 months of troglitazone treatment compared to before treatment.

[0248] PCR primers and Taqman Probe were designed to detect specifically the expression of MAST205. The sequences of the primers was as follows:

Forward primer:	717F - TTGGACAGTCTGCACCTTCTCTTA
Reverse primer:	801R - CGGTTACTTGTCCGACAAAAGC
Probe745:	TGGCCTGAAGGACTTGAGCCTTCCAGCCCACTGCCG

Lean Pre-Trog Relative Exp (%)	Lean Post-Trog Relative Exp (%)	Diabetic Pre-Trog Relative Exp (%)	Diabetic Post-Tro Relative Exp (%)	Fold Change (D-/L-)	Fold Change (D+/D-)	Gene name
100	98	242	66	2.42	0.27	MAST205

Legend: "Pre-Trog" and "Post-Trog" refer to samples taken before and after 3 months of troglitazone treatment respectively. "Relative Exp" refers to the expression of the gene relative to the Lean Pre-Trog sample, which is set to 100%. D-/L - refers to the ratio of relative expression in Diabetic Pre-Trog to relative expression in Lean Pre-Trog. D+/L+ refers to the ratio of relative expression in Diabetic Post-Trog compared to relative expression in in Diabetic Pre-Trog.

Example 6

[0249] This example shows that MAST205 is up-regulated in skeletal muscle of DBA/2J mice fed a high fat diet. These mice became insulin resistant after 28 weeks on a 32% or 42% fat diet, compared to littermates fed a chow diet, as measured by IPIST.

	Chow Diet	32% Fat Diet	42% Fat Diet	Gene name
Mean Rel Exp (%)	118	153	170	Mouse MAST205
SEM	13	13	9	
N	5	5	5	
Fold Change	—	1.3	1.4	
Students T-test	—	0.09	0.008	

Legend:
"Chow Diet" refers to standard mouse feed.
"32% Fat Diet" and "42% Fat Diet" refer to mouse feed from in 32% or 42% of the calories in the diet are obtained from fat, respectively.
"Mean Rel Exp (%)" refers to the average expression of the gene in muscles from 5 individual mice, relative to the expression in the muscle of a single mouse in the chow diet group.

Colon Kruppel-Like Factor (CKLF)

[0250] Probe set MBXHUMMUS28900 detects CKLF nucleic acid sequences. Expression of transcripts encoding CKLF was higher in diabetic patients as compared to lean, non-diabetic patients in the gene profiling analysis described above.

	Lean Pre-Trog			Diabetic Pre-Trog					
B/C	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene name
B	452	47	17	792	81	19	1.75	0.006	CKLF

Legend:
B/C indicates sample is from Basal or Clamp;
"Pre-Trog" indicates sample was taken before 3 months of Troglitazone treatment;
"Mean Expr" indicates mean expression;
"SEM" indicates standard error of mean;
"n" indicates number of patient samples;
"Fold change" indicates fold change of diabetic pre-trog/lean pre-trog.

[0251] Expression of transcripts encoding CKLF was also higher in obese patients as compared to lean patients in the gene profiling experiment.

	Lean Pre-Trog			Obese Pre-Trog					
B/C	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene name
B	321	46	7	753	104	7	2.34	0.005	CKLF

Legend:
B/C indicates sample is from Basal or Clamp;
"Pre-Trog" indicates sample was taken before 3 months of Troglitazone treatment;
"Mean Expr" indicates mean expression;
"SEM" indicates standard error of mean;
"n" indicates number of patient samples;
"Fold change" indicates fold change of obese pre-trog/lean pre-trog.

[0252] CKLF expression was also analyzed in diabetic individuals and lean individuals using real time PCR. The results showed that CKLF is significantly over-expressed in muscle from diabetic individuals when compared to muscle from lean individuals.

Comparison	Expression Fold change	t test
Diabetes (19)/Lean (17)	1.69	0.007

Legend
"Fold change" indicates fold change of mean diabetic/mean lean . . .
Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

[0253] Transcripts encoding CKLF contain the following protein domains (designated with reference to SEQ ID NO:34):

[0254] Zinc finger, C2H2 type domain encoded by amino acids 373 to 397

[0255] Zinc finger, C2H2 type domain encoded by amino acids 402 to 427

[0256] Zinc finger, C2H2 type domain encoded by amino acids 433 to 455.

Table of Sequences

[0257]

```
SEQ ID NO:1 Human Fritz nucleic acid sequence
HUM161414                    accession:U91903  coding sequence:70 . . . 1047
CGGAGACGGCGGAGCGGGCCTTGTTGGCGTCCACTGCGCGGNTGCACCCTGCCCCATCCTGCCGGGAT
CATGGTCTGCGGCAGCCCGGGAGGGATGCTGCTGCTGCGGGCCGGGCTGCTTGCCCTGGCTGCTCTCT
GCCTGCTCCGGGTGCCCGGGGCTCGGGCTGCAGCCTGTGAGCCCGTCCGCATCCCCCTGTGCAAGTCC
CTGCCCTGGAACATGACTAAGATGCCCAACCACCTGCACCACAGCACTCAGGACAACGCCATCCTGGC
CATCGAGCAGTTCGAAGGTCTGCTGGGCACCCACTGCAGCCCCGATCTGCTCTTCTTCCTCTGTGCCA
TGTACGCGCCCATCTGCACCATTGACTTCCAGCACGAGCCCATCAAGCCCTGTAAGTCTGTGTGCGAG
CGGGCCCGGCAGGGCTGTGAGCCCATACTCATCAAGTACCGCCACTCGTGGCCGGAGAACCTGGCCTG
CGAGGAGCTGCCAGTGTACGACAGGGGCGTGTGCATCTCTCCCGAGGCCATCGTTACTGCGGACGGAG
CTGATTTTCCTATGGATTCTAGTAACGGAAACTGTAGAGGGGCAAGCAGTGAACGCTGTAAATGTAAG
CCTATTAGAGCTACACAGAAGACCTATTTCCGGAACAATTACAACTATGTCATTCGGGCTAAAGTTAA
AGAGATAAAGACTAAGTGCCATGATGTGACTGCAGTAGTGGAGGTGAAGGAGATTCTAAAGTCCTCTC
TGGTAAACATTCCACGGGACACTGTCAACCTCTATACCAGCTCTGGCTGCCTCTGCCCTCCACTTAAT
GTTAATGAGGAATATATCATCATGGGCTATGAAGATGAGGAACGTTCCAGATTACTCTTGGTGGAAGG
CTCTATAGCTGAGAAGTGGAAGGATCGACTCGGTAAAAAAGTTAAGCGCTGGGATATGAAGCTTCGTC
ATCTTGGACTCAGTAAAAGTGATTCTAGCAATAGTGATTCCACTCAGAGTCAGAAGTCTGGCAGGAAC
```

-continued

TCGAACCCCCGGCAAGCACGCAACTAAATCCCGAAATACAAAAAGTAACACAGTGGACTTCCTATTAA
GACTTACTTGCATTGCTGGACTAGCAAAGGAAAATTGCACTATTGCACATCATATTCTATTGTTTACT
ATAAAAATCATGTGATAACTGATTATTACTTCTGTTTCTCTTTTGGTTTCTGCTTCTCTCTTCTCTCA
ACCCCTTTGTAATGGTTTGGGGGCAGACTCTTAAGTATATTGTGAGTTTTCTATTTCACTAATCATGA
GAAAAACTGTTCTTTTGCAATAATAATAAATTAAACATGCTGTTACCAGAGCCTCTTTGCTGGAGTCT
CCAGATGTTAATTTACTTTCTGCACCCCAATTGGGAATGCAATATTGGATGAAAAGAGAGGTTTCTGG
TATTCACAGAAAGCTAGATATGCCTTAAAACATACTCTGCCGATCTAATTACAGCCTTATTTTTGTAT
GCCTTTTGGGCATTCTCCTCATGCTTAGAAAGTTCCAAATGTTTATAAAGGTAAAATGGCAGTTTGAA
GTCAAATGTCACATAGGCAAAGCAATCAAGCACCAGGAAGTGTTTATGAGGAAACAACACCCAAGATG
AATTATTTTTGAGACTGTCAGGAAGTAAAATAAATAGGAGCTTAAGAAAGAACATTTTGCCTGATTGA
GAAGCACAACTGAAACCAGTAGCCGCTGGGGTGTTAATGGTAGCATTCTTCTTTTGGCAATACATTTG
ATTTGTTCATGAATATATTAATCAGCATTAGAGAAATGAATTATAACTAGACATCTGCTGTTATCACC
ATAGTTTTGTTTAATTTGCTTCCTTTTAAATAAACCCATTGGTGAAAGTCCCAAAAAAAAAAAAAAAA
AAAAA

SEQ ID NO:2 Human Fritz polypeptide sequence
Protein sequence protein_id:gi1917007
MVCGSPGGMLLLRAGLLALAALCLLRVPGARAAACEPVRIPLCKSLPWNMTKMPNHLHHSTQDNAILAIEQFEGL
LGTHCSPDLLFFLCAMYAPICTIDFQHEPIKPCKSVCERARQGCEPILIKYRHSWPENLACEELPVYDRGVCISP
EAIVTADGADFPMDSSNGNCRGASSERCKCKPIRATQKTYFRNNYNYVIRAKVKEIKTKCHDVTAVVEVKEILKS
SLVNIPRDTVNLYTSSGCLCPPLNVNEEYIIMGYEDEERSRLLLVEGSIAEKWKDRLGKKVKRWDMKLRHLGLSK
SDSSNSDSTQSQKSGRNSNPRQARN

SEQ ID NO:3 Mouse Fritz nucleic acid sequence
accession:NM_011356 coding sequence:365 . . . 1336
GAATTCGGCACGAGCTGAATTTGACTTTTGTTTTTATTTCTCTCTGGCTTCCTCTTCTGCCCCCTCATCTGATTG
ATGTGCTAAGGCTGATGTCTCTGCCAGAGCGAGAGGAATAAATAGATGCTGCCTCGCCTAGAGGCTTAGACGCTT
GGGAAGAGCAGCCGGCGCACGAGCGACCGGGCTCCGCCAAGCTAGTGGACCGGACCTGGGAGCACTTGGATCCAA
GAGAACTGTGATTGTCCCAGGGGTGGGGGCAGCTCCCCAGGTCGTTGGGATCACCCCTCGGAACCGCAGGGGGAG
ACTTCGGAACGAAAGTGTCTCCCGCGTCCGTCGCTCGTGCGCCCTGCCCCATCCTGCTGGGACCATGGTCTGCTG
CGGCCCGGGACGGATGCTGCTAGGATGGGCCGGGTTGCTAGTCCTGGCTGCTCTCTGCCTGCTCCAGGTGCCCGG
AGCTCAGGCTGCAGCCTGTGAGCCTGTCCGCATCCCGCTGTGCAAGTCCCTTCCCTGGAACATGACCAAGATGCC
CAACCACCTGCACCACAGCACCCAGGCTAACGCCATCCTGGCCATGGAACAGTTCGAAGGGCTGCTGGGCACCCA
CTGCAGCCCGGATCTTCTCTTCTTCCTCTGTGCAATGTACGCACCCATTTGCACCATCGACTTCCAGCACGAGCC
CATCAAGCCCTGCAAGTCTGTGTGTGAGCGCGCCCGACAGGGCTGCGAGCCCATTCTCATCAAGTACCGCCACTC
GTGGCCGGAAAGCTTGGCCTGCGACGAGCTGCCGGTGTACGACCGCGGCGTGTGCATCTCTCCTGAGGCCATCGT
CACCGCGGACGGAGCGGATTTTCCTATGGATTCAAGTACTGGACACTGCAGAGGGGCAAGCAGCGAACGTTGCAA
ATGTAAGCCTGTCAGAGCTACACAGAAGACCTATTTCCGGAACAATTACAACTATGTCATCCGGGCTAAAGTTAA
AGAGGTAAAGATGAAATGTCATGATGTGACCGCCGTTGTGGAAGTGAAGGAAATTCTAAAGGCATCACTGGTAAA
CATTCCAAGGGAcACCGTCAATCTTTATACCACCTCTGGCTGCCTCTGTCCTCCACTTACTGTCAATGAGGAATA
TGTCATCATGGGCTATGAAGACGAGGAACGTTCCAGGTTACTCTTGGTAGAAGGCTCTATAGCTGAGAAGTGGAA
GGATCGGCTTGGTAAGAAAGTCAAGCGCTGGGATATGAAACTCCGACACCTTGGACTGGGTAAAACTGATGCTAG
CGATTCCACTCAGAATCAGAAGTCTGGCAGGAACTCTAATCCCCGGCCAGCACGCAGCTAAATCCTGAAATGTAA
AAGGCCACACCCACGGACTCCCTTCTAAGACTGGCGCTGCTGGACTAACAAAGGAAAACGCACAGTTGTGCTCGT
GACCGATTGTTTACCGCAGACACCGCGTGGCTACCGAAGTTACTTCCGGTCCCCTTTCTCCTGCTTCTTAATGGC
CTGGGGTTAGATCCTTTAATATGTTATATATTCTGTTTCATCAATCACGTGGGGACTGTTCTTTTGCAACCAGAA
TAGTAAATTAAATATGTTGATGCTAAGGTTTCTGTACTGGACTCCCTGGGTTTAATTTGGTGTTCTGTACCCTGA
TTGAGAATGCAATGTTTCATGTAAAGAGAGAATCCTGGTCATATCTCAAGAACTAGATATTGCTGTAAGACAGCC
TCTGCTGCTGCGCTTATAGTCTTGTGTTTGTACCTGTTTGGCCATTTCCCTCATGCTGTGAAAGTTATACATGTT
TATAAAGGTAGAACGGCATTTTGAAATCAGACACTGCACAAGCAGAGTAGCCCAACACCAGGAAGCATTTATGAG
GAAACGCCACACAGCATGACTTATTTTCAAGATTGGCAGGCAGCAAAATAAATAGTGTTGGGAGCCAAGAAAAGA
ATATTTTGCCTGGTTAAGGGGCACACTGGAATCAGTAGCCTTGAGCCATTAACAGCAGTGTTCTTCTGGCACGTT
TTTGATTTGTTCATAAATGTATTCACGAGCATTAGAGATGAACTTATAACTAGACATCTGTTGTTATCACTATAG
CTCTGCTTCCTTCTAAATCAAACCCATTGTTGGATGCTCCCTCTCCATTCATAAATAAATTTGGCTTGCTGGTAT
TGGCCAGGAAAAGAAAGTATTAAAGTATGCATGCATGTGCACCAGGGTGTTATTTAACAGAGGTATGTAACTCTA
TAAAAGACTATAATTTACAGGACACGGAAATGTGCACATTTGTTTACTTTTTTTCTTCCTTTTGCTTTGGGCTTG
TGATTTTGGTTTTTGGTGTGTTTATGTCTGTATTTTGGGGGGTGGGTAGGTTTAANCCATTGCACATTCAAGTTG
NACTAGATTAGAGTAGACTAGGCTCATTGGCCTAGACATTATGATTTGAATTTGTGTTGTTTAATGCTCCATCAA
GATGTCTAATAAAAGGAATATGGTTGTCAACAGAGACGACAACACCCAAAAAAAAAAAAAAAAA

-continued

SEQ ID NO:4 Mouse Fritz polypeptide sequence
Protein sequence accession:gi6755476
MVCCGPGRMLLGWAGLLVLAALCLLQVPGAQAAACEPVRIPLCKSLPWNMTKMPNHLHHSTQANAILAMEQFEGL
LGTHCSPDLLFFLCAMYAPICTIDFQHEPIKPCKSVCERARQGCEPILIKYRHSWPESLACDELPVYDRGVCISP
EAIVTADGADFPMDSSTGHCRGASSERCKCKPVRATQKTYFRNNYNYVIRAKVKEVKMKCHDVTAVVEVKEILKA
SLVNIPRDTVNLYTTSGCLCPPLTVNEEYVIMGYEDEERSRLLLVEGSIAEKWKDRLGKKVKRWDMKLRHLGLGK
TDASDSTQNQKSGRNSNPRPARS

SEQ ID NO:5 Human PAK1B nucleic acid sequence
HUM137707 accession:U51120 coding sequence:55 . . . 1692
GGAGAGCCGAGAGGAGCTGAGCGAGCGCGGAAGTAGCTGCTGCTGGTGGTGACAATGTCAAATAACGGCCTAGAC
ATTCAAGACAAACCCCCAGCCCCTCCGATGAGAAATACCAGCACTATGATTGGAGCCGGCAGCAAAGATGCTGGA
ACCCTAAACCATGGTTCTAAACCTCTGCCTCCAAACCCAGAGGAGAAGAAAAAGAAGGACCGATTTTACCGATCC
ATTTTACCTGGAGATAAAACAAATAAAAAGAAAGAGAAAGAGCGGCCAGAGATTTCTCTCCCTTCAGATTTTGAA
CACACAATTCATGTCGGTTTTGATGCTGTCACAGGGGAGTTTACCGGAATGCCAGAGCAGTGGGCCCGCTTGCTT
CAGACATCAAATATCACTAAGTCGGAGCAGAAGAAAAACCCGCAGGCTGTTCTGGATGTGTTGGAGTTTTACAAC
TCGAAGAAGACATCCAACAGCCAGAAATACATGAGCTTTACAGATAAGTCAGCTGAGGATTACAATTCTTCTAAT
GCCTTGAATGTGAAGGCTGTGTCTGAGACTCCTGCAGTGCCACCAGTTTCAGAAGATGAGGATGATGATGATGAT
GATGCTACCCCACCACCAGTGATTGCTCCACGCCCAGAGCACACAAAATCTGTATACACACGGTCTGTGATTGAA
CCACTTCCTGTCACTCCAACTCGGGACGTGGCTACATCTCCCATTTCACCTACTGAAAATAACACCACTCCACCA
GATGCTTTGACCCTTAATACTGAGAAGCAGAAGAAGAAGCCTAAAATGTCTGATGAGGAGATCTTGGAGAAATTA
CGAAGCATAGTGAGTGTGGGCGATCCTAAGAAGAAATATACACGGTTTGAGAAGATTGGACAAGGTGCTTCAGGC
ACCGTGTACACAGCAATGGATGTGGCCACAGGACAGGAGGTGGCCATTAAGCAGATGAATCTTCAGCAGCAGCCC
AAGAAAGAGCTGATTATTAATGAGATCCTGGTCATGAGGGAAAACAAGAACCCAAACATTGTGAATTACTTGGAC
AGTTACCTCGTGGGAGATGAGCTGTGGGTTGTTATGGAATACTTGGCTGGAGGCTCCTTGACAGATGTGGTGACA
GAAACTTGCATGGATGAAGGCCAAATTGCAGCTGTGTGCCGTGAGTGTCTGCAGGCTCTGGAGTCTTTGCATTCG
AACCAGGTCATTCACAGAGACATCAAGAGTGACAATATTCTGTTGGGAATGGATGGCTCTGTCAAGCTAACTGAC
TTTGGATTCTGTGCACAGATAACCCCAGAGCAGAGCAAACGGAGCACCATGGTAGGAACCCCATACTGGATGGCA
CCAGAGGTTGTGACACGAAAGGCCTATGGGCCCAAGGTTGACATCTGGTCCCTGGGCATCATGGCCATCGAAATG
ATTGAAGGGGAGCCTCCATACCTCAATGAAAACCCTCTGAGAGCCTTGTACCTCATTGCCACCAATGGGACCCCA
GAACTTCAGAACCCAGAGAAGCTGTCAGCTATCTTCCGGGACTTTCTGAACCGCTGTCTCGAGATGGATGTGGAG
AAGAGAGGTTCAGCTAAAGAGCTGCTACAGCATCAATTCCTGAAGATTGCCAAGCCCCTCTCCAGCCTCACTCCA
CTGATTGCTGCAGCTAAGGAGGCAACAAAGAACAATCACTAAAACCACACTCACCCCAGCCTCATTGTGCCAAGC
CTTCTGTGAGATAAA

SEQ ID NO:6 Human PAK1B polypeptide sequence
protein_id:gi1256422
MSNNGLDIQDKPPAPPMRNTSTMIGAGSKDAGTLNHGSKPLPPNPEEKKKKDRFYRSILPGDKTNKKKEKERPEI
SLPSDFEHTIHVGFDAVTGEFTGMPEQWARLLQTSNITKSEQKKNPQAVLDVLEFYNSKKTSNSQKYMSFTDKSA
EDYNSSNALNVKAVSETPAVPPVSEDEDDDDDDATPPPVIAPRPEHTKSVYTRSVIEPLPVTPTRDVATSPISPT
ENNTTPPDALTLNTEKQKKKPKMSDEEILEKLRSIVSVGDPKKKYTRFEKIGQGASGTVYTAMDVATGQEVAIKQ
MNLQQQPKKELIINEILVMRENKNPNIVNYLDSYLVGDELWVVMEYLAGGSLTDVVTETCMDEGQIAAVCRECLQ
ALESLHSNQVIHRDIKSDNILLGMDGSVKLTDFGFCAQITPEQSKRSTMVGTPYWMAPEVVTRKAYGPKVDIWSL
GIMAIEMIEGEPPYLNENPLRALYLIATNGTPELQNPEKLSAIFRDFLNRCLEMDVEKRGSAKELLQHQFLKIAK
PLSSLTPLIAAAKEATKNNH

SEQ ID NO:7 Human PAK1B splice variant nucleic acid sequence
accession:AF0718844 coding sequence:12 . . . 1673
TGGTGGTGACAATGTCAAATAACGGCCTAGACATTCAAGACAAACCCCCAGCCCCTCCGATGAGAAATACCAGCA
CTATGATTGGAGTCGGCAGCAAAGATGCTGGAACCCTAAACCATGGTTCTAAACCTCTGCCTCCAAACCCAGAGG
AGAAGAAAAAGAAGGACCGATTTTACCGATCCATTTTACCTGGAGATAAAACAAATAAAAAGAAAGAGAAAGAGC
GGCCAGAGATTTCTCTCCCTTCAGATTTTGAACACACAATTCATGTCGGTTTTGATGCTGTCACAGGGGAGTTTA
CGGGAATGCCAGAGCAGTGGGCCCGCTTGCTTCAGACATCAAATATCACTAAGTCGGAGCAGAAGAAAAACCCGC
AGGCTGTTCTGGATGTGTTGGAGTTTTACAACTCGAAGAAGACATCCAACAGCCAGAAATACATGAGCTTTACAG
ATAAGTCAGCTGAGGATTACAATTCTTCTAATGCCTTGAATGTGAAGGCTGTGTCTGAGACTCCTGCAGTGCCAC
CAGTTTCAGAAGATGAGGATGATGATGATGATGATGCTACCCCACCACCAGTGATTGCTCCACGCCCAGAGCACA
CAAAATCTGTATACACACGGTCTGTGATTGAACCACTTCCTGTCACTCCAACTCGGGACGTGGCTACATCTCCCA
TTTCACCTACTGAAAATAACACCACTCCACCAGATGCTTTGACCCGGAATACTGAGAAGCAGAAGAAGAAGCCTA
AAATGTCTGATGAGGAGATCTTGGAGAAATTACGAAGCATAGTGAGTGTGGGCGATCCTAAGAAGAAATATACAC
GGTTTGAGAAGATTGGACAAGGTGCTTCAGGCACCGTGTACACAGCAATGGATGTGGCCACAGGACAGGAGGTGG
CCATTAAGCAGATGAATCTTCAGCAGCAGCCCAAGAAAGAGCTGATTATTAATGAGATCCTGGTCATGAGGGAAA
ACAAGAACCCAAACATTGTGAATTACTTGGACAGTTACCTCGTGGGAGATGAGCTGTGGGTTGTTATGGAATACT
TGGCTGGAGGCTCCTTGACAGATGTGGTGACAGAAACTTGCATGGATGAAGGCCAAATTGCAGCTGTGTGCCGTG
AGTGTCTGCAGGCTCTGGAGTTCTTGCATTCGAACCAGGTCATTCACAGAGACATCAAGAGTGACAATATTCTGT
TGGGAATGGATGGCTCTGTCAAGCTAACTGACTTTGGATTCTGTGCACAGATAACCCCAGAGCAGAGCAAACGGA
GCACCATGGTAGGAACCCCATACTGGATGGCACCAGAGGTTGTGACACGAAAGGCCTATGGGCCCAAGGTTGACA
TCTGGTCCCTGGGCATCATGGCCATCGAAATGATTGAAGGGGAGCCTCCATACCTCAATGAAAACCCTCTGAGAG
CCTTGTACCTCATTGCCACCAATGGGACCCCAGAACTTCAGAACCCAGAGAAGCTGTCAGCTATCTTCCGGGACT
TTCTGAACCGCTGTCTCGAGATGGATGTGGAGAAGAGAGGTTCAGCTAAAGAGCTGCTACAGGTGAGAAAACTGA
GGTTTCAAGTGTTTAGTAACTTTTCCATGATAGCTGCATCAATTCCTGAAGATTGCCAAGCCCCTCTCCAGCCTC
ACTCCACTGATTGCTGCAGCTAAGGAGGCAACAAAGAACAATCACTAAAACCACACTCACCCCAGCCTCATTGTG
CCAAGCTCTGTGAGATAAATGCACATTTCAGAAATTCCAACTCCTGATGCCCTCTTCTCCTTGCCTTGCTTCTCC
CATTTCCTGATCTAGCACTCCTCAAGACTTTGATCCTTGGAAACCGTGTGTCCAGCATTGAAGAGAACTGCAACT
GAATG

-continued

SEQ ID NO:8 Human PAK1B splice variant polypeptide sequence
protein_id:gi3265160
MSNNGLDIQDKPPAPPMRNTSTMIGVGSKDAGTLNHGSKPLPPNPEEKKKKDRFYRSILPGDKTNKKKEKERPEI
SLPSDFEHTIHVGFDAVTGEFTGMPEQWARLLQTSMITKSEQKKNPQAVLDVLEFYNSKKTSNSQKYMSFTDKSA
EDYNSSNALNVKAVSETPAVPPVSEDEDDDDDATPPPVIAPRPEHTKSVYTRSVIEPLPVTPTRDVATSPISPT
ENNTTPPDALTRNTEKQKKKPKMSDEEILEKLRSIVSVGDPKKKYTRFEKIGQGASGTVYTAMDVATGQEVAIKQ
MNLQQQPKKELIINEILVMRENKNPNIVNYLDSYLVGDELWVVMEYLAGGSLTDVVTETCMDEGQIAAVCRECLQ
ALEFLHSNQVIHRDIKSDNILLGMDGSVKLTDFGFCAQITPEQSKRSTMVGTPYWMAPEVVTRKAYGPKVDIWSL
GIMAIEMIEGEPPYLNENPLRALYLIATNGTPELQNPEKLSAIFRDFLNRCLEMDVEKRGSAKELLQVRKLRFQV
FSNFSMIAASIPEDCQAPLQPHSTDCCS

SEQ ID NO:9 Human PAK1B new splice variant nucleic acid sequence
TCAAATAACGGCCTAGACATTCAAGACAAACCCCCAGCCCCTCCGATGAGAAATACCAGCACTATGAT
TGGAGCCGGCAGCAAAGATGCTGGAACCCTAAACCATGGTTCTAAACCTCTGCCTCCAAACCCAGAGG
AGAAGAAAAAGAAGGACCGATTTTACCGATCCATTTTACCTGGAGATAAAACAAATAAAAAGAAAGAG
AAAGAGCGGCCAGAGATTTCTCTCCCTTCAGATTTTGAACACACAATTCATGTCGGTTTTGATGCTGT
CACAGGGGAGTTTACGGGAATGCCAGAGCAGTGGGCCCGCTTGCTTCAGACATCAAATATCACTAAGT
CGGAGCAGAAGAAAAACCCGCAGGCTGTTCTGGATGTGTTGGAGTTTTACAACTCGAAGAAGACATCC
AACAGCCAGAAATACATGAGCTTTACAGATAAGTCAGCTGAGGATTACAATTCTTCTAATGCCTTGAA
TGTGAAGGCTGTGTCTGAGACTCCTGCAGTGCCACCAGTTTCAGAAGATGAGGATGATGATGATGATG
ATGCTACCCCACCACCAGTGATTGCTCCACGCCCAGAGCACACAAAATCTGTGGCCATTAAGCAGATG
AATCTTCAGCAGCAGCCCAAGAAAGAGCTGATTATTAATGAGATCCTGGTCATGAGGGAAAACAAGAA
CCCAAACATTGTGAATTACTTGGACAGTTACCTCGTGGGAGATGAGCTGTGGGTTGTTATGGAATACT
TGGCTGGAGGCTCCTTGACAGATGTGGTGACAGAAACTTGCATGGATGAAGGCCAAATTGCAGCTGTG
TGCCGTGAGTGTCTGCAGGCTCTGGAGTTCTTGCATTCGAACCAGGTCATTCACAGAGACATCAAGAG
TGACAATATTCTGTTGGGAATGGATGGCTCTGTCAAGCTAACTGACTTTGGATTCTGTGCACAGATAA
CCCCAGAGCAGAGCAAACGGAGCACCATGGTAGGAACCCCATACTGGATGGCACCAGAGGTTGTGACA
CGAAAGGCCTATGGGCCCAAGGTTGACATCTGGTCCCTGGGCATCATGGCCATCGAAATGATTGAAGG
GGAGCCTCCATACCTCAATGAAAACCCTCTGAGAGCCTTGTACCTCATTGCCACCAATGGGACCCCAG
AACTTCAGAACCCAGAGAAGCTGTCAGCTATCTTCCGGGACTTTCTGAACCGCTGTCTCGGGATGGAT
GTGGAGAAGAGAGGTTCAGCTAAAGAGCTGCTACAGCATCAATTCCTGAAGATTGCCAAGCCCCTCTC
CAGCCTCACTCCACTGATTGCTGCAGCTAAGGAGGCAACAAAGAACAATCACTAA

SEQ ID NO:10 Human PAK1B new splice variant polypeptide sequence
MSNNGLDIQDKPPAPPMRNTSTMIGAGSKDAGTLNHGSKPLPPNPEEKKKKDRFYRSILPGDKTNKKK
EKERPEISLPSDFEHTIHVGFDAVTGEFTGMPEQWARLLQTSNITKSEQKKNPQAVLDVLEFYNSKKT
SNSQKYMSFTDKSAEDYNSSNALNVKAVSETPAVPPVSEDEDDDDATPPPVIAPRPEHTKSVAIKQ
MNLQQQPKKELIINEILVMRENKNPNIVNYLDSYLVGDELWVVMEYLAGGSLTDVVTETCMDEGQIAA
VCRECLQALEFLHSNQVIHRDIKSDNILLGMDGSVKLTDFGFCAQITPEQSKRSTMVGTPYWMAPEVV
TRKAYGPKVDIWSLGIMAIEMIEGEPPYLNENPLRALYLIATNGTPELQNPEKLSAIFRDFLNRCLGM
DVEKRGSAKELLQHQFLKIAKPLSSLTPLIAAAKEATKNNH

SEQ ID NO:11 Mouse PAK1B nucleic acid sequence
accession:NM_011035 coding sequence:190 . . . 1827
TGAGTGTGCTGCTCCTTGGGACGTGGGTGGCTGCGCTGCGGCTGTTAGCCGAATCCAGCCTGTCAGGCTGGTTCA
GCCCGGGCTTTGTCCCAACTGCTGAGCAGGGAGAGGTGCGGCTCTCGAGACTCACAGATACACAAGATCACGCCC
CGCACCCACCGCCAGTAGCTGCTGCTGCTGGTGGTGACAATGTCAAATAACGGCGTAGACATCCAGGACAAACCC
CCAGCCCCTCCGATGAGAAACACCAGCACTATGATTGGAGCCGGCAGCAAAGACACTGGAACCCTAAACCACGGC
TCCAAACCTCTGCCTCCAAACCCAGAGGAGAAGAAAAAGAAGGACCGGTTTTATCGATCCATCTTACCTGGAGAT
AAAACAAATAAAAAGAGGGAGAAGGAGCGACCAGAGATTTCTCTTCCTTCAGATTTTGAGCATACAATTCATGTT
GGTTTTGATGCTGTCACAGGGGAGTTTACGGGAATGCCAGAACAGTGGGCTCGCTTGCTTCAAACATCAAATATC
ACAAAGTCAGAGCAGAAGAAAAACCCACAGGCTGTTCTGGATGTGTTGGAATTTTATAACTCTAAGAAGACCTCC
AATAGTAAGAAGTACATGAGTTTTACAGATAAGTCAGCTGAAGATTATAATTCTTCTAACACTTTGAATGTGAAG
ACTGTGTCTGAGACCCCAGCAGTACCACCAGTGTCAGAAGATGATGAAGATGATGATGACGATGCTACCCCACCT
CCAGTGATTGCTCCACGCCCAGAACACACAAAATCTGTATATACACGATCTGTGATTGAACCACTTCCTGTTACT
CCAACTCGGGATGTGGCTACATCTCCTATTTCTCCTACTGAGAATAACACCACTCCGCCAGATGCTTTGACCCGG
AACACGGAAAAACAGAAGAAGAAGCCTAAAATGTCTGATGAGGAGATCTTAGAGAAATTACGGAGCATAGTGAGT
GTGGGTGACCCCAAGAAGAAGTACACACCGTTCGAGAAGATTGGACAAGGTGCTTCAGGCACAGTGTATACTGCA
ATGGATGTAGCCACAGGGCAGGAGGTGGCCATTAAACAGATGAATCTTCAGCAGCAGCCGAAGAAAGAGCTGATT
ATTAATGAGATCCTGGTCATGAGGGAAAACAAAAACCCAAATATTGTCAACTACCTGGACAGTTACCTTGTGGGA
GATGAGCTGTGGGTTGTTATGGAATACTTGGCTGGAGGCTCCTTGACAGATGTGGTGACAGAAACCTGTATGGAT
GAAGGCCAGATAGCAGCTGTGTGCCGAGAGTGTCTACAAGCTTTGGAGTTTCTACATTCAAACCAAGTCATTCAC
AGGGACATCAAGAGTGACAATATTCTGCTGGGAATGGATGGCTCTGTCAAGTTAACTGACTTTGGATTCTGTGCA
CAGATAACTCCAGAGCAGAGCAAAAGGAGCACCATGGTGGGAACTCCATATTGGATGGCACCTGAAGTTGTGACA
CGCAAGGCTTATGGACCCAAGGTTGACATCTGGTCCCTGGGCATTATGGCAATTGAAATGATTGAGGGGGAGCCC
CCATACCTCAATGAAAACCCTTTGAGAGCCTTGTACCTCATTGCTACCAATGGGACGCCAGAGCTTCAGAACCCA
GAGAAGTTGTCAGCTATCTTCCGGGACTTTCTGCAATGCTGTCTTGAGATGGATGTGGAGAAGAGAGGCTCAGCT
AAAGAGCTGCTGCAGCATCAGTTCCTGAAGATTGCCAAGCCCCTCTCTAGCCTGACTCCACTGATGCATGCTGCA
AAAGAGGCAACCAAGAACAATCACTGAAACCATGCTCATCCCAGCCTCATGTGCCAAGCCTTCTATGAAATAAAC
ACTTGTTTCGGGAACTCCGACACCTCATGTCCTCTTCTCCTTTCCTTGCTTCTCCCATTTCCTGATCTAGTGCTC
CCAAGACTTTGATCCTTGGAAACTGTC

SEQ ID NO:12 Mouse PAK1B polypeptide sequence
accession:gi6754978
MSNNGVDIQDKPPAPPMRNTSTMIGAGSKDTGTLNHGSKPLPPNPEEKKKKDRFYRSILPGDKTNKKREKERPEI
SLPSDFEHTIHVGFDAVTGEFTGMPEQWARLLQTSNITKSEQKKNPQAVLDVLEFYNSKKTSNSKKYMSFTDKSA

-continued

```
EDYNSSNTLNVKTVSETPAVPPVSEDDEDDDDDATPPPVIAPRPEHTKSVYTRSVIEPLPVTPTRDVATSPISPT
ENNTTPPDALTRNTEKQKKKPKMSDEEILEKLRSIVSVGDPKKKYTPFEKIGQGASGTVYTAMDVATGQEVAIKQ
MNLQQQPKKELIINEILVMRENKNPNIVNYLDSYLVGDELWVVMEYLAGGSLTDVVTETCMDEGQIAAVCRECLQ
ALEFLHSNQVIHRDIKSDNILLGMDGSVKLTDFGFCAQITPEQSKRSTMVGTPYWMAPEVVTRKAYGPKVDIWSL
GIMAIEMIEGEPPYLNENPLRALYLIATNGTPELQNPEKLSAIFRDFLQCCLEMDVEKRGSAKELLQHQFLKIAK
PLSSLTPLMHAAKEATKNNH
```

SEQ ID NO:13 Rat PAK1B nucleic acid sequence
accession:NM_017198 coding sequence:389–2023

```
CCCGGGCGGAAGGTACCGCCGCACGGTGGGCGCTCCTCACCGCTTCCCCACACCTCCCAGTCCCACTGCACCTCT
AGGCTGCGCGGAGTCCTCTCGGCTATTACCGGCTCTTTGGAGCCGCCTGTGCCCCCTCAGAGCCCGCGCTCTCCA
CAGTCTCCTCAGAGAAGGACCCCCTCAGAGAGCGCACCCCGCCCCTTCGCAGTCCGGGCGCGCCTCCCGCTCCCC
GCGGCCCGCAGCGTCCCGCCCTCCAGAGCCCGGCTCGAGCGGCGAGAGGAGAACGCGGGGCTGCCGCCCGCGCAC
ACTCCCGCGCTCTCTCTGGGTTTGGCTGCCGCCAGGAGGAGGAGGAGCGGAGCGAGCGCAGAAGTAGCTGCTGCT
GCTGGTGGTGACAATGTCAAATAACGGCTTAGACGTCCAGGACAAACCCCCAGCCCCTCCGATGAGAAACACCAG
CACTATGATTGGAGCTGGCAGCAAAGACCCTGGAACCCTAAACCACGGTTCCAAACCTCTGCCTCCAAACCCAGA
GGAGAAGAAAAAGAAGGACCGGTTCTATCGATCCATCTTAGCTGGAGATAAAACAAATAAAAAGAAGGAGAAGGA
GCGGCCAGAGATTTCTCTTCCTTCAGATTTTGAGCATACAATTCATGTTGGTTTTGATGCTGTCACAGGGGAGTT
TACGGGGATGCCAGAACAGTGGGCCCGCTTGCTTCAAACATCAAATATCACCAAGTCAGAGCAGAAGAAAAACCC
ACAGGCTGTTCTGGATGTGTTGGAATTTTATAACTCCAAGAAGACCTCCAATAGTCAGAAGTACATGAGTTTTAC
AGATAAGTCAGCTGAAGATTATAATTCTTCTAATACTTTGAATGTGAAGACTGTGTCTGAGACCCCAGCAGTGCC
ACCAGTGTCAGAAGATGAAGATGATGATGACGATGCTACCCCACCTCCAGTGATTGCTCCACGCCCAGAACACAC
AAAATCTGTATATACACGGTCTGTGATTGAACCACTTCCTGTTACTCCAACTCGGGATGTGGCTACATCTCCTAT
TTCTCCTACTGAGAATAACACCACTCCGCCAGATGCTTTGACCCGGAATACTGAAAAGCAGAAGAAGAAGCCTAA
AATGTCTGATGAGGAGATCTTAGAGAAATTACGGAGCATAGTGAGTGTGGGCGATCCTAAGAAGAAGTACACACG
CTTCGAGAAGATTGGACAAGGTGCTTCAGGCACAGTGTACACTGCAATGGATGTAGCCACAGGGCAGGAGGTGGC
CATTAAACAGATGAACCTTCAGCAGCAGCCGAAGAAAGAACTCATTATTAATGAGATCCTGGTCATGAGGGAAAA
CAAAAACCCAAACATTGTGAACTATCTGGACAGTTACCTTGTGGGAGATGAGCTATGGGTTGTCATGGAATACTT
GGCCGGAGGCTCCTTAACAGATGTGGTGACAGAAACCTGTATGGATGAAGGCCAGATAGCAGCTGTGTGCCGAGA
GTGTCTACAAGCTTTGGAGTTCCTGCATTCAAACCAAGTCATTCACAGAGACATCAAGAGTGACAATATTCTGCT
GGGAATGGATGGCTCTGTCAAATTAACTGACTTTGGATTCTGTGCACAGATAACTCCAGAGCAGAGCAAAAGGAG
CACCATGGTGGGAACTCCATATTGGATGGCACCTGAAGTTGTGACACGCAAGGCCTATGGACCCAAGGTTGACAT
CTGGTCCCTGGGTATTATGGCAATTGAAATGATTGAGGGGGAGCCCCCATACCTCAATGAAAACCCTTTGAGAGC
CTTGTACCTCATTGCTACCAATGGGACGCCAGAGCTTCAGAACCCAGAGAAGTTGTCAGCTATTTTCCGGGACTT
TTTGAACCGCTGTCTTGAGATGGATGTGGAGAAGAGAGGTTCAGCTAAAGAGCTACTGCAGCATCAATTCCTGAA
GATTGCCAAGCCTCTCTCCAGCCTGACTCCACTGATTGCTGCAGCAAAAGAGGCAACCAAGAACAATCACTGAAA
CCACGCTCACCCCAGCCTCATGTGCCAAGCCTTCTATGAAATAAACACTCGTTTCGGGAACTCCGACCCCTCATG
TCCTCTTCTCCTTTCCTTGCTTCTCCCATTTCCTGATCTAGTGCTCTCAAGACTTTGATCCTTGGAAACCGTCTA
GCACTGAAGAGAACCGCAACCGGATGACTAATCCAGCAGAGGCCATTTCTAAATAGGAATTCCCTTTAGCTTGTG
GGCATGGAGGGGACTGATGAGCAAGGGTTTACCTGAATAAACCTGTTTCTACGAAACAGAAATCTCAACCATCCC
ATTCCTTACCCCTCACAATCAGTTCTTAACTCTATAAACTTATGGTTTGATAGCATTATCAATTTGCTATCAGTT
GAAATTGCTTTTGTTTTTTATTTCTGTGACCAAATTGCCCAAACACTTCATTGTATTTGAAAACCAGAACAGCTT
TGAAATGCCACGGGGCCTGATAATCTGCCAGGGACATGAAGAGGTCTTGTTTCCCTGAACCCAC
```

SEQ ID NO:14 Rat PAK1B polypeptide sequence
accession:gi8393901

```
MSNNGLDVQDKPPAPPMRNTSTMIGAGSKDPGTLNHGSKPLPPNPEEKKKKDRFYRSILAGDKTNKKKEKERPEI
SLPSDFEHTIHVGFDAVTGEFTGMPEQWARLLQTSNITKSEQKKNPQAVLDVLEFYNSKKTSNSQKYMSFTDKSA
EDYNSSNTLNVKTVSETPAVPPVSEDEDDDDATPPPVIAPRPEHTKSVYTRSVIEPLPVTPTRDVATSPISPTE
NNTTPPDALTRNTEKQKKKPKMSDEEILEKLRSIVSVGDPKKKYTRFEKIGQGASGTVYTAMDVATGQEVAIKQM
NLQQQPKKELIINEILVMRENKNPNIVNYLDSYLVGDELWVVMEYLAGGSLTDVVTETCMDEGQIAAVCRECLQA
LEFLHSNQVIHRDIKSDNILLGMDGSVKLTDFGFCAQITPEQSKRSTMVGTPYWMAPEVVTRKAYGPKVDIWSLG
IMAIEMIEGEPPYLNENPLRALYLIATNGTPELQNPEKLSAIFRDFLNRCLEMDVEKRGSAKELLQHQFLKIAKP
LSSLTPLIAAAKEATKNNH
```

SEQ ID NO:15 Human SPUVE nucleic acid sequence
HUM168767 accession:BC001278 coding sequence:121 . . . 1272

```
GACCCACGCGTCCGGGGGCAGGCATGGGAGCCGCGCGCTCTCTCCCGGCGCCCACACCTGTCTGAGCGGCGCAGC
GAGCCGCGGCCCGGGCGGGCTGCTCGGCGCGGAACAGTGCTCGGCATGGCAGGGATTCCAGGGCTCCTCTTCCTT
CTCTTCTTTCTGCTCTGTGCTGTTGGGCAAGTGAGCCCTTACAGTGCCCCCTGGAAACCCACTTGGCCTGCATAC
CGCCTCCCTGTCGTCTTGCCCCAGTCTACCCTCAATTTAGCCAAGCCAGACTTTGGAGCCGAAGCCAAATTAGAA
GTATCTTCTTCATGTGGACCCCAGTGTCATAAGGGAACTCCACTGCCCACTTACGAAGAGGCCAAGCAATATCTG
TCTTATGAAACGCTCTATGCCAATGGCAGCCGCACAGAGACGCAGGTGGGCATCTACATCCTCAGCAGTAGTGGA
GATGGGGCCCAACACCGAGACTCAGGGTCTTCAGGAAAGTCTCGAAGGAAGCGGCAGATTTATGGCTATGACAGC
AGGTTCAGCATTTTTGGGAAGGACTTCCTGCTCAACTACCCTTTCTCAACATCAGTGAAGTTATCCACGGGCTGC
ACCGGCACCCTGGTGGCAGAGAAGCATGTCCTCACAGCTGCCCACTGCATACACGATGGAAAAACCTATGTGAAA
GGAACCCAGAAGCTTCGAGTGGGCTTCCTAAAGCCCAAGTTTAAAGATGGTGGTCGAGGGGCCAACGACTCCACT
TCAGCCATGCCCGAGCAGATGAAATTTCAGTGGATCCGGGTGAAACGCACCCATGTGCCCAAGGGTTGGATCAAG
GGCAATGCCAATGACATCGGCATGGATTATGATTATGCCCTCCTGGAACTCAAAAAGCCCCACAAGAGAAAATTT
ATGAAGATTGGGGTGAGCCCTCCTGCTAAGCAGCTGCCAGGGGGCAGAATTCACTTCTCTGGTTATGACAATGAC
CGACCAGGCAATTTGGTGTATCGCTTCTGTGACGTCAAAGACGAGACCTATGACTTGCTCTACCAGCAATGCGAT
GCCCAGCCAGGGGCCAGCGGGTCTGGGGTCTATGTGAGGATGTGGAAGAGACAGCAGCAGAAGTGGGGAGCGAAAA
ATTATTGGCATTTTTTCAGGGCACCAGTGGGTGGACATGAATGGTTCCCCACAGGATTTCAACGTGGCTGTCAGA
ATCACTCCTCTCAAATATGCCCAGATTTGCTATTGGATTAAAGGAAACTACCTGGATTGTAGGGAGGGGTGACAC
AGTGTTCCCTCCTGGCAGCAATTAAGGGTCTTCATGTTCTTATTTAGGAGAGGCCAAATTGTTTTTTGTCATTG
GCGTGCACACGTGTGTGTGTGTGTGTGTGTGTGTGTAAGGTGTCTTATAATCTTTTACCTATTTCTTACAATTGC
AAGATGACTGGCTTTACTATTTGAAAACTGGTTTGTGTATCATATCATATATCATTTAAGCAGTTTGAAGGCATA
```

-continued

CTTTTGCATAGAAATAAAAAAAATACTGATTTGGGGCAATGAGGAATATTTGACAATTAAGTTAATCTTCACGTT
TTTGCAAACTTTGATTTTTATTTCATCTGAACTTGTTTCAAAGATTTATATTAAATATTTGGCATACAAGAGAAA
AAAAAAAAAAAA

SEQ ID NO:16 Human SPUVE polypeptide sequence
Protein sequence protein_id:gi12654869
MAGIPGLLFLLFFLLCAVGQVSPYSAPWKPTWPAYRLPVVLPQSTLNLAKPDFGAEAKLEVSSSCGPQCHKGTPL
PTYEEAKQYLSYETLYANGSRTETQVGIYILSSSGDGAQHRDSGSSGKSRRKRQIYGYDSRFSIFGKDFLLNYPF
STSVKLSTGCTGTLVAEKHVLTAAHCIHDGKTYVKGTQKLRVGFLKPKFKDGGRGANDSTSAMPEQMKFQWIRVK
RTHVPKGWIKGNANDIGMDYDYALLELKKPHKRKFMKIGVSPPAKQLPGGRIHFSGYDNDRPGNLVYRFCDVKDE
TYDLLYQQCDAQPGASGSGVYVRMWKRQQQKWERKIIGIFSGHQWVDMNGSPQDFNVAVRITPLKYAQICYWIKG
NYLDCREG

SEQ ID NO:17 Mouse SPUVE nucleic acid sequence
accession:BC018517 coding sequence:170 . . . 1318
GTGCTGGTGTCAGTACTAGACTCTTCTGGCAGGGTGACCTTCAATGACTCTGCCTTATGAGAAGATTCTCAGAAC
TGACACCCTCCCGGAGCCCAGGGACAGAGCTGCAGGAGAGGCCTGCCTGCCTGCCTGCCTGCCTGCATGCCTGCC
TTCCTATGTAGCACTCAGCATGGCTGGAATCCCGGGGCTCTTCATCCTTCTTGTCCTGCTCTGTGTGTTCATGCA
GGTGAGTCCCTACACCGTTCCGTGGAAACCCACATGGCCGGCTTATCGCCTCCCTGTAGTCTTGCCTCAGTCTAC
CCTCAACTTAGCTAAGGCAGACTTCGACGCCAAAGCGAAATTGGAGGTGTCCTCCTCATGTGGACCTCAGTGTCA
CAAGGGAACACCACTGCCCACCTACGAAGAGGCCAAGCAGTACCTTTCCTATGAAACCCTTTATGCCAATGGCAG
CCGCACAGAGACTCGGGTGGGCATCTACATCCTCAGCAATGGTGAAGGCAGGGCACGAGGCAGAGACTCGGAGGC
CACAGGGAGATCTCGCAGGAAGAGGCAGATTTATGGCTACGATGGCAGGTTTAGCATTTTTGGGAAGGACTTCCT
GCTCAATTATCCTTTCTCAACATCGGTGAAGTTGTCTACTGGCTGCACTGGCACCCTGGTGGCAGAGAAGCACGT
CCTCACTGCTGCCCACTGCATACACGATGGGAAAACCTATGTGAAAGGGACACAGAAACTCCGAGTGGGCTTCCT
GAAGCCCAAGTATAAAGATGGTGCCGGAGGGGACAACAGCTCGAGCTCAGCCATGCCAGACAAGATGAAGTTTCA
GTGGATCCGCGTGAAACGCACCCATGTGCCCAAGGGGTGGATCAAGGGCAATGCCAATGACATCGGCATGGATTA
TGACTACGCCCTGCTGGAACTCAAGAAACCCCACAAAAGACAGTTCATGAAGATTGGTGTGAGTCCTCCAGCGAA
GCAGCTCCCAGGGGGCAGGATCCACTTCTCTGGTTATGACAATGACCGGCCCGGCAATTTGGTGTACCGCTTCTG
TGATGTCAAAGATGAGACCTACGACCTTCTCTACCAGCAGTGTGACGCCCAGCCCGGGGCCAGTGGTTCAGGGGT
CTATGTGAGGATGTGGAAGAGACCACAGCAGAAATGGGAAAGAAAAATTATCGGCATCTTTTCAGGGCACCAGTG
GGTGGACATGAATGGCTCTCCACAGGATTTCAACGTGGCAGTTAGAATCACGCCTCTTAAATATGCCCAGATTTG
CTATTGGATTAAAGGAAACTACCTAGATTGCAGGGAGGGGTGACATGCGTCTTCTTGCCAGCACCAATGGTCTTT
TTGCACTCATTGTAGGAGAGGCTAGCTTTTTATCATTGACTCTTGTGGTGTGAGTCACATAGTATCTTTTACCTA
GTATTCTTCAAATGGCAAAAATTATTGGCTATATTATTTTAAAACTGTTGTGTGCGTTATAGCATTTAAGCAGTC
TGAAAGCATACTTTTGCATAGAGACTTTAAAGTATTCGGGTAATAGGGCCTATTTGACAAGGAAGTTAAACTTTC
AGTTTTTGGAGAATTCTAATTTTTGTCTGATCCAAACTTGCTTCAGAGGTTTATATCAAATACGTGACACACAGG
GAATATGAATTCTTATGTTTGTATATGTATATGTTTTCTTCTGAGAGTCATATATTGATATTTTTGTAATGTGTG
GTTATTATGCTTCCAGATAATGATAGCAAAGTCTTCAATAGGCAATTTATAATGTTTTGGATTCAAACATTTACG
TAGTAGTCCTTGAAGAGAACAATAATTTATTGGCTATATTGATACCCATATAAGACTGTATCTTACAGTGCACAG
AATTCCCACGCTGCTTTTAGTTTTGAAAATAAAACTTTCCCTTGTAAAAAAAAAAAAAAAA

SEQ ID NO:18 Mouse SPUVE polypeptide sequence
Protein sequence accession:gi20178034
MAGIPGLFILLVLLCVFMQVSPYTVPWKPTWPAYRLPVVLPQSTLNLAKADFDAKAKLEVSSSCGPQCHKGTPLP
TYEEAKQYLSYETLYANGSRTETRVGIYILSNGEGRARGRDSEATGRSRRKRQIYGYDGRFSIFGKDFLLNYPFS
TSVKLSTGCTGTLVAEKHVLTAAHCIHDGKTYVKGTQKLRVGFLKPKYKDGAGGDNSSSSAMPDKMKFQWIRVKR
THVPKGWIKGNANDIGMDYDYALLELKKPHKRQFMKIGVSPPAKQLPGGRIHFSGYDNDRPGNLVYRFCDVKDET
YDLLYQQCDAQPGASGSGVYVRMWKRPQQKWERKIIGIFSGHQWVDMNGSPQDFNVAVRITPLKYAQICYWIKGN
YLDCREG

SEQ ID NO:19 Human NK4 nucleic acid sequence
HUM241786 accession:BC018782 CDS:59 . . . 625
GACTGTCTCAGTGGAGCTGGGTCATCTCAGGCCTTGGCTCCTTGAACTTTTGGCCGCCATGTGCTTCC
CGAAGGTCCTCTCTGATGACATGAAGAAGCTGAAGGCCCGAATGCACCAGGCCATAGAAAGATTTTAT
GATAAAATGCAAAATGCAGAATCAGGACGTGGACAGGTGATGTCGAGCCTGGCAGAGCTGGAGGACGA
CTTCAAAGAGGGCTACCTGGAGACAGTGGCGGCTTATTATGAGGAGCAGCACCCAGAGCTCACTCCTC
TACTTGAAAAAGAAAGAGATGGATTACGGTGCCGAGGCAACAGATCCCCTGTCCCGGATGTTGAGGAT
CCCGCAACCGAGGAGCCTGGGGAGAGCTTTTGTGACAAGGTCATGAGATGGTTCCAGGCCATGCTGCA
GCGGCTGCAGACCTGGTGGCACGGGGTTCTGGCCTGGGTGAAGGAGAAGGTGGTGGCCCTGGTCCATG
CAGTGCAGGCCCTCTGGAAACAGTTCCAGAGTTTCTGCTGCTCTCTGTCAGAGCTCTTCATGTCCTCT
TTCCAGTCCTACGGAGCCCCACGGGGGGACAAGGAGGAGCTGACACCCCAGAAGTGCTCTGAACCCCA
ATCCTCAAAATGAAGATACTGACACCACCTTTGCCCTCCCCGTCACCGCGCACCCACCCTGACCCCTC
CCTCAGCTGTCCTGTGCCCCGCCCTCTCCCGCACACTCAGTCCCCCTGCCTGGCGTTCCTGCCGCAGC
TCTGACCTGGTGCTGTCGCCCTGGCATCTTAATAAAACCTGCTTATACTTCCCTGGAAAAAAAAAAAA
AAAAAAAAAAA

SEQ ID NO:20 Human NK4 polypeptide sequence
protein_id:gi17511868
MCFPKVLSDDMKKLKARMHQAIERFYDKMQNAESGRGOVMSSLAELEDDFKEGYLETVAAYYEEQHPE
LTPLLEKERDGLRCRGNRSPVPDVEDPATEEPGESFCDKVMRWFQAMLQRLQTWWHGVLAWVKEKVVA
LVHAVQALWKQFQSFCCSLSELFMSSFQSYGAPRGDKEELTPQKCSEPQSSK

SEQ ID NO:21 Human Protein C inhibitor nucleic acid sequence
Nucleotide sequence HUM207659 accession:BC008915 CDS:140 . . . 1360
GGCACGAGGCTCAGCTACGGGACACATTTCAGGTATCCAAGGCAGCAGAGGTGAGTGGGTCCCCCGAG
CTCTGTGACCTTATGCTCCACACTAACTCTGGCAGAGCCTCCGTTTCCTCATAGAACAAAGAACAGCC

-continued

ACCATGCAGCTCTTCCTCCTCTTGTGCCTGGTGCTTCTCAGCCCTCAGGGGGCCTCCCTTCACCGCCA
CCACCCCCGGGAGATGAAGAAGAGAGTCGAGGACCTCCATGTAGGTGCCACGGTGGCCCCCAGCAGCA
GAAGGGACTTTACCTTTGACCTCTACAGGGCCTTGGCTTCCGCTGCCCCCAGCCAGAACATCTTCTTC
TCCCCTGTGAGCATCTCCATGAGCCTGGCCATGCTCTCCCTGGGGGCTGGGTCCAGCACAAAGATGCA
GATCCTGGAGGGCCTGGGCCTCAACCTCCAGAAAAGCTCAGAGAAGGAGCTGCACAGAGGCTTTCAGC
AGCTCCTTCAGGAACTCAACCAGCCCAGAGATGGCTTCCAGCTGAGCCTCGGCAATGCCCTTTTCACC
GACCTGGTGGTAGACCTGCAGGACACCTTCGTAAGTGCCATGAAGACGCTGTACCTGGCAGACACTTT
CCCCACCAACTTTAGGGACTCTGCAGGGGCCATGAAGCAGATCAATGATTATGTGGCAAAGCAAACGA
AGGGCAAGATTGTGGACTTGCTTAAGAACCTCGATAGCAATGCGGTCGTGATCATGGTGAATTACATC
TTCTTTAAAGCTAAGTGGGAGACAAGCTTCAACCACAAAGGCACCCAAGAGCAAGACTTCTACGTGAC
CTCGGAGACTGTGGTGCGGGTACCCATGATGAGCCGCGAGGATCAGTATCACTACCTCCTGGACCGGA
ACCTCTCCTGCAGGGTGGTGGGGGTCCCCTACCAAGGCAATGCCACGGCTTTGTTCATTCTCCCCAGT
GAGGGAAAGATGCAGCAGGTGGAGAATGGACTGAGTGAGAAAACGCTGAGGAAGTGGCTTAAGATGTT
CAAAAAGAGGCAGCTCGAGCTTTACCTTCCCAAATTCTCCATTGAGGGCTCCTATCAGCTGGAGAAAG
TCCTCCCCAGTCTGGGGATCAGTAACGTCTTCACCTCCCATGCTGATCTGTCCGGCATCAGCAACCAC
TCAAATATCCAGGTGTCTGAGATGGTGCACAAAGCTGTGGTGGAGGTGGACGAGTCGGGAACCAGAGC
AGCGGCAGCCACGGGGACAATCTTCACTTTCAGGTCGGCCCGCCTGAACTCTCAGAGGCTAGTGTTCA
ACAGGCCCTTTCTGATGTTCATTGTGGATAACAACATCCTCTTCCTTGGCAAAGTGAACCGCCCCTGA
GGTGGGGCTTCTCCTGAAATCTACAGGCCTCAGGGTGGGAGATGAAGGGGGCTATGCTATGGCCCATC
TGTATGCTGGTAGCTAGTGATTTACACAGGTTTAGTTGACTAATGAGGCATTACAAATAATATTACTC
TATGATGATTGCTTCCACCCACACGACTGCAACATACAGGTGCCTTGGGGAAATGTGGAGAACATTCA
ATCTTGCCGTCACTATTCATCAATGAAGATTAGCACTGAGATCCAGAGAGGCTGGATGACTTGCTCAA
GTTCACCAGCATGGTAGTGGCAAAGAGAGGTCCAGAGTCCTGGCCCTTGATGCCCAGCTCAGTGCCAC
AAAGCTCAGTAGGAGGGATGTTCCAGTGGATGAGGGCCACCAGGAAGCACAGGTCCAAGGCTGGTCCC
ACACTTATCAGCAGCAACAACTGTCAGTTCATCCTGCATGGGAAAAATGTTGGAATGGGAGTCTGAAA
TGGGGCTACTGTTTCAGTCCTAACGTGCTGTGTGACATTGGGACAACACTTTCCCTCTCTGGACCTCA
GTTTCCCTCTGTATACAAGGATCAGATTCTTGCTGTGACCCAAGAACTCCTGAAATCATATAGAAAGG
CTGGGGTGGGCCCTGTCATTCGTGGTTGATTTCAATACACTCAAGTGCCATTCATCCTTTAAGAAAAA
CATCTGGATATCAAGGTGGAAATGGCCCATTTAATGATTGATTATATCATTTTGTGGATATAGTTATA
ATCTGATGGGCCTGGCTGGGAGTGGAAGAAGGGAAGCCTTTTGCAPATAGTAGAGTGTCAGTTGCAGG
TGCCAATGACTAACTTTTTGAATTCTATGTTGGCATTAACAATAAAGCATTTTGCAAACAAAAAAAAA
AAAAAAAAAA

SEQ ID NO:22 HumanProtein C inhibitor polypeptide sequence
protein_id:gi14286238
MQLFLLLCLVLLSPQGASLHRHHPREMKKRVEDLHVGATVAPSSRRDFTFDLYRALASAAPSQNIFFS
PVSISMSLAMLSLGAGSSTKMQILEGLGLNLQKSSEKELHRGFQQLLQELNQPRDGFQLSLGNALFTD
LVVDLQDTFVSAMKTLYLADTFPTNFRDSAGAMKQINDYVAKQTKGKIVDLLKNLDSNAVVIMVNYIF
FKAKWETSFNHKGTQEQDFYVTSETVVRVPMMSREDQYHYLLDRNLSCRVVGVPYQGNATALFILPSE
GKMQQVENGLSEKTLRKWLKMFKKRQLELYLPKFSIEGSYQLEKVLPSLGISNVFTSHADLSGISNHS
NIQVSEMVHKAVVEVDESGTRAAAATGTIFTFRSARLNSQRLVFNRPFLMFIVDNNILFLGKVNRP

SEQ ID NO:23 Mouse Protein C inhibitor nucleic acid sequence
accession:NM_008785 CDS:125 . . . 1342
CAGAGTCTCTGTGGTCAGTGACATTTGCCGGCTGCAATTCCTCTGCTCTGCTCTCTGTCCTCTTGTAA
GCCAGGTGGACTATACCTTGACACTCAGCCACCACGATACAGAAGCTGTCACCACAATGAGGTTCTTC
CCCATTCTGTGCCTGGTGCTGTTCATCAGCCATGGGGTGGCTTCCCGCCGACACTCCCATTCCAAGAA
GAAGAAGGCTAAAGAGTCCTCGGTGGGTGCTGTGGGACCTCCCAGTAGCAAAGACTTTGCTTTCAGAC
TCTACAGGGCCTTGGCTTCTGAATCCCCTGGTCAGAATGTCTTCTTCTCCCCCTTGAGCGTGTCTATG
AGTTTGGGCATGCTCTCCCTGGGGGCTGGCTTGAAGACGAAGACCCAGATCCTAGATGGCCTAGGCCT
CAGCCTCCAGCAAGGCCAAGAAGACAAGCTCCACAAGGGCTTCCAACAGCTGCTACAGAGATTCAGGC
AGCCTAGTGATGGCCTGCAGCTGAGCCTGGGCAGTGCCCTTTTTAAAGACCCAGCAGTACATATCCGG
GACGACTTCCTGAGTGCCATGAAGACACTGTACATGTCAGACACTTTCTCTACCAACTTTGGGAACCC
TGAAATTGCCAAGAAGCAGATCAACAACTATGTAGCCAAGCAGACCAAGGGCAAGATTGTAGACTTGA
TCAAGGACCTTGACAGCACCCATGTCATGATAGTGGTAAATTACATCTTCTTCAAAGCCAAGTGGCAG
ACGGCCTTCAGTGAGACAAACACCCACAAGATGGATTTCCATGTGACCCCCAAAAAGACCATACGGGT
GCCCATGATGAACCGTGAAGATGAGTATTCCTACTACCTGGACCAAAACATCTCCTGCACGGTGGTGG
GGATCCCTTATCAAGGCAATGCCATTGCTCTATTTATTCTCCCCAGCGAGGGCAAGATGAAGCAGGTG
GAGGATGGCCTGGACGAGAGAACATTGAGGAACTGGCTTAAGATGTTCACCAAGAGACGCTTAGATCT
TTACCTCCCCAAGTTCTCCATTGAGGCTACCTACAAACTGGAAAATGTCCTCCCCAAGCTGGGCATCC
AGGACGTCTTCACCACCCATGCTGACTTGTCTGGCATTACTGACCATACCAATATCAAGTTGTCTGAG
ATGGTGCACAAATCCATGATGGAGGTTGAAGAGTCAGGAACCACAGCAGCTGCCATCACAGGAGCCAT
CTTCACATTCAGATCTGCTCGGCCGAGCTCCCTGAAGATAGAATTCACCAGACCCTTTCTGCTGACCC
TTATGGAGGATTCACATATACTTTTCGTTGGCAAGGTGACCCGGCCCTGAGGTGGGACTCCTTCTGAA
ACTCATGGGCTTACACAGAGGGAGCCAGGCATGTTCTTCAGCCCATCCATCTCTTGGTTGCTAGTGAT
TTTCATAGAGTAGGTTGACTAGTAAGGTGTTGTAAGTGATAAGATCAATATCCCTATGGCTGCTTCCT
TGATCACAGTTGCAAGATGTGTTCTTGCTCTCATCCTTCATCATTGACACTGACCCTAAAGAGGTCAC
TTGATTTCTCCAAGGTCCTACCATCTTATATGTGACATAGTCGTGCCAAGACTCCAGGCCTCTCCTGC
TCAGTCTAGCTCTACAAAGTTCCCTAAGAGACTGGTCCGCATAGATAAAGGTCACCAGGAAGCACAGT
CATAAACCTGGTCCCCCACCTATCAGCACCAACTCTCAGTTCATCAGGGTTGAGAAAGGACACTGGAT
TGAGCTTCTGAAATGCCGCTGGTAGCTCTGTAACATGGGAATGGAAATGTCCCTCTCTGGGCCTGTGT
TTCTTTGTGTGTCCAGTGAAGGAACAGAACTAGAAGCCCATTCAGCAACCTGTGAGATTCTTACACAG
GGGTTGGCTTCTCTATGTTAGCCAGGGTTGCTGTAAGAACTCAGGTTAGTAGCATACTTGAGTAATCA
GGGTCAAAGTGTCTCACCACAAACAAAGGTATCAAATCACAAGCCTGGCTGGCTGAGGTGGTGTGGTA
GAGGTCTGTGGTAAACAGCAGAACTTAGGGTGCAGACTGACTTATTGATTTCCGCCTTGGCTGGCAAT
AAAGGCATTTTGCAAAG

-continued

SEQ ID NO:24 Mouse Protein C inhibitor polypeptide sequence
protein_id:6679215
MRFFPILCLVLFISHGVASRRHSHSKKKKAKESSVGAVGPPSSKDFAFRLYRALASESPGQNVFFSPL
SVSMSLGMLSLGAGLKTKTQILDGLGLSLQQGQEDKLHKGFQQLLQRFRQPSDGLQLSLGSALFKDPA
VHIRDDFLSAMKTLYMSDTFSTNFGNPEIAKKQINNYVAKQTKGKIVDLIKDLDSTHVMIVVNYIFFK
AKWQTAFSETNTHKMDFHVTPKKTIRVPMMNREDEYSYYLDQNISCTVVGIPYQGNAIALFILPSEGK
MKQVEDGLDERTLRNWLKMFTKRRLDLYLPKFSIEATYKLENVLPKLGIQDVFTTHADLSGITDHTNI
KLSEMVHKSMMEVEESGTTAAAITGAIFTFRSARPSSLKIEFTRPFLLTLMEDSHILFVGKVTRP

SEQ ID NO:25 Rat Protein C nucleic acid sequence
accession:NM_022957 CDS:48 . . . 1268
TACCCGTCAAGACACTCAGCCACCACAATACAGAAATGGCAGCCACAATGAGGTTCTTCCCCATTCTG
TGCCTGGTGCTGTTCTTCAGTCACGGGGTGGCTTCCCGCCAACGCTCCCATTCTAAGGAGAAGAAGAA
GTCTAAAGAGTCCTCTGTGGGTGCTGTGGGGACTTCCCGAAGCAGAGACTTTGCCTTCAGGCTCTACA
GGGCCTTGGCTTCTGAAGCCCCTGGTCAGAATGTTTTCTTCTCCCCCATGAGCGTGTCTATGAGCTTG
GGTATGCTCTCCCTGGGGTCTGGCTTGAAGACTAAGGCACAGATCCTGGAGGGCCTGGGCCTCAGCCT
CCAGCAAGGCCAAGAGGATATGCTCCACAAGGGCTTCCAACAGCTGCTGCTGCAGTTCAGCCAGCCTA
GTGATGGCCTCCAGCTGAGCCTAGGCAGTGCCCTTTTTACAGACCCAGCAGTACATATCCGGGACCAC
TTTCTGAGTGCCATGAAGACATTGTACATGTCAGACATGTTCTCTACCAACTTTGGGAACCCTGAAAG
TGCTAAGAAGCAGATCAATGACTATGTAGCCAAGAAGACCAACGGCAAGATTGTAGACTTGATCAAGG
ATCTTGACAGCACCCACGTCATGGTGGTGGTAAATTATATCTTCTTCAAAGCCAAGTGGCAGACGGCC
TTCAGTAGCACCAACACCCACAAGATGGATTTCCATGTGACTCCCAAGAAGACCATTCAGGTGCCAAT
GATGAACCGCGAAGATATATATTCCTACATCCTCGACCAAAACATTTCCTGCACGGTGGTGGGGATCC
CTTACCAAGGCAATACATTTGCTCTGTTTATTCTTCCCAGCGAGGGCAAGATGAAGCGGGTAGAGGAT
GGCCTGGATGAGAGAACGTTGAGGAACTGGCTTAAGATGTTCACAAAGAGACAGTTAGATCTTTACCT
CCCCAAGTTCTCCATTGAGGGTACCTATAAACTGGAAAAAATCCTCCCCAAGCTGGGCATCCAGGACA
TCTTCACCACCCATGCTGACTTGTCTGGCCTTACTGACCACACCAATATTAAGTTGTCTGAGATGGTA
CACAAATCCATGGTGGAGGTAGACGAGTCAGGAACAACAGCAGCTGCCTCCACAGGAATTCTCTTCAC
ACTCAGATCTGCTCGACCGAGCTCTCTGAAGGTAGAATTCACCAGACCCTTTCTGGTGGTCATTATGG
ATGGTACAAACCTCTATTTCATTGGCAAAGTGATTCAGCCCTGAGGTGGGACTCCTTCTGAAACTCAC
AGGCCTATACAGAAGGAGCCAGGCATATTCTACAGACTATCCACCTCTTGGTAGCTAGTGATTTTCAT
AGAGTAGATTGACTAATAAGGTGTTGTAAGTGAGAAGATCAATATCCTGGTGGAAGCTTCCTTGATCA
CAGTTGCAAGATGTGCTCTTGCTCTATTTTTTCATTATTGACACTGATGCTAAAGAGGCCACTTGATT
GCTCCAAGGTCCTATCACCATATTTGTGACATAGTTATGCCCAGACCTCAGACCTCTCTTGCTCAGTC
TAGCGCTACAAAGCTCTGCAAGAGGCTGGTCAGCATAGATAAAGGCCACCAGGAAGCACGGTTATAAA
CCTGGTCTCACACCTATTAGCACCATGGTTGAGAAAGGACATTGGACTGAGATTCTGAAATGGTGCTA
ATAGCTATGTAACTCGAAAATGAAAATGTTCCTCTCTGCCTGTGTTTCTTTGAGTATCCAGTCAATGA
TCAGAACAAGTAGCCCATTCAGTGATCTGTGAAATTGTTACACAAGGGTTGGCTACTCTATGCTAGCC
ATGGTTGATGTAAGAACTCAGTTTAATAGCATACTTGAGCAATCAGGGTAAAAGTGTTCCACCATGAA
CAAAGGTATCAAATCACAAACCTGGCTGACTGCAGTTGGATGTTAGAGGTCTGTTGTAAATAGTAGAA
CATAGATTGCAGACTGACTTGTTGATTTCTACTTTGGCAATAGCAATAAAGCACTTTGCAAAG

SEQ ID NO:26 Rat Protein C inhibitor polypeptide sequence
protein_id:12621138 CDS:171 . . . 1265
MRFFPILCLVLFFSHGVASRQRSHSKEKKKSKESSVGAVGTSRSRDFAFRLYRALASEAPGQNVFFSP
MSVSMSLGMLSLGSGLKTKAQILEGLGLSLQQGQEDMLHKGFQQLLQQFSQPSDGLQLSLGSALFTDP
AVHIRDHFLSAMKTLYMSDMFSTNFGNPESAKKQINDYVAKKTNGKIVDLIKDLDSTHVMVVVNYIFF
KAKWQTAFSSTNTHKMDFHVTPKKTIQVPMMNREDIYSYILDQNISCTVVGIPYQGNTFALFILPSEG
KMKRVEDGLDERTLRNWLKMFTKRQLDLYLPKFSIEGTYKLEKILPKLGIQDIFTTHADLSGLTDHTN
IKLSEMVHKSMVEVDESGTTAAASTGILFTLRSARPSSLKVEFTRPFLVVIMDGTNLYFIGKVIQP

SEQ ID NO:27 Human MAST205b nucleic acid sequence
Nucleotide sequence Novel Variant MAST205b CDS:1–5073
ATGTTTTCACCCACATCTGCTCCAGCCCTCTTCCTCACTAAAGTCCCATTTAGTGCTGATTGTGCTTT
GGCTACTTCTCCTCTTGCCATTTTCCTGAACCCACGAGCCCACAGCAGTCCTGGCACTCCTTGTTCCA
GCCGCCCACTGCCGTGGAGTTGTCGGACAAGTAACCGCAAGAGCTTGATTGTGACCTCTAGCACATCA
CCTACACTACCACGGCCACACTCACCACTCCATGGCCACACAGGTAACAGTCCTTTGGACAGCCCCCG
GAATTTCTCTCCAAATGCACCTGCTCACTTTTCTTTTGTTCCTGCCCGTAGCCATAGCCACAGAGCTG
ACAGGACTGATGGGCGGCGCTGGTCTTTGGCCTCTTTGCCCTCTTCAGGATATGGAACTAACACTCCT
AGCTCCACTGTCTCATCATCATGCTCCTCACAGGAAAAGCTGCATCAGTTGCTTTTCCAGCCTACAGC
TGATGAGCTGCACTTTTTGACGAAGCATTTCAGCACAGAGAGCGTACCAGATGAGGAAGGACGGCAGT
CCCCAGCCATGCGGCCTCGCTCCCGGAGCCTCAGTCCCGGACGATCCCCAGTATCCTTTGACAGTGAA
ATAATAATGATGAATCATGTTTACAAAGAAAGATTCCCAAAGGCCACCGCACAAATGGAAGAGCGACT
AGCAGAGTTTATTTCCTCCAACACTCCAGACAGCGTGCTGCCCTTGGCAGATGGAGCCCTGAGCTTTA
TTCATCATCAGGTGATTGAGATGGCCCGAGACTGCCTGGATAAATCTCGGAGTGGCCTCATTACATCA
CAATACTTCTACGAACTTCAAGAGAATTTGGAGAAACTTTTACAAGATGCTCATGAGCGCTCAGAGAG
CTCAGAAGTGGCCTTTGTGATGCAGCTGGTGAAAAAGCTGATGATTATCATTGCCCGCCCAGCACGTC
TCCTGGAATGCCTGGAGTTTGACCCTGAAGAGTTCTACCACCTTTTAGAAGCAGCTGAGGGCCACGCC
AAAGAGGGACAAGGGATTAAATGTGACATTCCCCGCTACATCGTTAGCCAGCTGGGCCTCACCCGGGA
TCCCCTAGAAGAAATGGCCCAGTTGAGCAGCTGTGACAGTCCTGACACTCCAGAGACAGATGATTCTA
TTGAGGGCCATGGGGCATCTCTGCCATCTAAAAAGACACCCTCTGAAGAGGACTTCGAGACCATTAAG
CTCATCAGCAATGGCGCCTATGGGGCTGTATTTCTGGTGCGGCACAAGTCCACCCGGCAGCGCTTTGC
CATGAAGAAGATCAACAAGCAGAACCTGATCCTACGGAACCAGATCCAGCAGGCCTTCGTGGAGCGTG
ACATACTGACTTTCGCTGAGAACCCCTTTGTGGTCAGCATGTTCTGCTCCTTTGATACCAAGCGCCAC
TTGTGCATGGTGATGGAGTACGTTGAAGGGGGAGACTGTGCCACTCTGCTGAAGAATATTGGGGCCCT
GCCTGTGGACATGGTGCGTCTATACTTTGCGGAAACTGTGCTGGCCCTGGAGTACTTACACAACTATG
GCATCGTGCACCGTGACCTCAAGCCTGACAACCTCCTAATTACATCCATGGGGCACATCAAGCTCACG

-continued

GACTTTGGACTGTCCAAAATGGGCCTCATGAGTCTGACAACGAACTTGTATGAGGGTCATATTGAAAA
GGATGCCCGGGAATTCCTGGACAAGCAGGTATGCGGACCCCAGAATACATTGCGCCTGAGGTGATCC
TGCGCCAGGGCTATGGGAAGCCAGTGGACTGGTGGGCCATGGGCATTATCCTGTATGAGTTCCTGGTG
GGCTGCGTCCCTTTTTTTGGAGATACTCCGGAGGAGCTCTTTGGGCAGGTGATGAGTGATGAGATTGT
GTGGCCTGAGGGTGATGAGGCACTGCCCCCAGACGCCCAGGACCTCACCTCCAAACTGCTCCACCAGA
ACCCTCTGGAGAGACTTGGCACAGGCAGTGCCTATGAGGTGAAGCAGCACCCATTCTTTACTGGTCTG
GACTGGACAGGACTTCTCCGCCAGAAGGCTGAATTTATTCCTCAGTTGGAGTCAGAGGATGATACTAG
CTATTTTGACACCCGCTCAGAGCGATACCACCACATGGACTCGGAGGATGAGGAAGAAGTGAGTGAGG
ATGGCTGCCTTGAGATCCGCCAGTTCTCTTCCTGCTCTCCAAGGTTCAACAAGGTGTACAGCAGCATG
GAGCGGCTCTCACTGCTCGAGGAGCGCCGGACACCACCCCCGACCAAGCGCAGCCTGAGTGAGGAGAA
GGAGGACCATTCAGATGGCCTGGCAGGGCTCAAAGGCCGAGACCGGAGCTGGGTGATTGGCTCCCCTG
AGATATTACGGAAGCGGCTGTCGGTGTCTGAGTCGTCCCACACAGAGAGTGACTCAAGCCCTCCAATG
ACAGTGCGACGCCGCTGCTCAGGCCTCCTGGATGCGCCTCGGTTCCCGGAGGGCCCTGAGGAGGCCAG
CAGCACCCTCAGGAGGCAACCACAGGAGGGTATATGGGTCCTGACACCCCCATCTGGAGAGGGGGTAT
CTGGGCCTGTCACTGAACACTCAGGGGAGCAGCGGCCAAAGCTGGATGAGGAAGCTGTTGGCCGGAGC
AGTGGTTCCAGTCCAGCTATGGAGACCCGAGGCCGTGGGACCTCACAGCTGGCTGAGGGAGCCACAGC
CAAGGCCATCAGTGACCTGGCTGTGCGTAGGGCCCGCCACCGGCTGCTCTCTGGGGACTCAACAGAGA
AGCGCACTGCTCGCCCTGTCAACAAAGTGATCAAGTCCGCCTCAGCCACAGCCCTCTCACTCCTCATT
CCTTCGGAACACCACACCTGCTCCCCGTTGGCCAGCCCCATGTCCCCACATTCTCAGTCGTCCAACCC
ATCATCCCGGGACTCTTCTCCAAGCAGGGACTTCTTGCCAGCCCTTGGCAGCATGAGGCCTCCCATCA
TCATCCACCGAGCTGGCAAGAAGTATGGCTTCACCCTGCGGGCCATTCGCGTCTACATGGGTGACTCC
GATGTCTACACCGTGCACCATATGGTGTGGCACGTGGAGGATGGAGGTCCGGCCAGTGAGGCAGGGCT
TCGTCAAGGTGACCTCATCACCCATGTCAATGGGGAACCTGTGCATGGCCTGGTGCACACGGAGGTGG
TGGAGCTGATCCTGAAGAGTGGAAACAAGGTGGCCATTTCAACAACTCCCCTGGAGAACACATCCATT
AAAGTGGGGCCAGCTCGGAAGGGCAGCTACAAGGCCAAGATGGCCCGAAGGAGCAAGAGGAGCCGCGG
CAAGGATGGGCAAGAAAGCAGAAAAAGGAGCTCCCTGTTCCGCAAGATCACCAAGCAAGCATCCCTGC
TCCACACCAGCCGCAGCCTTTCTTCCCTTAACCGCTCCTTGTCATCAGGGGAGAGTGGGCCAGGCTCT
CCCACACACAGCCACAGCCTTTCCCCCCGATCTCCCACTCAAGGCTACCGGGTGACCCCCGATGCTGT
GCATTCAGTGGGAGGGAATTCATCACAGAGCAGCTCCCCCAGCTCCAGCGTGCCCAGTTCCCCAGCCG
GCTCTGGGCACACACGGCCCAGCTCCCTCCACGGTCTGGCACCCAAGCTCCAACGCCAGTACCGCTCT
CCACGGCGCAAGTCAGCAGGCAGCATCCCACTGTCACCACTGGCCCACACCCCTTCTCCCCCACCCCC
AACAGCTTCACCTCAGCGGTCCCCATCGCCCCTGTCTGGCCATGTAGCCCAGGCCTTTCCCACAAAGC
TTCACTTGTCACCTCCCCTGGGCAGGCAACTCTCACGGCCCAAGAGTGCGGAGCCACCCCGTTCACCA
CTACTCAAGAGGGTGCAGTCGGCTGAGAAACTGGCAGCAGCACTTGCCGCCTCTGAGAAGAAGCTAGC
CACTTCTCGCAAGCACAGCCTTGACCTGCCCCACTCTGAACTAAAGAAGGAACTGCCGCCCAGGGAAG
TGAGCCCTCTGGAGGTAGTTGGAGCCAGGAGTGTGCTGTCTGGCAAGGGGGCCCTGCCAGGGAAGGGG
GTGCTGCAGCCTGCTCCCTCACGGGCCCTAGGCACCCTCCGGCAGGACCGAGCCGAACGACGGGAGTC
GCTGCAGAAGCAAGAAGCCATTCGTGAGGTGGACTCCTCAGAGGACGACACCGAGGAAGGGCCTGAGA
ACAGCCAGGGTGCACAGGAGCTGAGCTTGGCACCTCACCCAGAAGTGAGCCAGAGTGTGGCCCCTAAA
GGAGCAGGAGAGAGTGGGGAAGAGGATCCTTTCCCGTCCAGAGGCCCTAGGAGCCTGGGCCCAATGGT
CCCAAGCCTATTGACAGGGATCACACTGGGGCCTCCCAGAATGGAAAGTCCCAGTGGTCCCCACAGGA
GGCTCGGGAGCCCACAAGCCATTGAGGAGGCTGCCAGCTCCTCCTCAGCAGGCCCCAACCTAGGTCAG
TCTGGAGCCACAGACCCCATCCCTCCTGAAGGTTGCTGGAAGGCCCAGCACCTCCACACCCAGGCACT
AACAGCACTTTCTCCCAGCACTTCGGGACTCACCCCCACCAGCAGTTGCTCTCCTCCCAGCTCCACCT
CTGGGAAGCTGAGCATGTGGTCCTGGAAATCCCTTATTGAGGGCCCAGACAGGGCATCCCCAAGCAGA
AAGGCAACCATGGCAGGTGGGCTAGCCAACCTCCAGGATTTGGAAAACACAACTCCAGCCCAGCCTAA
GAACCTGTCTCCCAGGGAGCAGGGGAAGACACAGCCACCTAGTGCCCCCAGACTGGCCCATCCATCTT
ATGAGGATCCCAGCCAGGGCTGGCTATGGGAGTCTGAGTGTGCACAAGCAGTGAAAGAGGATCCAGCC
CTGAGCATCACCCAAGTGCCTGATGCCTCAGGTGACAGAAGGCAGGACGTTCCATGCCGAGGCTGCCC
CCTCACCCAGAAGTCTGAGCCCAGCCTCAGGAGGGGCCAAGAACCAGGGGGCCATCAAAAGCATCGGG
ATTTGGCATTGGTTCCAGATGAGCTTTTAAAGCAAACATAG

SEQ ID NO:28 Human MAST205b novel variant polypeptide sequence

MFSPTSAPALFLTKVPFSADCALATSPLAIFLNPRAHSSPGTPCSSRPLPWSCRTSNRKSLIVTSSTS
PTLPRPHSPLHGHTGNSPLDSPRNFSPNAPAHFSFVPARSHSHRADRTDGRRWSLASLPSSGYGTNTP
SSTVSSSCSSQEKLHQLLFQPTADELHFLTKHFSTESVPDEEGRQSPAMRPRSRSLSPGRSPVSFDSE
IIMMNHVYKERFPKATAQMEERLAEFISSNTPDSVLPLADGALSFIHHQVIEMARDCLDKSRSGLITS
QYFYELQENLEKLLQDAHERSESSEVAFVMQLVKKLMIIIARPARLLECLEFDPEEFYHLLEAAEGHA
KEGQGIKCDIPRYIVSQLGLTRDPLEEMAQLSSCDSPDTPETDDSIEGHGASLPSKKTPSEEDFETIK
LISNGAYGAVFLVRHKSTRQRFAMKKINKQNLILRNQIQQAFVERDILTFAENPFVVSMFCSFDTKRH
LCMVMEYVEGGDCATLLKNIGALPVDMVRLYFAETVLALEYLHNYGIVHRDLKPDNLLITSMGHIKLT
DFGLSKMGLMSLTTNLYEGHIEKDAREFLDKQVCGTPEYIAPEVILRQGYGKPVDWWAMGIILYEFLV
GCVPFFGDTPEELFGQVISDEIVWPEGDEALPPDAQDLTSKLLHQNPLERLGTGSAYEVKQHPFFTGL
DWTGLLRQKAEFIPQLESEDDTSYFDTRSERYHHMDSEDEEEVSEDGCLEIRQFSSCSPRFNKVYSSM
ERLSLLEERRTPPPTKRSLSEEKEDHSDGLAGLKGRDRSWVIGSPEILRKRLSVSESSHTESDSSPPM
TVRRRCSGLLDAPRFPEGPEEASSTLRRQPQEGIWVLTPPSGEGVSGPVTEHSGEQRPKLDEEAVGRS
SGSSPAMETRGRGTSQLAEGATAKAISDLAVRRARHRLLSGDSTEKRTARPVNKVIKSASATALSLLI
PSEHHTCSPLASPMSPHSQSSNPSSRDSSPSRDFLPALGSMRPPIIIHRAGKKYGFTLRAIRVYMGDS
DVYTVHHMVWHVEDGGPASEAGLRQGDLITHVNGEPVHGLVHTEVVELILKSGNKVAISTTPLENTSI
KVGPARKGSYKAKMARRSKRSRGKDGQESRKRSSLFRKITKQASLLHTSRSLSSLNRSLSSGESGPGS
PTHSHSLSPRSPTQGYRVTPDAVHSVGGNSSQSSSPSSSVPSSPAGSGHTRPSSLHGLAPKLQRQYRS
PRRKSAGSIPLSPLAHTPSPPPPTASPQRSPSPLSGHVAQAFPTKLHLSPPLGRQLSRPKSAEPPRSP
LLKRVQSAEKLAAALAASEKKLATSRKHSLDLPHSELKKELPPREVSPLEVVGARSVLSGKGALPGKG
VLQPAPSRALGTLRQDRAERRESLQKQEAIREVDSSEDDTEEGPENSQGAQELSLAPHPEVSQSVAPK
GAGESGEEDPFPSRGPRSLGPMVPSLLTGITLGPPRMESPSGPHRRLGSPQAIEEAASSSSAGPNLGQ

-continued

SGATDPIPPEGCWKAQHLHTQALTALSPSTSGLTPTSSCSPPSSTSGKLSMWSWKSLIEGPDRASPSR
KATMAGGLANLQDLENTTPAQPKNLSPREQGKTQPPSAPRLAHPSYEDPSQGWLWESECAQAVKEDPA
LSITQVPDASGDRRQDVPCRGCPLTQKSEPSLRRGQEPGGHQKHRDLALVPDELLKQT

SEQ ID NO:29 Human MAST205 nucleic acid sequence
Nucleotide sequence HUM163512 accession:AB047005 CDS:284 . . . 5488
TAGGCAGGCGGCTGAGCCGGCGGCGGGTGGCCTGCCCAACGTGTGCTGGGTGGGAGAAGGCGAGGCGG
CAGCGATGCTGTCTCTTCCGTGAGGAGCGCAGAGGAGGTCGCGGCGCCGGAGGCCCCAGAAGGCTCGA
AGGCGCCGCGGGCTGGGGTCGGTGGCTTAGGGAGCCCGTCCGGCCATGGTGGCCGCGGGTGGTGGTTG
GCGCGGCTGCGCTGCGGCCCGGGGCAGTGCGGAGCCGGGACAGTCGCGGCGCTGACGCCCGCGGGCCC
CAGCTGCAGATATGAAGCGGAGCCGCTGCCGCGACCGACCGCAGCCGCCGCCGCCCGACCGCCGGGAG
GATGGAGTTCAGCGGGCAGCGGAGCTGTCTCAGTCTTTGCCGCCGCGCCGGCGAGCGCCGCCCGGGAG
GCAGCGGCTGGAGGAGCGGACGGGCCCCGCGGGGCCCGAGGGCAAGGAGCAGGATGTAGTAACTGGAG
TTAGTCCCCTGCTCTTCAGGAAACTCAGTAATCCTGACATATTTTCATCCACTGGAAAAGTTAAACTT
CAGCGACAACTGAGTCAGGATGATTGTAAGTTATGGAGAGGAAACCTGGCCAGCTCTCTATCGGGTAA
GCAGCTGCTCCCTTTGTCCAGCAGTGTACATAGCAGTGTGGGACAGGTGACTTGGCAGTCGTCAGGAG
AAGCATCAAACCTGGTTCGAATGAGAAACCAGTCCCTTGGACAGTCTGCACCTTCTCTTACTGCTGGC
CTGAAGGAGTTGAGCCTTCCAAGAAGAGGCAGCTTTTGTCGGACAAGTAACCGCAAGAGCTTGATTGT
GACCTCTAGCACATCACCTACACTACCACGGCCACACTCACCACTCCATGGCCACACAGGTAACAGTC
CTTTGGACAGCCCCCGGAATTTCTCTCCAAATGCACCTGCTCACTTTTCTTTTGTTCCTGCCCGTAGG
ACTGATGGGCGGCGCTGGTCTTTGGCCTCTTTGCCCTCTTCAGGATATGGAACTAACACTCCTAGCTC
CACTGTCTCATCATCATGCTCCTCACAGGAAAAGCTGCATCAGTTGCCTTTCCAGCCTACAGCTGATG
AGCTGCACTTTTTGACGAAGCATTTCAGCACAGAGAGCGTACCAGATGAGGAAGGACGGCAGTCCCCA
GCCATGCGGCCTCGCTCCCGGAGCCTCAGTCCCGGACGATCCCCAGTATCCTTTGACAGTGAAATAAT
AATGATGAATCACGTTTACAAAGAAAGATTCCCAAAGGCCACCGCACAAATGGAAGAGCGACTAGCAG
AGTTTATTTCCTCCAACACTCCAGACAGCGTGCTGCCCTTGGCAGATGGAGCCCTGAGCTTTATTCAT
CATCAGGTGATTGAGATGGCCCGAGACTGCCTGGATAAATCTCGGAGTGGCCTCATTACATCACAATA
CTTCTACGAACTTCAAGAGAATTTGGAGAAACTTTTACAAGATGCTCATGAGCGCTCAGAGAGCTCAG
AAGTGGCTTTTGTGATGCAGCTGGTGAAAAAGCTGATGATTATCATTGCCCGCCCAGCACGTCTCCTG
GAATGCCTGGAGTTTGACCCTGAAGAGTTCTACCACCTTTTAGAAGCAGCTGAGGGCCACGCCAAAGA
GGGACAAGGGATTAAATGTGACATTCCCCGCTACATCGTTAGCCAGCTGGGCCTCACCCGGGATCCCC
TAGAAGAAATGGCCCAGTTGAGCAGCTGTGACAGTCCTGACACTCCAGAGACAGATGATTCTATTGAG
GGCCATGGGGCATCTCTGCCATCTAAAAAGACACCCTCTGAAGAGGACTTCGAGACCATTAAGCTCAT
CAGCAATGGCGCCTATGGGGCTGTATTTCTGGTGCGGCACAAGTCCACCCGGCAGCGCTTTGCCATGA
AGAAGATCAACAAGCAGAACCTGATCCTACGGAACCAGATCCAGCAGGCCTTCGTGGAGCGTGACATA
CTGACTTTCGCTGAGAACCCCTTTGTGGTCAGCATGTTCTGCTCCTTTGATACCAAGCGCCACTTGTG
CATGGTGATGGAGTACGTTGAAGGGGGAGACTGTGCCACTCTGCTGAAGAATATTGGGGCCCTGCCTG
TGGACATGGTGCGTCTATACTTTGCGGAAACTGTGCTGGCCCTGGAGTACTTACACAACTATGGCATC
GTGCACCGTGACCTCAAGCCTGACAACCTCCTAATTACATCCATGGGGCACATCAAGCTCACGGACTT
TGGACTGTCCAAAATGGGCCTCATGAGTCTGACAACGAACTTGTATGAGGGTCATATTGAAAAGGATG
CCCGGGAATTCCTGGACAAGCAGGTATGCGGGACCCCAGAATACATTGCGCCTGAGGTGATCCTGCGC
CAGGGCTATGGGAAGCCAGTGGACTGGTGGGCCATGGGCATTATCCTGTATGAGTTCCTGGTGGGCTG
CGTCCCTTTTTTTGGAGATACTCCGGAGGAGCTCTTTGGGCAGGTGATCAGTGATGAGATTGTGTGGC
CTGAGGGTGATGAGGCACTGCCCCCAGACGCCCAGGACCTCACCTCCAAACTGCTCCACCAGAACCCT
CTGGAGAGACTTGGCACAGGCAGTGCCTATGAGGTGAAGCAGCACCCATTCTTTACTGGTCTGGACTG
GACAGGACTTCTCCGCCAGAAGGCTGAATTTATTCCTCAGTTGGAGTCAGAGGATGATACTAGCTATT
TTGACACCCGCTCAGAGCGATACCACCACATGGACTCGGAGGATGAGGAAGAAGTGAGTGAGGATGGC
TGCCTTGAGATCCGCCAGTTCTCTTCCTGCTCTCCAAGGTTCAACAAGGTGTACAGCAGCATGGAGCG
GCTCTCACTGCTCGAGGAGCGCCGGACACCACCCCCGACCAAGCGCAGCCTGAGTGAGGAGAAGGAGG
ACCATTCAGATGGCCTGGCAGGGCTCAAAGGCCGAGACCGGAGCTGGGTGATTGGCTCCCCTGAGATA
TTACGGAAGCGGCTGTCGGTGTCTGAGTCGTCCCACACAGAGAGTGACTCAAGCCCTCCAATGACAGT
GCGACGCCGCTGCTCAGGCCTCCTGGATGCGCCTCGGTTCCCGGAGGGCCCTGAGGAGGCCAGCAGCA
CCCTCAGGAGGCAACCACAGGAGGGTATATGGGTCCTGACACCCCCATCTGGAGAGGGGGTATCTGGG
CCTGTCACTGAACACTCAGGGGAGCAGCGGCCAAAGCTGGATGAGGAAGCTGTTGGCCGGAGCAGTGG
TTCCAGTCCAGCTATGGAGACCCGAGGCCGTGGGACCTCACAGCTGGCTGAGGGAGCCACAGCCAAGG
CCATCAGTGACCTGGCTGTGCGTAGGGCCCGCCACCGGCTGCTCTCTGGGGACTCAACAGAGAAGCGC
ACTGCTCGCCCTGTCAACAAAGTGATCAAGTCCGCCTCAGCCACAGCCCTCTCACTCCTCATTCCTTC
GGAACACCACACCTGCTCCCCGTTGGCCAGCCCCATGTCCCCACATTCTCAGTCGTCCAACCCATCAT
CCCGGGACTCTTCTCCAAGCAGGGACTTCTTGCCAGCCCTTGGCAGCATGAGGCCTCCCATCATCATC
CACCGAGCTGGCAAGAAGTATGGCTTCACCCTGCGGGCCATTCGCGTCTACATGGGTGACTCCGATGT
CTACACCGTGCACCATATGGTGTGGCACGTGGAGGATGGAGGTCCGGCCAGTGAGGCAGGGCTTCGTC
AAGGTGACCTCATCACCCATGTCAATGGGGAACCTGTGCATGGCCTGGTGCACACGGAGGTGGTGGAG
CTGATCCTGAAGAGTGGAAACAAGGTGGCCATTTCAACAACTCCCCTGGAGAACACATCCATTAAAGT
GGGGCCAGCTCGGAAGGGCAGCTACAAGGCCAAGATGGCCCGAAGGAGCAAGAGGAGCCGCGGCAAGG
ATGGGCAAGAAAGCAGAAAAAGGAGCTCCCTGTTCCGCAAGATCACCAAGCAAGCATCCCTGCTCCAC
ACCAGCCGCAGCCTTTCTTCCCTTAACCGCTCCTTGTCATCAGGGGAGAGTGGGCCAGGCTCTCCCAC
ACACAGCCACAGCCTTTCCCCCCGATCTCCCACTCAAGGCTACCGGGTGACCCCCGATGCTGTGCATT
CAGTGGGAGGGAATTCATCACAGAGCAGCTCCCCCAGCTCCAGCGTGCCCAGTTCCCCAGCCGGCTCT
GGGCACACACGGCCCAGCTCCCTCCACGGTCTGGCACCCAAGCTCCAACGCCAGTACCGCTCTCCACG
GCGCAAGTCAGCAGGCAGCATCCCACTGTCACCACTGGCCCACACCCCTTCTCCCCCACCCCCAACAG
CTTCACCTCAGCGGTCCCCATCGCCCCTGTCTGGCCATGTAGCCCAGGCCTTTCCCACAAAGCTTCAC
TTGTCACCTCCCCTGGGCAGGCAACTCTCACGGCCCAAGAGTGCGGAGCCACCCCGTTCACCACTACT
CAAGAGGGTGCAGTCGGCTGAGAAACTGGCAGCAGCACTTGCCGCCTCTGAGAAGAAGCTAGCCACTT
CTCGCAAGCACAGCCTTGACCTGCCCCACTCTGAACTAAAGAAGGAACTGCCGCCCAGGGAAGTGAGC
CCTCTGGAGGTAGTTGGAGCCAGGAGTGTGCTGTCTGGCAAGGGGGCCCTGCCAGGGAAGGGGGTGCT
GCAGCCTGCTCCCTCACGGGCCCTAGGCACCCTCCGGCAGGACCGAGCCGAACGACGGGAGTCGCTGC
AGAAGCAAGAAGCCATTCGTGAGGTGGACTCCTCAGAGGACGACACCGAGGAAGGGCCTGAGAACAGC

-continued

CAGGGTGCACAGGAGCTGAGCTTGGCACCTCACCCAGAAGTGAGCCAGAGTGTGGCCCCTAAAGGAGC
AGGAGAGAGTGGGGAAGAGGATCCTTTCCCGTCCAGAGGCCCTAGGAGCCTGGGCCCAATGGTCCCAA
GCCTATTGACAGGGATCACACTGGGGCCTCCCAGAATGGAAAGTCCCAGTGGTCCCCACAGGAGGCTC
GGGAGCCCACAAGCCATTGAGGAGGCTGCCAGCTCCTCCTCAGCAGGCCCCAACCTAGGTCAGTCTGG
AGCCACAGACCCCATCCCTCCTGAAGGTTGCTGGAAGGCCCAGCACCTCCACACCCAGGCACTAACAG
CACTTTCTCCCAGCACTTCGGGACTCACCCCCACCAGCAGTTGCTCTCCTCCCAGCTCCACCTCTGGG
AAGCTGAGCATGTGGTCCTGGAAATCCCTTATTGAGGGCCCAGACAGGGCATCCCCAAGCAGAAAGGC
AACCATGGCAGGTGGGCTAGCCAACCTCCAGGATTTGGAAACACAACTCCAGCCCAGCCTAAGAACCT
GTCTCCCAGGGAGCAGGGGAAGACACAGCCACCTAGTGCCCCCAGACTGGCCCATCCATCTTATGAGG
ATCCCAGCCAGGGCTGGCTATGGGAGTCTGAGTGTGCACAAGCAGTGAAAGAGGATCCAGCCCTGAGC
ATCACCCAAGTGCCTGATGCCTCAGGTGACAGAAGGCAGGACGTTCCATGCCGAGGCTGCCCCCTCAC
CCAGAAGTCTGAGCCCAGCCTCAGGAGGGGCCAAGAACCAGGGGGCCATCAAAAGCATCGGGATTTGG
CATTGGTTCCAGATGAGCTTTTAAAGCAAACATAGCAGTTGTTTGCCATTTCTTGCACTCAGACCTGT
GTAATATATGCTCCTGGAAACCATC

SEQ ID NO:30 Human MAST205 polypeptide sequence
protein_id:gi13537204
MKRSRCRDRPQPPPPDRREDGVQRAAELSQSLPPRRRAPPGRQRLEERTGPAGPEGKEQDVVTGVSPL
LFRKLSNPDIFSSTGKVKLQRQLSQDDCKLWRGNLASSLSGKQLLPLSSSVHSSVGQVTWQSSGEASN
LVRMRNQSLGQSAPSLTAGLKELSLPRRGSFCRTSNRKSLIVTSSTSPTLPRPHSPLHGHTGNSPLDS
PRNFSPNAPAHFSFVPARRTDGRRWSLASLPSSGYGTNTPSSTVSSSCSSQEKLHQLPFQPTADELHF
LTKHFSTESVPDEEGRQSPAMRPRSRSLSPGRSPVSFDSEIIMMNHVYKERFPKATAQMEERLAEFIS
SNTPDSVLPLADGALSFIHHQVIEMARDCLDKSRSGLITSQYFYELQENLEKLLQDAHERSESSEVAF
VMQLVKKLMIIIARPARLLECLEFDPEEFYHLLEAAEGHAKEGQGIKCDIPRYIVSQLGLTRDPLEEM
AQLSSCDSPDTPETDDSIEGHGASLPSKKTPSEEDFETIKLISNGAYGAVFLVRHKSTRQRFAMKKIN
KQNLILRNQIQQAFVERDILTFAENPFVVSMFCSFDTKRHLCMVMEYVEGGDCATLLKNIGALPVDMV
RLYFAETVLALEYLHNYGIVHRDLKPDNLLITSMGHIKLTDFGLSKMGLMSLTTNLYEGHIEKDAREF
LDKQVCGTPEYIAPEVILRQGYGKPVDWWAMGIILYEFLVGCVPFFGDTPEELFGQVISDEIVWPEGD
EALPPDAQDLTSKLLHQNPLERLGTGSAYEVKQHPFFTGLDWTGLLRQKAEFIPQLESEDDTSYFDTR
SERYHHMDSEDEEEVSEDGCLEIRQFSSCSPRFNKVYSSMERLSLLEERRTPPPTKRSLSEEKEDHSD
GLAGLKGRDRSWVIGSPEILRKRLSVSESSHTESDSSPPMTVRRRCSGLLDAPRFPEGPEEASSTLRR
QPQEGIWVLTPPSGEGVSGPVTEHSGEQRPKLDEEAVGRSSGSSPAMETRGRGTSQLAEGATAKAISD
LAVRRARHRLLSGDSTEKRTARPVNKVIKSASATALSLLIPSEHHTCSPLASPMSPHSQSSNPSSRDS
SPSRDFLPALGSMRPPIIIHRAGKKYGFTLRAIRVYMGDSDVYTVHHMVWHVEDGGPASEAGLRQGDL
ITHVNGEPVHGLVHTEVVELILKSGNKVAISTTPLENTSIKVGPARKGSYKAKMARRSKRSRGKDGQE
SRKRSSLFRKITKQASLLHTSRSLSSLNRSLSSGESGPGSPTHSHSLSPRSPTQGYRVTPDAVHSVGG
NSSQSSSPSSSVPSSPAGSGHTRPSSLHGLAPKLQRQYRSPRRKSAGSIPLSPLAHTPSPPPPTASPQ
RSPSPLSGHVAQAFPTKLHLSPPLGRQLSRPKSAEPPRSPLLKRVQSAEKLAAALAASEKKLATSRKH
SLDLPHSELKKELPPREVSPLEVVGARSVLSGKGALPGKGVLQPAPSRALGTLRQDRAERRESLQKQE
AIREVDSSEDDTEEGPENSQGAQELSLAPHPEVSQSVAPKGAGESGEEDPFPSRGPRSLGPMVPSLLT
GITLGPPRMESPSGPHRRLGSPQAIEEAASSSSAGPNLGQSGATDPIPPEGCWKAQHLHTQALTALSP
STSGLTPTSSCSPPSSTSGKLSMWSWKSLIEGPDRASPSRKATMAGGLANLQDLETQLQPSLRTCLPG
SRGRHSHLVPPDWPIHLMRIPARAGYGSLSVHKQ

SEQ ID NO:31 mouse MAST205 nucleic acid sequence
accession:NM_008641
GGACGGCCGAGTTGCCCCAGCCGCAGTCTTTGCCCCCGCGCCGGCGAGCGCCGCCCGGGAGGCAGCTA
CTGGAGGAGCGGAGCGGGCCTTTGGGGCACGACAGCAGGGAGCAGGATATGGTTACTGGACTTAGTCC
TCTGCTCTTCAGGAAGCTTAGTAATCCTGACATATTTGCACCCACTGGAAAAGTTAAACTCCAGCGAC
AACTTAGTCAGGATGACTGTAAGTTACGGAGAGGAAGCCTGGCAGTTCTCTGTCGGGTCGGGCAGCTG
CTCCCTTTGTCCAGCAGTGTACACAGCAGTGTGGGACAGGTAACTTGGCAGTCTACAGGAGAAGCATC
AAACCTGGTTCGAATGAGAAACCAATCCCTTGGACAGTCTGCACCTTCCCTTACGGCTGGCTTGAAGG
AATTGAGCCTTCCAAGGAGAGGCAGCTTTTGTCGGACAAGTAACCGCAAGAGCTTGATTGTAACCTCC
AGCACATCACCTACGCTACCAAGGCCACACTCCCCACTCCATGGCCACACAGGTAACAGTCCCTTGGA
CAGCCCCCGGAATTTCTCTCCAAATGCTCCTGCTCACTTTTCCTTCGTTCCTGCCCGTAGGACTGATG
GACGGCGCTGGTCTTTGGCCTCGTTGCCTTCTTCAGGCTATGGAACCAACACTCCTAGCTCTACAGTC
TCTTCATCATGTTCCTCACAAGAAAAGCTTCATCAGTTACCTTTCCAGCCAACAGCTGATGAACTACA
CTTTCTGACGAAGCATTTTAGCACAGAAAACGTACCAGATGAGGAGGGACGTCGGTCCCCACGGATGC
GGCCCCGTTCCCGCAGCCTCAGTCCTGGACGGTCCCCAGTTTCCTTTGACAGTGAAATAATAATGATG
AATCATGTGTACAAAGAAAGATTCCCCAAGGCCACTGCACAGATGGAAGAGCGACCTAGCCTGACCTT
CATTTCCTCTAACACTCCAGATAGTGTGTTGCCCTTGGCAGATGGAGCACTAAGCTTTATTCATCATC
AGGTGATTGAGATGGCCCGAGACTGCCTGGATAAATCTCGGAGTGGCCTCATTACGTCACACTATTTC
TATGAACTTCAAGAGAATTTGGAAAAGCTTCTGCAAGATGCTCACGAACGCTCAGAAAGCTCAGATGT
AGCCTTTGTGATACAGCTGGTAAAAAAGTTGATGATCATCATTGCTCGCCCAGCTCGCCTCCTGGGAT
GCCTGGAGTTTGACCCTGAAGAATTTTACCACCTGTTAGAAGCAGCTGAAGGTCATGCCAAAGAGGGA
CATGGAATTAAATGTGACATTCCCCGCTACATCGTTAGCCAGCTGGGCCTAACTCGGGATCCCTTGGA
GGAAATGGCCCAGTTGAGCAGCTATGACAGTCCAGATACTCCAGAGACAGATGATTCAGTTGAGGGTC
GTGGGGTATCTCAGCCATCTCAGAAGACCCCCTCTGAAGAGGACTTTGAAACCATTAAGCTCATCAGC
AATGGCGCCTATGGGGCTGTCTTTCTGGTGCGGCACAAGTCCACGCGGCAGCGCTTTGCAATGAAGAA
GATTAATAAGCAAAACCTAATCCTACGGAACCAGATCCAGCAAGCATTTGTGGAACGCGACATACTGA
CTTTCGCTGAAAACCCCTTTGTGGTCAGCATGTTCTGCTCCTTTGAGACCAAGCGTCACTTATGCATG
GTGATGGAATACGTAGAAGGGGGAGACTGTGCCACTCTGCTCAAGAACATCGGGGCCCTACCTGTGGA
CATGGTACGCCTGTACTTTGCGGAAACTGTGCTGGCTTTGGAATACTTACACAACTACGGCATCGTGC
ACCGTGACCTCAAGCCTGACAACCTTCTGATTACATCCATGGGACACATCAAACTCACTGACTTTGGA
CTTTCTAAAATCGGCCTCATGAGTTTGACAACCAACTTGTATGAGGGTCATATTGAAAAGGATGCCCG
GGAGTTCCTAGACAAGCAGGTATGCGGGACCCCGGAATACATAGCACCTGAGGTGATCCTGCGTCAGG
GATATGGGAAGCCAGTGGACTGGTGGGCCATGGGCATAATCCTGTATGAGTTCCTGGTGGGTTGCGTC

-continued

CCTTTCTTTGGAGACACTCCGGAGGAGCTCTTTGGGCAAGTGATCAGTGATGAGATCGTGTGGCCCGA
AGGTGATGACGCGCTTCCCCCAGATGCCCAAGACCTCACTTCCAAACTGCTTCATCAGAATCCACTAG
AAAGACTGGGCACAAGTAGTGCCTATGAAGTGAAGCAGCACCCATTCTTCATGGGTCTGGACTGGACA
GGACTTCTACGGCAGAAGGCTGAATTTATACCTCAGCTGGAGTCAGAGGATGACACCAGCTACTTTGA
TACCCGTTCAGAACGATACCACCATGTGGACTCTGAGGATGAGGAGGAAGTGAGTGAGGATGGCTGCC
TTGAGATTCGCCAGTTTTCTTCCTGCTCTCCAAGGTTCAGCAAGGTTTACAGTAGCATGGAACGGCTT
TCCCTGCTTGAGGAACGCCGGACACCACCTCCAACCAAGCGCAGCCTCAGTGAGGAAAAGGAAGATCA
CTCAGACGGCTTGGCAGGACTGAAGGGCCGAGATCGCAGCTGGGTGATTGGGTCCCCTGAGATATTAC
GGAAGCGGTTATCTGTGTCCGAGTCATCCCACACAGAGAGTGATTCGAGTCCTCCAATGACAGTGCGA
CATCGGTGTTCAGGCCTCCCAGATGGGCCTCATTGCCCTGAGGAGACTAGTAGCACCCCTCGGAAGCA
ACAACAGGAAGGTATATGGGTCCTCATTCCCCCATCTGGAGAGGGGTCATCTAGGCCTGTTCCTGAAC
GACCGTTGGAGAGGCAACTGAAGCTGGATGAGGAGCCTCCTGGCCAAAGCGGGTCGGTGTTGCCAGCC
CTGGAGACACGAGGCCGTGGGACCCCTCAGCTAGCTGAGGAAGCTACAGCCAAAGCCATCAGCGACCT
AGCTGTGCGTAGGGCCCGTCACCGGCTGCTCTCTGGGGACTCTATAGAGAAGCGCACCACTCGCCCTG
TCAACAAAGTAATCAAGTCAGCCTCAGCTACGGCCCTTTCCCTCCTCATTCCTTCAGAACACCATGCC
TGCTCACCATTGGCCAGCCCTATGTCCCCACATTCCCAGTCATCCAATCCATCATCCAGGGACTCTTC
TCCAAGCAGGGACTTCTTGCCAGCCCTTGGCAGCTTGAGGCCTCCCATCATCATCCACCGAGCTGGCA
AGAAGTATGGCTTCACCCTGCGGGCCATTCGAGTCTACATGGGTGACACTGATGTCTACACCGTACAC
CACATGGTGTGGCATGTGGAGGATGGTGGTCCAGCCAGTGAAGCAGGGCTTCGTCAGGGTGACCTCAT
CACCCATGTCAATGGTGAGCCTGTGCATGGGCTAGTCCACACAGAAGTGGTGGAGCTGGTTCTGAAGA
GTGGAAACAAGGTATCAATTTCAACAACTCCCTTGGAGAACACGTCAATCAAAGTGGGGCCAGCTAGG
AAAGGCAGCTATAAAGCCAAGATGGCCCGAAGGAGCAAACGGAGCAAAGGCAAGGATGGGCAAGAAAG
CCGAAAAAGAAGCTCCCTATTCCGGAAAATCACAAAGCAGGCCTCCTTGCTCCACACCAGCCGCAGCC
TTTCTTCCCTTAACCGCTCCTTGTCATCAGGGGAGAGTGGTCCAGGCTCTCCCACACACAGCCACAGC
CTCTCTCCCAGATCTCCTCCTCAGGGATACCGGGTAGCCCCAGATGCTGTGCACTCAGTAGGAGGGAA
TTCCTCGCAGAGCAGCTCTCCCAGCTCCAGTGTGCCCAGTTCTCCTGCTGGCTCTGGACATACACGGC
CCAGCTCTCTTCACGGTCTGGCACCCAAGCTCCAACGCCAATACCGCTCACCACGGCGCAAGTCAGCA
GGCAGCATCCCACTGTCACCGTTGGCCCACACCCCTTCCCCACCAGCAACGGCAGCTTCACCTCAGCG
TTCCCCATCACCCTTGTCTGGCCATGGGTCTCAGTCCTTTCCTACCAAACTTCACTTGTCTCCTCCGC
TAGGTAGGCAGCTCTCACGGCCCAAGAGTGCAGAGCCACCCCGCTCTCCCCTACTTAAGAGGGTGCAG
TCTGCTGAGAAGCTGGCGGCTGCACTGGCAGCTGCTGAGAAGAAGTTAGCACCTTCCCGCAAACATAG
TCTTGACCTGCCCCATGGTGAACTAAAGAAGGAACTGACACCCAGGGAAGCCAGCCCTCTGGAGGTAG
TTGGAACCAGAAGTGTGCTATCCGGGAAAGGGCCACTTCCAGGAAAGGGGGTACTGCAGCCTGCTCCT
TCACGGGCCCTTGGGACCCTACGGCAGGATCGAGCTGAACGCCGTGAGTCACTGCAAAAACAAGAAGC
AATCCGGGAAGTAGACTCCTCAGAAGATGACACTGATGAGGAGCCTGAGAACAGCCAGGCCACACAGG
AGCCAAGATTGTCCCCCCACCCAGAAGCAAGCCACAATCTACTCCCTAAAGGTTCAGGAGAGGGTACA
GAAGAGGACACTTTCTTGCACAGGGATCTAAAGAAGCAGGGCCCTGTACTCTCAGGTCTAGTGACAGG
GGCCACACTAGGCTCCCCCCGAGTAGACGTTCCTGGGCTCTCCCCAAGGAAGGTCAGCAGGCCACAAG
CCTTTGAGGAAGCTACCAACCCCTTACAAGTCCCTAGCCTGAGCAGGTCTGGACCCACAAGCCCCACC
CCCTCTGAAGGCTGCTGGAAGGCCCAGCACCTCCACACACAGGCACTAACTGCACTTTGTCCCAGCTT
TTCAGAACTTACCCCTACCGGTTGTTCTGCTGCCACCTCCACCTCTGGAAAGCCAGGGACATGGTCCT
GGAAATTCCTTATTGAGGGTCCAGACAGAGCATCCACGAACAAGACCATAACAAGGAAAGGTGAACCA
GCTAACTCCCAAGATACGAATACCACGGTCCCAAATCTTCTGAAGAACCTGTCTCCTGAGGAGGAGAA
GCCACAGCCACCAAGTGTGCCTGGGCTGACCCATCCGCTTCTTGAGGTCCCCAGCCAGAACTGGCCAT
GGGAGTCTGAATGTGAACAAATGGAGAAAGAAGAACCATCCCTGAGCATCACCGAAGTGCCTGATTCC
TCAGGCGACAGGAGGCAGGACATTCCATGCAGAGCCCACCCCCTGAGCCCAGAAACCCGGCCCAGCCT
GCTCTGGAAAAGCCAAGAACTTGGGGGCCAGCAAGATCATCAGGACTTAGCACTGACATCAGATGAGC
TCTTAAAGCAAACCTAGCACTTGTTTGCTTCCCTTACATTCACCTGTGTAATACACCCTCCTGGAAAC
CA

SEQ ID NO:32 mouse MAST205 polypeptide sequence
accession:gi6678958
MVTGLSPLLFRKLSNPDIFAPTGKVKLQRQLSQDDCKLRRGSLASSLSGKQLLPLSSSVHSSVGQVTW
QSTGEASNLVRMRNQSLGQSAPSLTAGLKELSLPRRGSFCRTSNRKSLIVTSSTSPTLPRPHSPLHGH
TGNSPLDSPRNFSPNAPAHFSFVPARRTDGRRWSLASLPSSGYGTNTPSSTVSSSCSSQEKLHQLPFQ
PTADELHFLTKHFSTENVPDEEGRRSPRMRPRSRSLSPGRSPVSFDSEIIMMNHVYKERFPKATAQME
ERPSLTFISSNTPDSVLPLADGALSFIHHQVIEMARDCLDKSRSGLITSHYFYELQENLEKLLQDAHE
RSESSDVAFVIQLVKKLMIIIARPARLLECLEFDPEEFYHLLEAAEGHAKEGHGIKCDIPRYIVSQLG
LTRDPLEEMAQLSSYDSPDTPETDDSVEGRGVSQPSQKTPSEEDFETIKLISNGAYGAVFLVRHKSTR
QRFAMKKINKQNLILRNQIQQAFVERDILTFAENPFVVSMFCSFETKRHLCMVMEYVEGGDCATLLKN
IGALPVDMVRLYFAETVLALEYLHNYGIVHRDLKPDNLLITSMGHIKLTDFGLSKIGLMSLTTNLYEG
HIEKDAREFLDKQVCGTPEYIAPEVILRQGYGKPVDWWAMGIILYEFLVGCVPFFGDTPEELFGQVIS
DEIVWPEGDDALPPDAQDLTSKLLHQNPLERLGTSSAYEVKQHPFFMGLDWTGLLRQKAEFIPQLESE
DDTSYFDTRSERYHHVDSEDEEEVSEDGCLEIRQFSSCSPRFSKVYSSMERLSLLEERRTPPPTKRSL
SEEKEDHSDGLAGLKGRDRSWVIGSPEILRKRLSVSESSHTESDSSPPMTVRHRCSGLPDGPHCPEET
SSTPRKQQQEGIWVLIPPSGEGSSRPVPERPLERQLKLDEEPPGQSSRCCPALETRGRGTPQLAEEAT
AKAISDLAVRRARHRLLSGDSIEKRTTRPVNKVIKSASATALSLLIPSEHHACSPLASPMSPHSQSSN
PSSRDSSPSRDFLPALGSLRPPTIIHRAGKKYGFTLRAIRVYMGDTDVYTVHHMVWHVEDGGPASEAG
LRQGDLITHVNGEPVHGLVHTEVVELVLKSGNKVSISTTPLENTSIKVGPARKGSYKAKMARRSKRSK
GKDGQESRKRSSLFRKITKQASLLHTSRSLSSLNRSLSSGESGPGSPTHSHSLSPRSPPQGYRVAPDA
VHSVGGNSSQSSSPSSSVPSSPAGSGHTRPSSLHGLAPKLQRQYRSPRRKSAGSIPLSPLAHTPSPPA
TAASPQRSPSPLSGHGSQSFPTKLHLSPPLGRQLSRPKSAEPPRSPLLKRVQSAEKLAAALAAAEKKL
APSRKHSLDLPHGELKKELTPREASPLEVVGTRSVLSGKGPLPGKGVLQPAPSRALGTLRQDRAERRE
SLQKQEAIREVDSSEDDTDEEPENSQATQEPRLSPHPEASHNLLPKGSGEGTEEDTFLHRDLKKQGPV

-continued

LSGLVTGATLGSPRVDVPGLSPRKVSRPQAFEEATNPLQVPSLSRSGPTSPTPSEGCWKAQHLHTQAL
TALCPSFSELTPTGCSAATSTSGKPGTWSWKFLIEGPDRASTNKTITRKGEPANSQDTNTTVPNLLKN
LSPEEEKPQPPSVPGLTHPLLEVPSQNWPWESECEQMEKEEPSLSITEVPDSSGDRRQDIPCRAHPLS
PETRPSLLWKSQELGGQQDHQDLALTSDELLKQT

SEQ ID NO:33 Human CKLF nucleotide sequence
HUM224910 accession:AF132818 CDS:537 . . . 1910
GCGGCCGCCCTGCGCGCGAAGCTCGTGGCCCGAGAGGGGTGCGGTCGGGCCGACGGAGGCGGGGCCCT
GGCTGCCTCTCTCCCTGCTCATAGGCTGGCCGCTCAGGCCTGGCCGGCCTCGGGGCCTCGGGATTCGC
GGCGGCGCTGCCAATCAGGCGATCGGGCCCCGCCCCCCCGGAGTTGGGTGAAATAGAGGCGGGCGTCA
AGTGTCAGTAGTCGCGGGGCAGGTACGTGCGCTCGCGGTTCTCTCGCGGAGGTCGGCGGTGGCGGGAG
CGGGCTCCGGAGAGCCTGAGAGCACGGTGGGGCGGGGCGGGAGAAAGTGGCCGCCCGGAGGACGTTGG
CGTTTACGTGTGGAAGAGCGGAAGAGTTTTGCTTTTCGTGCGCGCCTTCGAAAACTGCCTGCCGCTGT
CTGAGGAGTCCACCCGAAACCTCCCCTCCTCCGCCGGCAGCCCCGCGCTGAGCTCGCCGACCCAAGCC
AGCGTGGGCGAGGTGGGAAGTGCGCCCGACCCGCGCCTGGAGCTGCGCCCCCGAGTGCCCATGGCTAC
AAGGGTGCTGAGCATGAGCGCCCGCCTGGGACCCGTGCCCCAGCCGCCGGCGCCGCAGGACGAGCCGG
TGTTCGCGCAGCTCAAGCCGGTGCTGGGCGCCGCGAATCCGGCCCGCGACGCGGCGCTCTTCCCCGGC
GAGGAGCTGAAGCACGCGCACCACCGCCCGCAGGCGCAGCCCGCGCCCGCGCAGGCCCCGCAGCCGGC
CCAGCCGCCCGCCACCGGCCCGCGGCTGCCTCCAGAGGACCTGGTCCAAACAAGATGTGAAATGGAGA
AGTATCTGACACCTCAGCTTCCTCCAGTTCCTATAATTCCAGAGCATAAAAAGTATAGACGAGACAGT
GCCTCAGTCGTAGACCAGTTCTTCACTGACACTGAAGGGTTACCTTACAGTATCAACATGAACGTCTT
CCTCCCTGACATCACTCACCTGAGAACTGGCCTCTACAAATCCCAGAGACCGTGCGTAACACACATCA
AGACAGAACCTGTTGCCATTTTCAGCCACCAGAGTGAAACGACTGCCCCTCCTCCGGCCCCGACCCAG
GCCCTCCCTGAGTTCACCAGTATATTCAGCTCACACCAGACCGCAGCTCCAGAGGTGAACAATATTTT
CATCAAACAAGAACTTCCTACACCAGATCTTCATCTTTCTGTCCCTACCCAGCAGGGCCACCTGTACC
AGCTACTGAATACACCGGATCTAGATATGCCCAGTTCTACAAATCAGACAGCAGCAATGGACACTCTT
AATGTTTCTATGTCAGCTGCCATGGCAGGCCTTAACACACACACCTCTGCTGTTCCGCAGACTGCAGT
GAAACAATTCCAGGGCATGCCCCCTTGCACATACACAATGCCAAGTCAGTTTCTTCCACAACAGGCCA
CTTACTTTCCCCCGTCACCACCAAGCTCAGAGCCTGGAAGTCCAGATAGACAAGCAGAGATGCTCCAG
AATTTAACCCCACCTCCATCCTATGCTGCTACAATTGCTTCTAAACTGGCAATTCACAATCCAAATTT
ACCCACCACCCTGCCAGTTAACTCACAAAACATCCAACCTGTCAGATACAATAGAAGGAGTAACCCCG
ATTTGGAGAAACGACGCATCCACTACTGCGATTACCCTGGTTGCACAAAAGTTTATACCAAGTCTTCT
CATTTAAAAGCTCACCTGAGGACTCACACTGGTGAAAAGCCATACAAGTGTACCTGGGAAGGCTGCGA
CTGGAGGTTCGCGCGATCGGATGAGCTGACCCGCCACTACCGGAAGCACACAGGCGCCAAGCCCTTCC
AGTGCGGGGTGTGCAACCGCAGCTTCTCGCGCTCTGACCACCTGGCCCTGCATATGAAGAGGCACCAG
AACTGAGCACTGCCCGTGTGACCCGTTCCAGGTCCCCTGGGCTCCCTCAAATGACAGACCTAACTATT
CCTGTGTAAAAACAACAAAAAACAAACAAAAGCAAGAAAACCACAACTAAAACTGGAAATGTATATTT
TGTATATTTGAGAAAACAGGGAATACATTGTATTAATACCAAAGTGTTTGGTCATTTTAAGAATCTGG
AATGCTTGCTGTAATGTATATGGCTTTACTCAAGCAGATCTCATCTCATGACAGGCAGCCACGTCTCA
ACATGGGTAAGGGGTGGGGGTGGAGGGGAGTGTGTGCAGCGTTTTTACCTAGGCACCATCATTTAATG
TGACAGTGTTCAGTAAACAAATCAGTTGGCAGGCACCAGAAGAAGAATGGATTGTATGTCAAGATTTT
ACTTGGCATTGAGTAGTTTTTTTCAATAGTAGGTAATTCCTTAGAGATACAGTATACCTGGCAATTCA
CAAATAGCCATTGAACAAATGTGTGGGTTTTTAAAAATTATATACTATATGAGTTGCCTATATTTGCT
ATTCAAAATTTTGTAAATATGCAAATCAGCTTTATAGGTTTATTACAAGTTTTTTAGGATTCTTTTGG
GGAAGAGTCATAATTCTTTTGAAAATAACCATGAATACACTTACAGTTAGGATTTGTGGTAAGGTACC
TCTCAACATTACCAAAATCATTTCTTTAGAGGGAAGGAATAATCATTCAAATGAACTTTAAAAAAGCA
AATTTCATGCACTGATTAAAATAGGATTATTTTAAATACAAAAGGCATTTTATATGAATTATAAACTG
AAGAGCTTAAAGATAGTTACAAATACAAAAGTTCAACCTCTTACAATAAGCTAAACGCAATGTCTTTT
TAAAAAGAGGACTTAGGGTGTCGTTTTTCACATATGACAATGTTGCATTTATGATGCAGTTTCAAGTA
CCAAAACGTTGAATTGATGATGCAGTTTTCATATATCGAGATGTTCGCTCGTGCAGTACTGTTGGTTA
AATGACAATTTATGTGGATTTTGCATGTAATACACAGTGAGACACAGTAATTTTATCTAAATTACAGT
GCAGTTTAGTTAATCTATTAATACTGACTCAGTGTCTGCCTTTAAATATAAATGATATGTTGAAAACT
TAAGGAAGCAAATGCTACATATATGCAATATAAAATAGTAATGTGATGCTGATGCTGTTAACCAAAGG
GCAGAATAAATAAGCAAAATGCCAAAAGGGGTCTTAATTGAAATGAAAATTTAATTTTGTTTTTAAAA
TATTGTTTATCTTTATTTATTTGGGGTAATATTGTAAGTTTTTTAGAAGACAATTTTCATAACTTGA
TAAATTATAGTTTTGTTTGTTAGAAAAGTAGCTCTTAAAAGATGTAAATAGATGACAAACGATGTAAA
TAATTTTGTAAGAGGCTTCAAAATGTTTATACGTGGAAACACACCTACATGAAAAGCAGAAATCGGTT
GCTGTTTTGCTTCTTTTTCCCTCTTATTTTTGTATTGTGGTCATTTCCTATGCAAATAATGGAGCAAA
CAGCTGTATAGTTGTAGAATTTTTTGAGAGAATGAGATGTTTATATATTAACGACAATTTTTTTTTGG
AAAATAAAAAGTGCCCTAAAAGAAAAAAAAAA

SEQ ID NO:34 Human CKLF protein sequence
protein_id:gi6580835
MATRVLSMSARLGPVPQPPAPQDEPVFAQLKPVLGAANPARDAALFPGEELKHAHHRPQAQPAPAQAP
QPAQPPATGPRLPPEDLVQTRCEMEKYLTPQLPPVPIIPEHKKYRRDSASVVDQFFTDTEGLPYSINM
NVFLPDITHLRTGLYKSQRPCVTHIKTEPVAIFSHQSETTAPPPAPTQALPEFTSIFSSHQTAAPEVN
NIFIKQELPTPDLHLSVPTQQGHLYQLLNTPDLDMPSSTNQTAAMDTLNVSMSAAMAGLNTHTSAVPQ
TAVKQFQGMPPCTYTMPSQFLPQQATYFPPSPPSSEPGSPDRQAEMLQNLTPPPSYAATIASKLAIHN
PNLPTTLPVNSQNIQPVRYNRRSMPDLEKRRIHYCDYPGCTKVYTKSSHLKAHLRTHTGEKPYKCTWE
GCDWRFARSDELTRHYRKHTGAKPFQCGVCNRSFSRSDHLALHMKRHQN

SEQ ID NO:35 Mouse CKLF nucleotide sequence
accession:AF079852 CDS:167 . . . 1507
CCGAGCCCAGGAGCCCCGATCTCCGTGCCCGCCTTCGTGAGCGTCTGGCTGCCGGCCCAGGGGTCCCC
CGCCGCGGCCCCCCGCCGAGTCCGCCGTCCCGTGCCAGCCCGAGCGAGGTGGGATCGCGATCGCTCCG
TGTCCCGCTCCCGTAATCCCCAGACCGTCCATGCCCACGCGGGTGCTGACCATGAGCGCCCGCCTGGG
ACCACTGCCCCAGCCGCCGGCCGCGCAGGCCGAGCCCGTGTTCGCGCAGCTCAAGCCGGTGCTGGGCG

-continued

CTGCGAACCCGGCCCGCGACGCGGCGCTCTTCTCCGGAGACGATCTGAAACACGCGCACCACCACCCG
CCTGCGCCGCCGCCAGCCGCTGGCCCGCGACTGCCCTCGGAGGAGCTGGTCCAGACAAGATGTGAAAT
GGAGAAGTATCTGACCCCTCAGCTCCCTCCAGTTCCGATAATTTCAGAGCATAAAAAGTATAGACGAG
ACAGTGCCTCAGTGGTAGACCAGTTCTTCACTGACACTGAAGGCATACCTTACAGCATCAACATGAAC
GTCTTCCTCCCTGACATCACTCACCTGAGAACTGGCCTCTACAAATCCCAGAGACCATGCGTAACACA
GATCAAGACAGAACCTGTTACCATTTTCAGCCACCAGAGCGAGTCGACGGCCCCTCCTCCTCCTCCGG
CCCCCACCCAGGCTCTCCCCGAGTTCACTAGTATCTTCAGCTCCCACCAGACCACAGCGCCACCACAG
GAGGTGAACAATATCTTCATCAAACAAGAACTTCCTATACCAGATCTTCATCTCTCTGTCCCTTCCCA
GCAGGGCCACCTGTACCAGCTGTTGAATACACCGGATCTAGACATGCCCAGTTCGACAAACCAGACGG
CAGTAATGGACACCCTTAATGTTTCTATGGCAGGCCTTAACCCACACCCCTCTGCTGTTCCACAGACG
TCAATGAAACAGTTCCAGGGCATGCCCCCTTGCACGTACACCATGCCAAGTCAGTTTCTTCCACAGCA
GGCCACTTATTTCCCCCGTCACCACCAAGCTCAGAGCCTGGAAGTCCCGATAGACAAGCTGAGATGC
TGCAGAATCTCACCCCACCTCCGTCCTATGCCGCTACAATTGCTTCCAAACTGGCGATTCACAACCCA
AATTTACCTGCCACTCTGCCAGTTAATTCGCCAACTCTCCCACCTGTCAGATACAACAGAAGGAGTAA
CCCGGATCTGGAGAAGCGACGTATCCACTTCTGCGATTATAATGGTTGCACAAAAGTTTATACAAAGT
CGTCTCACTTAAAAGCTCACCTGAGGACTCATACGGGCGAGAAGCCCTACAAGTGCACCTGGGAGGGC
TGCGACTGGAGGTTTGCCCGGTCGGATGAGCTGACCCGCCACTACAGGAAGCACACGGGCGCCAAGCC
GTTCCAGTGCATGGTGTGCCAACGCAGCTTCTCCCGCTCCGACCACCTCGCGCTGCACATGAAGCGCC
ACCAGAACTGAGCGAGCGAACGCTGCGCCCACCCGCCTGACGCCTTGCAGTCCGCTTTGCCATCCTTT
AAACCGCAGACCTAACTTCATAAAAAG

SEQ ID NO:36 Mouse CKLF protein sequence
accession:gi4336209
MPTRVLTMSARLGPLPQPPAAQAEPVFAQLKPVLGAANPARDAALFSGDDLKHAHHHPPAPPPAAGPR
LPSEELVQTRCEMEKYLTPQLPPVPIISEHKKYRRDSASVVDQFFTDTEGIPYSINMNVFLPDITHLR
TGLYKSQRPCVTQIKTEPVTIFSHQSESTAPPPPPAPTQALPEFTSIFSSHQTTAPPQEVNNIFIKQE
LPIPDLHLSVPSQQGHLYQLLNTPDLDMPSSTNQTAVMDTLNVSMAGLNPHPSAVPQTSMKQFQGMPP
CTYTMPSQFLPQQATYFPPSPPSSEPGSPDRQAEMLQNLTPPPSYAATIASKLAIHNPNLPATLPVNS
PTLPPVRYNRRSNPDLEKRRIHFCDYNGCTKVYTKSSHLKAHLRTHTGEKPYKCTWEGCDWRFARSDE
LTRHYRKHTGAKPFQCMVCQRSFSRSDHLALHMKRHQN

SEQ ID NO:37 Rat CKLF nucleotide sequence
accession:NM_053394
CGGTATTTCAGCTCCCACCAGACCACAGCGCCAGAGGTGAACAATATCTTCATCAAACAAGAACTTCC
TATACCAGATCTTCATCTCTCGGTCCCTTCCCAGCAGGGCCACCTGTACCAGCTGTTGAATACACCTG
ATCTAGACATGCCCAGTTCGACAAACCAGACAGCAGTCATGGACACCCTTAATGTCTCTATGGCTGGC
CTTAACTCACACCCCTCTGCTGTGCCACAGACGTCCATGAAACAGTTCCAGGGCATGCCTCCTTGCAC
GTACACCATGCCGAGTCAGTTTCTTCCAAGCAGGCCACCTACTTTCCCCCATCACCACCGAGCTGCAG
AGCCTGGAAGTCCTGATAGACAAGCTGAGATGCTCCAGAATCTGACCCCACCTCCGTCCTATGCTGCT
ACAATTGCTTCGAAACTGGCAATTCACAATCCAAATTTACCTGCCACTCTGCCAGTTAATTCGCCAAA
TATCCAACCTGTCCGATACAACAGAAGGAGTAACCCGGATCTGGAGAAGCGACGCATCCATTTCTGTG
ATTATGATGGTTGCACAAAAGTTTATACAAAGTCGTCTCATTTAAAAGCTCACCTGAGGACTCATACG
GGCGAGAAGCCCTACAAGTGCACCTGGGAGGGCTGCGACTGGAGGTTTGCCCGGTCGGACGAGCTGAC
CCGCCACTACAGGAAGCACACGGGTGCCAAGCCGTTCCAGTGCGTGGTGTGCAACCGCAGCTTCTCCC
GCTCCGACCACCTGGCGCTGCACATGAAGCGCCACCAGAACTGAGCACTGCGCACAACCGGCTCGACG
CCTCGCAGTCCGCTCGCCATCCTTTAAACCGCAGACCTAACTTCATATAAAAAAAAAAAAA

SEQ ID NO:38 Rat CKLF protein sequence
accession:gi16758122
MPSSTNQTAVMDTLNVSMAGLNSHPSAVPQTSMKQFQGMPPCTYTMPSQFLPQQATYFPPSPPSSEPG
SPDRQAEMLQNLTPPPSYAATIASKLAIHNPNLPATLPVNSPNIQPVRYNRRSNPDLEKRRIHFCDYD
GCTKVYTKSSHLKAHLRTHTGEKPYKCTWEGCDWRFARSDELTRHYRKHTGAKPFQCVVCNRSFSRSD
HLALHMKRHQN

[0258]

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 46

<210> SEQ ID NO 1
<211> LENGTH: 1909
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human Fritz (frizzled protein homolog) cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (70)..(1047)
<223> OTHER INFORMATION: Fritz
<220> FEATURE:

-continued

<221> NAME/KEY: modified_base
<222> LOCATION: (42)
<223> OTHER INFORMATION: n = g, a, c or t

<400> SEQUENCE: 1

```
cggagacggc ggagcgggcc ttgttggcgt ccactgcgcg gntgcaccct gccccatcct     60
gccgggatca tggtctgcgg cagcccggga gggatgctgc tgctgcgggc cgggctgctt    120
gccctggctg ctctctgcct gctccgggtg cccggggctc gggctgcagc ctgtgagccc    180
gtccgcatcc ccctgtgcaa gtccctgccc tggaacatga ctaagatgcc caaccacctg    240
caccacagca ctcaggacaa cgccatcctg gccatcgagc agttcgaagg tctgctgggc    300
acccactgca gccccgatct gctcttcttc ctctgtgcca tgtacgcgcc catctgcacc    360
attgacttcc agcacgagcc catcaagccc tgtaagtctg tgtgcgagcg ggcccggcag    420
ggctgtgagc ccatactcat caagtaccgc cactcgtggc cggagaacct ggcctgcgag    480
gagctgccag tgtacgacag gggcgtgtgc atctctcccg aggccatcgt tactgcggac    540
ggagctgatt ttcctatgga ttctagtaac ggaaactgta gaggggcaag cagtgaacgc    600
tgtaaatgta agcctattag agctacacag aagacctatt tccggaacaa ttacaactat    660
gtcattcggg ctaaagttaa agagataaag actaagtgcc atgatgtgac tgcagtagtg    720
gaggtgaagg agattctaaa gtcctctctg gtaaacattc cacgggacac tgtcaacctc    780
tataccagct ctggctgcct ctgccctcca cttaatgtta atgaggaata tatcatcatg    840
ggctatgaag atgaggaacg ttccagatta ctcttggtgg aaggctctat agctgagaag    900
tggaaggatc gactcggtaa aaaagttaag cgctgggata tgaagcttcg tcatcttgga    960
ctcagtaaaa gtgattctag caatagtgat tccactcaga gtcagaagtc tggcaggaac   1020
tcgaaccccc ggcaagcacg caactaaatc ccgaaataca aaaagtaaca cagtggactt   1080
cctattaaga cttacttgca ttgctggact agcaaaggaa aattgcacta ttgcacatca   1140
tattctattg tttactataa aaatcatgtg ataactgatt attacttctg tttctctttt   1200
ggtttctgct tctctcttct ctcaacccct ttgtaatggt ttgggggcag actcttaagt   1260
atattgtgag ttttctattt cactaatcat gagaaaaact gttcttttgc aataataata   1320
aattaaacat gctgttacca gagcctcttt gctggagtct ccagatgtta atttactttc   1380
tgcaccccaa ttgggaatgc aatattggat gaaaagagag gtttctggta ttcacagaaa   1440
gctagatatg ccttaaaaca tactctgccg atctaattac agccttattt ttgtatgcct   1500
tttgggcatt ctcctcatgc ttagaaagtt ccaaatgttt ataaaggtaa aatggcagtt   1560
tgaagtcaaa tgtcacatag gcaaagcaat caagcaccag gaagtgttta tgaggaaaca   1620
acacccaaga tgaattattt ttgagactgt caggaagtaa aataaatagg agcttaagaa   1680
agaacatttt gcctgattga gaagcacaac tgaaaccagt agccgctggg gtgttaatgg   1740
tagcattctt cttttggcaa tacatttgat ttgttcatga atatattaat cagcattaga   1800
gaaatgaatt ataactagac atctgctgtt atcaccatag ttttgtttaa tttgcttcct   1860
tttaaataaa cccattggtg aaagtcccaa aaaaaaaaaa aaaaaaaaa               1909
```

<210> SEQ ID NO 2
<211> LENGTH: 325
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:

-continued

<223> OTHER INFORMATION: human Fritz (frizzled protein homolog)

<400> SEQUENCE: 2

```
Met Val Cys Gly Ser Pro Gly Gly Met Leu Leu Leu Arg Ala Gly Leu
  1               5                  10                  15

Leu Ala Leu Ala Ala Leu Cys Leu Leu Arg Val Pro Gly Ala Arg Ala
             20                  25                  30

Ala Ala Cys Glu Pro Val Arg Ile Pro Leu Cys Lys Ser Leu Pro Trp
         35                  40                  45

Asn Met Thr Lys Met Pro Asn His Leu His His Ser Thr Gln Asp Asn
     50                  55                  60

Ala Ile Leu Ala Ile Glu Gln Phe Glu Gly Leu Leu Gly Thr His Cys
 65                  70                  75                  80

Ser Pro Asp Leu Leu Phe Phe Leu Cys Ala Met Tyr Ala Pro Ile Cys
                 85                  90                  95

Thr Ile Asp Phe Gln His Glu Pro Ile Lys Pro Cys Lys Ser Val Cys
            100                 105                 110

Glu Arg Ala Arg Gln Gly Cys Glu Pro Ile Leu Ile Lys Tyr Arg His
        115                 120                 125

Ser Trp Pro Glu Asn Leu Ala Cys Glu Glu Leu Pro Val Tyr Asp Arg
    130                 135                 140

Gly Val Cys Ile Ser Pro Glu Ala Ile Val Thr Ala Asp Gly Ala Asp
145                 150                 155                 160

Phe Pro Met Asp Ser Ser Asn Gly Asn Cys Arg Gly Ala Ser Ser Glu
                165                 170                 175

Arg Cys Lys Cys Lys Pro Ile Arg Ala Thr Gln Lys Thr Tyr Phe Arg
            180                 185                 190

Asn Asn Tyr Asn Tyr Val Ile Arg Ala Lys Val Lys Glu Ile Lys Thr
        195                 200                 205

Lys Cys His Asp Val Thr Ala Val Val Glu Val Lys Glu Ile Leu Lys
    210                 215                 220

Ser Ser Leu Val Asn Ile Pro Arg Asp Thr Val Asn Leu Tyr Thr Ser
225                 230                 235                 240

Ser Gly Cys Leu Cys Pro Pro Leu Asn Val Asn Glu Glu Tyr Ile Ile
                245                 250                 255

Met Gly Tyr Glu Asp Glu Glu Arg Ser Arg Leu Leu Leu Val Glu Gly
            260                 265                 270

Ser Ile Ala Glu Lys Trp Lys Asp Arg Leu Gly Lys Lys Val Lys Arg
        275                 280                 285

Trp Asp Met Lys Leu Arg His Leu Gly Leu Ser Lys Ser Asp Ser Ser
    290                 295                 300

Asn Ser Asp Ser Thr Gln Ser Gln Lys Ser Gly Arg Asn Ser Asn Pro
305                 310                 315                 320

Arg Gln Ala Arg Asn
                325
```

<210> SEQ ID NO 3
<211> LENGTH: 2540
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse Fritz (frizzled-related protein) cDNA
<220> FEATURE:
<221> NAME/KEY: CDS

-continued

<222> LOCATION: (365)..(1336)
<223> OTHER INFORMATION: Fritz
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(2540)
<223> OTHER INFORMATION: n = g, a, c or t

<400> SEQUENCE: 3

gaattcggca cgagctgaat ttgacttttg tttttatttc tctctggctt cctcttctgc 60

cccctcatct gattgatgtg ctaaggctga tgtctctgcc agagcgagag gaataaatag 120

atgctgcctc gcctagaggc ttagacgctt gggaagagca gccggcgcac gagcgaccgg 180

gctccgccaa gctagtggac cggacctggg agcacttgga tccaagagaa ctgtgattgt 240

cccaggggtg ggggcagctc cccaggtcgt tgggatcacc cctcggaacc gcaggggggag 300

acttcggaac gaaagtgtct cccgcgtccg tcgctcgtgc gccctgcccc atcctgctgg 360

gaccatggtc tgctgcggcc cgggacggat gctgctagga tgggccgggt tgctagtcct 420

ggctgctctc tgcctgctcc aggtgcccgg agctcaggct gcagcctgtg agcctgtccg 480

catcccgctg tgcaagtccc ttccctggaa catgaccaag atgcccaacc acctgcacca 540

cagcacccag gctaacgcca tcctggccat ggaacagttc gaagggctgc tgggcaccca 600

ctgcagcccg gatcttctct tcttcctctg tgcaatgtac gcacccattt gcaccatcga 660

cttccagcac gagcccatca agccctgcaa gtctgtgtgt gagcgcgccc gacagggctg 720

cgagcccatt ctcatcaagt accgccactc gtggccggaa agcttggcct gcgacgagct 780

gccggtgtac gaccgcggcg tgtgcatctc tcctgaggcc atcgtcaccg cggacggagc 840

ggattttcct atggattcaa gtactggaca ctgcagaggg gcaagcagcg aacgttgcaa 900

atgtaagcct gtcagagcta cacagaagac ctatttccgg aacaattaca actatgtcat 960

ccgggctaaa gttaaagagg taaagatgaa atgtcatgat gtgaccgccg ttgtggaagt 1020

gaaggaaatt ctaaaggcat cactggtaaa cattccaagg gacaccgtca atctttatac 1080

cacctctggc tgcctctgtc ctccacttac tgtcaatgag gaatatgtca tcatgggcta 1140

tgaagacgag gaacgttcca ggttactctt ggtagaaggc tctatagctg agaagtggaa 1200

ggatcggctt ggtaagaaag tcaagcgctg ggatatgaaa ctccgacacc ttggactggg 1260

taaaactgat gctagcgatt ccactcagaa tcagaagtct ggcaggaact ctaatccccg 1320

gccagcacgc agctaaatcc tgaaatgtaa aaggccacac ccacggactc ccttctaaga 1380

ctggcgctgc tggactaaca aaggaaaacg cacagttgtg ctcgtgaccg attgtttacc 1440

gcagacaccg cgtggctacc gaagttactt ccggtcccct ttctcctgct tcttaatggc 1500

ctggggttag atcctttaat atgttatata ttctgtttca tcaatcacgt ggggactgtt 1560

cttttgcaac cagaatagta aattaaatat gttgatgcta aggtttctgt actggactcc 1620

ctgggtttaa tttggtgttc tgtaccctga ttgagaatgc aatgtttcat gtaaagagag 1680

aatcctggtc atatctcaag aactagatat tgctgtaaga cagcctctgc tgctgcgctt 1740

atagtcttgt gtttgtacct gtttggccat ttccctcatg ctgtgaaagt tatacatgtt 1800

tataaaggta gaacggcatt ttgaaatcag acactgcaca agcagagtag cccaacacca 1860

ggaagcattt atgaggaaac gccacacagc atgacttatt ttcaagattg gcaggcagca 1920

aaataaatag tgttgggagc caagaaaaga atattttgcc tggttaaggg gcacactgga 1980

atcagtagcc ttgagccatt aacagcagtg ttcttctggc acgtttttga tttgttcata 2040

-continued

```
aatgtattca cgagcattag agatgaactt ataactagac atctgttgtt atcactatag   2100

ctctgcttcc ttctaaatca aacccattgt tggatgctcc ctctccattc ataaataaat   2160

ttggcttgct ggtattggcc aggaaaagaa agtattaaag tatgcatgca tgtgcaccag   2220

ggtgttattt aacagaggta tgtaactcta taaaagacta taatttacag gacacggaaa   2280

tgtgcacatt tgtttacttt ttttcttcct tttgctttgg gcttgtgatt ttggtttttg   2340

gtgtgtttat gtctgtattt tggggggtgg gtaggtttaa nccattgcac attcaagttg   2400

nactagatta gagtagacta ggctcattgg cctagacatt atgatttgaa tttgtgttgt   2460

ttaatgctcc atcaagatgt ctaataaaag gaatatggtt gtcaacagag acgacaacac   2520

ccaaaaaaaa aaaaaaaaaa                                               2540
```

```
<210> SEQ ID NO 4
<211> LENGTH: 323
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse Fritz (frizzled-related protein)

<400> SEQUENCE: 4

Met Val Cys Cys Gly Pro Gly Arg Met Leu Leu Gly Trp Ala Gly Leu
  1               5                  10                  15

Leu Val Leu Ala Ala Leu Cys Leu Leu Gln Val Pro Gly Ala Gln Ala
             20                  25                  30

Ala Ala Cys Glu Pro Val Arg Ile Pro Leu Cys Lys Ser Leu Pro Trp
         35                  40                  45

Asn Met Thr Lys Met Pro Asn His Leu His His Ser Thr Gln Ala Asn
     50                  55                  60

Ala Ile Leu Ala Met Glu Gln Phe Glu Gly Leu Leu Gly Thr His Cys
 65                  70                  75                  80

Ser Pro Asp Leu Leu Phe Phe Leu Cys Ala Met Tyr Ala Pro Ile Cys
                 85                  90                  95

Thr Ile Asp Phe Gln His Glu Pro Ile Lys Pro Cys Lys Ser Val Cys
            100                 105                 110

Glu Arg Ala Arg Gln Gly Cys Glu Pro Ile Leu Ile Lys Tyr Arg His
        115                 120                 125

Ser Trp Pro Glu Ser Leu Ala Cys Asp Glu Leu Pro Val Tyr Asp Arg
    130                 135                 140

Gly Val Cys Ile Ser Pro Glu Ala Ile Val Thr Ala Asp Gly Ala Asp
145                 150                 155                 160

Phe Pro Met Asp Ser Ser Thr Gly His Cys Arg Gly Ala Ser Ser Glu
                165                 170                 175

Arg Cys Lys Cys Lys Pro Val Arg Ala Thr Gln Lys Thr Tyr Phe Arg
            180                 185                 190

Asn Asn Tyr Asn Tyr Val Ile Arg Ala Lys Val Lys Glu Val Lys Met
        195                 200                 205

Lys Cys His Asp Val Thr Ala Val Val Glu Val Lys Glu Ile Leu Lys
    210                 215                 220

Ala Ser Leu Val Asn Ile Pro Arg Asp Thr Val Asn Leu Tyr Thr Thr
225                 230                 235                 240

Ser Gly Cys Leu Cys Pro Pro Leu Thr Val Asn Glu Glu Tyr Val Ile
                245                 250                 255
```

-continued

```
Met Gly Tyr Glu Asp Glu Glu Arg Ser Arg Leu Leu Leu Val Glu Gly
            260                 265                 270

Ser Ile Ala Glu Lys Trp Lys Asp Arg Leu Gly Lys Lys Val Lys Arg
        275                 280                 285

Trp Asp Met Lys Leu Arg His Leu Gly Leu Gly Lys Thr Asp Ala Ser
    290                 295                 300

Asp Ser Thr Gln Asn Gln Lys Ser Gly Arg Asn Ser Asn Pro Arg Pro
305                 310                 315                 320

Ala Arg Ser


<210> SEQ ID NO 5
<211> LENGTH: 1740
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human p21 activated kinase 1B (PAK1B) cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (55)..(1692)
<223> OTHER INFORMATION: PAK1B

<400> SEQUENCE: 5

ggagagccga gaggagctga gcgagcgcgg aagtagctgc tgctggtggt gacaatgtca     60

aataacggcc tagacattca agacaaaccc ccagcccctc cgatgagaaa taccagcact    120

atgattggag ccggcagcaa agatgctgga accctaaacc atggttctaa acctctgcct    180

ccaaacccag aggagaagaa aaagaaggac cgattttacc gatccatttt acctggagat    240

aaaacaaata aaaagaaaga gaaagagcgg ccagagattt ctctcccttc agattttgaa    300

cacacaattc atgtcggttt tgatgctgtc acaggggagt ttaccggaat gccagagcag    360

tgggcccgct tgcttcagac atcaaatatc actaagtcgg agcagaagaa aaacccgcag    420

gctgttctgg atgtgttgga gttttacaac tcgaagaaga catccaacag ccagaaatac    480

atgagcttta cagataagtc agctgaggat tacaattctt ctaatgcctt gaatgtgaag    540

gctgtgtctg agactcctgc agtgccacca gtttcagaag atgaggatga tgatgatgat    600

gatgctaccc caccaccagt gattgctcca cgcccagagc acacaaaatc tgtatacaca    660

cggtctgtga ttgaaccact tcctgtcact ccaactcggg acgtggctac atctcccatt    720

tcacctactg aaaataacac cactccacca gatgctttga cccttaatac tgagaagcag    780

aagaagaagc ctaaaatgtc tgatgaggag atcttggaga aattacgaag catagtgagt    840

gtgggcgatc ctaagaagaa atatacacgg tttgagaaga ttggacaagg tgcttcaggc    900

accgtgtaca cagcaatgga tgtggccaca ggacaggagg tggccattaa gcagatgaat    960

cttcagcagc agcccaagaa agagctgatt attaatgaga tcctggtcat gagggaaaac   1020

aagaacccaa acattgtgaa ttacttggac agttacctcg tgggagatga gctgtgggtt   1080

gttatggaat acttggctgg aggctccttg acagatgtgg tgacagaaac ttgcatggat   1140

gaaggccaaa ttgcagctgt gtgccgtgag tgtctgcagg ctctggagtc tttgcattcg   1200

aaccaggtca ttcacagaga catcaagagt gacaatattc tgttgggaat ggatggctct   1260

gtcaagctaa ctgactttgg attctgtgca cagataaccc cagagcagag caaacggagc   1320

accatggtag gaaccccata ctggatggca ccagaggttg tgacacgaaa ggcctatggg   1380

cccaaggttg acatctggtc cctgggcatc atggccatcg aaatgattga aggggagcct   1440
```

-continued

```
ccatacctca atgaaaaccc tctgagagcc ttgtacctca ttgccaccaa tgggacccca   1500

gaacttcaga acccagagaa gctgtcagct atcttccggg actttctgaa ccgctgtctc   1560

gagatggatg tggagaagag aggttcagct aaagagctgc tacagcatca attcctgaag   1620

attgccaagc ccctctccag cctcactcca ctgattgctg cagctaagga ggcaacaaag   1680

aacaatcact aaaaccacac tcaccccagc ctcattgtgc caagccttct gtgagataaa   1740


<210> SEQ ID NO 6
<211> LENGTH: 545
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human p21 activated kinase 1B (PAK1B)

<400> SEQUENCE: 6

Met Ser Asn Asn Gly Leu Asp Ile Gln Asp Lys Pro Pro Ala Pro Pro
  1               5                  10                  15

Met Arg Asn Thr Ser Thr Met Ile Gly Ala Gly Ser Lys Asp Ala Gly
             20                  25                  30

Thr Leu Asn His Gly Ser Lys Pro Leu Pro Pro Asn Pro Glu Glu Lys
         35                  40                  45

Lys Lys Lys Asp Arg Phe Tyr Arg Ser Ile Leu Pro Gly Asp Lys Thr
     50                  55                  60

Asn Lys Lys Lys Glu Lys Glu Arg Pro Glu Ile Ser Leu Pro Ser Asp
 65                  70                  75                  80

Phe Glu His Thr Ile His Val Gly Phe Asp Ala Val Thr Gly Glu Phe
                 85                  90                  95

Thr Gly Met Pro Glu Gln Trp Ala Arg Leu Leu Gln Thr Ser Asn Ile
            100                 105                 110

Thr Lys Ser Glu Gln Lys Lys Asn Pro Gln Ala Val Leu Asp Val Leu
        115                 120                 125

Glu Phe Tyr Asn Ser Lys Lys Thr Ser Asn Ser Gln Lys Tyr Met Ser
    130                 135                 140

Phe Thr Asp Lys Ser Ala Glu Asp Tyr Asn Ser Ser Asn Ala Leu Asn
145                 150                 155                 160

Val Lys Ala Val Ser Glu Thr Pro Ala Val Pro Pro Val Ser Glu Asp
                165                 170                 175

Glu Asp Asp Asp Asp Asp Asp Ala Thr Pro Pro Pro Val Ile Ala Pro
            180                 185                 190

Arg Pro Glu His Thr Lys Ser Val Tyr Thr Arg Ser Val Ile Glu Pro
        195                 200                 205

Leu Pro Val Thr Pro Thr Arg Asp Val Ala Thr Ser Pro Ile Ser Pro
    210                 215                 220

Thr Glu Asn Asn Thr Thr Pro Pro Asp Ala Leu Thr Leu Asn Thr Glu
225                 230                 235                 240

Lys Gln Lys Lys Lys Pro Lys Met Ser Asp Glu Glu Ile Leu Glu Lys
                245                 250                 255

Leu Arg Ser Ile Val Ser Val Gly Asp Pro Lys Lys Lys Tyr Thr Arg
            260                 265                 270

Phe Glu Lys Ile Gly Gln Gly Ala Ser Gly Thr Val Tyr Thr Ala Met
        275                 280                 285

Asp Val Ala Thr Gly Gln Glu Val Ala Ile Lys Gln Met Asn Leu Gln
```

-continued

```
    290                 295                 300

Gln Gln Pro Lys Lys Glu Leu Ile Ile Asn Glu Ile Leu Val Met Arg
305                 310                 315                 320

Glu Asn Lys Asn Pro Asn Ile Val Asn Tyr Leu Asp Ser Tyr Leu Val
                325                 330                 335

Gly Asp Glu Leu Trp Val Val Met Glu Tyr Leu Ala Gly Gly Ser Leu
            340                 345                 350

Thr Asp Val Val Thr Glu Thr Cys Met Asp Glu Gly Gln Ile Ala Ala
        355                 360                 365

Val Cys Arg Glu Cys Leu Gln Ala Leu Glu Ser Leu His Ser Asn Gln
    370                 375                 380

Val Ile His Arg Asp Ile Lys Ser Asp Asn Ile Leu Leu Gly Met Asp
385                 390                 395                 400

Gly Ser Val Lys Leu Thr Asp Phe Gly Phe Cys Ala Gln Ile Thr Pro
                405                 410                 415

Glu Gln Ser Lys Arg Ser Thr Met Val Gly Thr Pro Tyr Trp Met Ala
            420                 425                 430

Pro Glu Val Val Thr Arg Lys Ala Tyr Gly Pro Lys Val Asp Ile Trp
        435                 440                 445

Ser Leu Gly Ile Met Ala Ile Glu Met Ile Glu Gly Glu Pro Pro Tyr
    450                 455                 460

Leu Asn Glu Asn Pro Leu Arg Ala Leu Tyr Leu Ile Ala Thr Asn Gly
465                 470                 475                 480

Thr Pro Glu Leu Gln Asn Pro Glu Lys Leu Ser Ala Ile Phe Arg Asp
                485                 490                 495

Phe Leu Asn Arg Cys Leu Glu Met Asp Val Glu Lys Arg Gly Ser Ala
            500                 505                 510

Lys Glu Leu Leu Gln His Gln Phe Leu Lys Ile Ala Lys Pro Leu Ser
        515                 520                 525

Ser Leu Thr Pro Leu Ile Ala Ala Ala Lys Glu Ala Thr Lys Asn Asn
    530                 535                 540

His
545


<210> SEQ ID NO 7
<211> LENGTH: 1880
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human p21 activated kinase 1B (PAK1B) splice
      variant cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (12)..(1673)
<223> OTHER INFORMATION: PAK1B splice variant

<400> SEQUENCE: 7

tggtggtgac aatgtcaaat aacggcctag acattcaaga caaacccccca gcccctccga     60

tgagaaatac cagcactatg attggagtcg gcagcaaaga tgctggaacc ctaaaccatg    120

gttctaaacc tctgcctcca aacccagagg agaagaaaaa gaaggaccga ttttaccgat    180

ccattttacc tggagataaa acaaataaaa agaaagagaa agagcggcca gagatttctc    240

tcccttcaga ttttgaacac acaattcatg tcggttttga tgctgtcaca ggggagttta    300

cgggaatgcc agagcagtgg gcccgcttgc ttcagacatc aaatatcact aagtcggagc    360
```

-continued

```
agaagaaaaa cccgcaggct gttctggatg tgttggagtt ttacaactcg aagaagacat    420

ccaacagcca gaaatacatg agctttacag ataagtcagc tgaggattac aattcttcta    480

atgccttgaa tgtgaaggct gtgtctgaga ctcctgcagt gccaccagtt tcagaagatg    540

aggatgatga tgatgatgat gctaccccac caccagtgat tgctccacgc ccagagcaca    600

caaaatctgt atacacacgg tctgtgattg aaccacttcc tgtcactcca actcgggacg    660

tggctacatc tcccatttca cctactgaaa ataacaccac tccaccagat gctttgaccc    720

ggaatactga gaagcagaag aagaagccta aaatgtctga tgaggagatc ttggagaaat    780

tacgaagcat agtgagtgtg ggcgatccta agaagaaata tacacggttt gagaagattg    840

gacaaggtgc ttcaggcacc gtgtacacag caatggatgt ggccacagga caggaggtgg    900

ccattaagca gatgaatctt cagcagcagc ccaagaaaga gctgattatt aatgagatcc    960

tggtcatgag ggaaaacaag aacccaaaca ttgtgaatta cttggacagt tacctcgtgg   1020

gagatgagct gtgggttgtt atggaatact tggctggagg ctccttgaca gatgtggtga   1080

cagaaacttg catggatgaa ggccaaattg cagctgtgtg ccgtgagtgt ctgcaggctc   1140

tggagttctt gcattcgaac caggtcattc acagagacat caagagtgac aatattctgt   1200

tgggaatgga tggctctgtc aagctaactg actttggatt ctgtgcacag ataaccccag   1260

agcagagcaa acggagcacc atggtaggaa ccccatactg gatggcacca gaggttgtga   1320

cacgaaaggc ctatgggccc aaggttgaca tctggtccct gggcatcatg gccatcgaaa   1380

tgattgaagg ggagcctcca tacctcaatg aaaaccctct gagagccttg tacctcattg   1440

ccaccaatgg gaccccagaa cttcagaacc cagagaagct gtcagctatc ttccgggact   1500

ttctgaaccg ctgtctcgag atggatgtgg agaagagagg ttcagctaaa gagctgctac   1560

aggtgagaaa actgaggttt caagtgttta gtaacttttc catgatagct gcatcaattc   1620

ctgaagattg ccaagcccct ctccagcctc actccactga ttgctgcagc taaggaggca   1680

acaaagaaca atcactaaaa ccacactcac cccagcctca ttgtgccaag ctctgtgaga   1740

taaatgcaca tttcagaaat tccaactcct gatgccctct tctccttgcc ttgcttctcc   1800

catttcctga tctagcactc ctcaagactt tgatccttgg aaaccgtgtg tccagcattg   1860

aagagaactg caactgaatg                                                1880


<210> SEQ ID NO 8
<211> LENGTH: 553
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human p21 activated kinase 1B (PAK1B) splice
      variant

<400> SEQUENCE: 8

Met Ser Asn Asn Gly Leu Asp Ile Gln Asp Lys Pro Pro Ala Pro Pro
  1               5                  10                  15

Met Arg Asn Thr Ser Thr Met Ile Gly Val Gly Ser Lys Asp Ala Gly
             20                  25                  30

Thr Leu Asn His Gly Ser Lys Pro Leu Pro Pro Asn Pro Glu Glu Lys
         35                  40                  45

Lys Lys Lys Asp Arg Phe Tyr Arg Ser Ile Leu Pro Gly Asp Lys Thr
     50                  55                  60
```

-continued

```
Asn Lys Lys Lys Glu Lys Glu Arg Pro Glu Ile Ser Leu Pro Ser Asp
 65                  70                  75                  80

Phe Glu His Thr Ile His Val Gly Phe Asp Ala Val Thr Gly Glu Phe
                85                  90                  95

Thr Gly Met Pro Glu Gln Trp Ala Arg Leu Leu Gln Thr Ser Asn Ile
            100                 105                 110

Thr Lys Ser Glu Gln Lys Lys Asn Pro Gln Ala Val Leu Asp Val Leu
        115                 120                 125

Glu Phe Tyr Asn Ser Lys Lys Thr Ser Asn Ser Gln Lys Tyr Met Ser
    130                 135                 140

Phe Thr Asp Lys Ser Ala Glu Asp Tyr Asn Ser Ser Asn Ala Leu Asn
145                 150                 155                 160

Val Lys Ala Val Ser Glu Thr Pro Ala Val Pro Pro Val Ser Glu Asp
                165                 170                 175

Glu Asp Asp Asp Asp Asp Asp Ala Thr Pro Pro Pro Val Ile Ala Pro
            180                 185                 190

Arg Pro Glu His Thr Lys Ser Val Tyr Thr Arg Ser Val Ile Glu Pro
        195                 200                 205

Leu Pro Val Thr Pro Thr Arg Asp Val Ala Thr Ser Pro Ile Ser Pro
    210                 215                 220

Thr Glu Asn Asn Thr Thr Pro Pro Asp Ala Leu Thr Arg Asn Thr Glu
225                 230                 235                 240

Lys Gln Lys Lys Lys Pro Lys Met Ser Asp Glu Glu Ile Leu Glu Lys
                245                 250                 255

Leu Arg Ser Ile Val Ser Val Gly Asp Pro Lys Lys Lys Tyr Thr Arg
            260                 265                 270

Phe Glu Lys Ile Gly Gln Gly Ala Ser Gly Thr Val Tyr Thr Ala Met
        275                 280                 285

Asp Val Ala Thr Gly Gln Glu Val Ala Ile Lys Gln Met Asn Leu Gln
    290                 295                 300

Gln Gln Pro Lys Lys Glu Leu Ile Ile Asn Glu Ile Leu Val Met Arg
305                 310                 315                 320

Glu Asn Lys Asn Pro Asn Ile Val Asn Tyr Leu Asp Ser Tyr Leu Val
                325                 330                 335

Gly Asp Glu Leu Trp Val Val Met Glu Tyr Leu Ala Gly Gly Ser Leu
            340                 345                 350

Thr Asp Val Val Thr Glu Thr Cys Met Asp Glu Gly Gln Ile Ala Ala
        355                 360                 365

Val Cys Arg Glu Cys Leu Gln Ala Leu Glu Phe Leu His Ser Asn Gln
    370                 375                 380

Val Ile His Arg Asp Ile Lys Ser Asp Asn Ile Leu Leu Gly Met Asp
385                 390                 395                 400

Gly Ser Val Lys Leu Thr Asp Phe Gly Phe Cys Ala Gln Ile Thr Pro
                405                 410                 415

Glu Gln Ser Lys Arg Ser Thr Met Val Gly Thr Pro Tyr Trp Met Ala
            420                 425                 430

Pro Glu Val Val Thr Arg Lys Ala Tyr Gly Pro Lys Val Asp Ile Trp
        435                 440                 445

Ser Leu Gly Ile Met Ala Ile Glu Met Ile Glu Gly Glu Pro Pro Tyr
    450                 455                 460

Leu Asn Glu Asn Pro Leu Arg Ala Leu Tyr Leu Ile Ala Thr Asn Gly
```

-continued

```
465                 470                 475                 480

Thr Pro Glu Leu Gln Asn Pro Glu Lys Leu Ser Ala Ile Phe Arg Asp
                485                 490                 495

Phe Leu Asn Arg Cys Leu Glu Met Asp Val Glu Lys Arg Gly Ser Ala
            500                 505                 510

Lys Glu Leu Leu Gln Val Arg Lys Leu Arg Phe Gln Val Phe Ser Asn
        515                 520                 525

Phe Ser Met Ile Ala Ala Ser Ile Pro Glu Asp Cys Gln Ala Pro Leu
    530                 535                 540

Gln Pro His Ser Thr Asp Cys Cys Ser
545                 550


<210> SEQ ID NO 9
<211> LENGTH: 1347
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human p21 activated kinase 1B (PAK1B) new
      splice variant

<400> SEQUENCE: 9

tcaaataacg gcctagacat tcaagacaaa cccccagccc ctccgatgag aaataccagc     60

actatgattg gagccggcag caaagatgct ggaaccctaa accatggttc taaacctctg    120

cctccaaacc cagaggagaa gaaaaagaag gaccgatttt accgatccat tttacctgga    180

gataaaacaa ataaaaagaa agagaaagag cggccagaga tttctctccc ttcagatttt    240

gaacacacaa ttcatgtcgg ttttgatgct gtcacagggg agtttacggg aatgccagag    300

cagtgggccc gcttgcttca gacatcaaat atcactaagt cggagcagaa gaaaaacccg    360

caggctgttc tggatgtgtt ggagttttac aactcgaaga agacatccaa cagccagaaa    420

tacatgagct ttacagataa gtcagctgag gattacaatt cttctaatgc cttgaatgtg    480

aaggctgtgt ctgagactcc tgcagtgcca ccagtttcag aagatgagga tgatgatgat    540

gatgatgcta ccccaccacc agtgattgct ccacgcccag agcacacaaa atctgtggcc    600

attaagcaga tgaatcttca gcagcagccc aagaaagagc tgattattaa tgagatcctg    660

gtcatgaggg aaaacaagaa cccaaacatt gtgaattact tggacagtta cctcgtggga    720

gatgagctgt gggttgttat ggaatacttg gctggaggct ccttgacaga tgtggtgaca    780

gaaacttgca tggatgaagg ccaaattgca gctgtgtgcc gtgagtgtct gcaggctctg    840

gagttcttgc attcgaacca ggtcattcac agagacatca agagtgacaa tattctgttg    900

ggaatggatg gctctgtcaa gctaactgac tttggattct gtgcacagat aaccccagag    960

cagagcaaac ggagcaccat ggtaggaacc ccatactgga tggcaccaga ggttgtgaca   1020

cgaaaggcct atgggcccaa ggttgacatc tggtccctgg gcatcatggc catcgaaatg   1080

attgaagggg agcctccata cctcaatgaa aaccctctga gagccttgta cctcattgcc   1140

accaatggga ccccagaact tcagaaccca gagaagctgt cagctatctt ccgggacttt   1200

ctgaaccgct gtctcgggat ggatgtggag aagagaggtt cagctaaaga gctgctacag   1260

catcaattcc tgaagattgc caagcccctc tccagcctca ctccactgat tgctgcagct   1320

aaggaggcaa caaagaacaa tcactaa                                       1347


<210> SEQ ID NO 10
```

-continued

```
<211> LENGTH: 449
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human p21 activated kinase 1B (PAK1B) new
      splice variant

<400> SEQUENCE: 10

Met Ser Asn Asn Gly Leu Asp Ile Gln Asp Lys Pro Pro Ala Pro Pro
  1               5                  10                  15

Met Arg Asn Thr Ser Thr Met Ile Gly Ala Gly Ser Lys Asp Ala Gly
             20                  25                  30

Thr Leu Asn His Gly Ser Lys Pro Leu Pro Pro Asn Pro Glu Glu Lys
         35                  40                  45

Lys Lys Lys Asp Arg Phe Tyr Arg Ser Ile Leu Pro Gly Asp Lys Thr
     50                  55                  60

Asn Lys Lys Lys Glu Lys Glu Arg Pro Glu Ile Ser Leu Pro Ser Asp
 65                  70                  75                  80

Phe Glu His Thr Ile His Val Gly Phe Asp Ala Val Thr Gly Glu Phe
                 85                  90                  95

Thr Gly Met Pro Glu Gln Trp Ala Arg Leu Leu Gln Thr Ser Asn Ile
            100                 105                 110

Thr Lys Ser Glu Gln Lys Lys Asn Pro Gln Ala Val Leu Asp Val Leu
        115                 120                 125

Glu Phe Tyr Asn Ser Lys Lys Thr Ser Asn Ser Gln Lys Tyr Met Ser
    130                 135                 140

Phe Thr Asp Lys Ser Ala Glu Asp Tyr Asn Ser Ser Asn Ala Leu Asn
145                 150                 155                 160

Val Lys Ala Val Ser Glu Thr Pro Ala Val Pro Pro Val Ser Glu Asp
                165                 170                 175

Glu Asp Asp Asp Asp Asp Asp Ala Thr Pro Pro Pro Val Ile Ala Pro
            180                 185                 190

Arg Pro Glu His Thr Lys Ser Val Ala Ile Lys Gln Met Asn Leu Gln
        195                 200                 205

Gln Gln Pro Lys Lys Glu Leu Ile Ile Asn Glu Ile Leu Val Met Arg
    210                 215                 220

Glu Asn Lys Asn Pro Asn Ile Val Asn Tyr Leu Asp Ser Tyr Leu Val
225                 230                 235                 240

Gly Asp Glu Leu Trp Val Val Met Glu Tyr Leu Ala Gly Gly Ser Leu
                245                 250                 255

Thr Asp Val Val Thr Glu Thr Cys Met Asp Glu Gly Gln Ile Ala Ala
            260                 265                 270

Val Cys Arg Glu Cys Leu Gln Ala Leu Glu Phe Leu His Ser Asn Gln
        275                 280                 285

Val Ile His Arg Asp Ile Lys Ser Asp Asn Ile Leu Leu Gly Met Asp
    290                 295                 300

Gly Ser Val Lys Leu Thr Asp Phe Gly Phe Cys Ala Gln Ile Thr Pro
305                 310                 315                 320

Glu Gln Ser Lys Arg Ser Thr Met Val Gly Thr Pro Tyr Trp Met Ala
                325                 330                 335

Pro Glu Val Val Thr Arg Lys Ala Tyr Gly Pro Lys Val Asp Ile Trp
            340                 345                 350

Ser Leu Gly Ile Met Ala Ile Glu Met Ile Glu Gly Glu Pro Pro Tyr
```

-continued

```
        355                 360                 365

Leu Asn Glu Asn Pro Leu Arg Ala Leu Tyr Leu Ile Ala Thr Asn Gly
    370                 375                 380

Thr Pro Glu Leu Gln Asn Pro Glu Lys Leu Ser Ala Ile Phe Arg Asp
385                 390                 395                 400

Phe Leu Asn Arg Cys Leu Gly Met Asp Val Glu Lys Arg Gly Ser Ala
                405                 410                 415

Lys Glu Leu Leu Gln His Gln Phe Leu Lys Ile Ala Lys Pro Leu Ser
            420                 425                 430

Ser Leu Thr Pro Leu Ile Ala Ala Ala Lys Glu Ala Thr Lys Asn Asn
        435                 440                 445

His
```

<210> SEQ ID NO 11
<211> LENGTH: 1977
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse p21 (CDKN1A)-activated kinase 1B (PAK1B) cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (190)..(1827)
<223> OTHER INFORMATION: PAK1B

<400> SEQUENCE: 11

```
tgagtgtgct gctccttggg acgtgggtgg ctgcgctgcg gctgttagcc gaatccagcc     60

tgtcaggctg gttcagcccg ggctttgtcc caactgctga gcagggagag gtgcggctct    120

cgagactcac agatacacaa gatcacgccc cgcacccacc gccagtagct gctgctgctg    180

gtggtgacaa tgtcaaataa cggcgtagac atccaggaca aaccccccagc ccctccgatg    240

agaaacacca gcactatgat tggagccggc agcaaagaca ctggaaccct aaaccacggc    300

tccaaacctc tgcctccaaa cccagaggag aagaaaaaga aggaccggtt ttatcgatcc    360

atcttacctg gagataaaac aaataaaaag agggagaagg agcgaccaga gatttctctt    420

ccttcagatt ttgagcatac aattcatgtt ggttttgatg ctgtcacagg ggagtttacg    480

ggaatgccag aacagtgggc tcgcttgctt caaacatcaa atatcacaaa gtcagagcag    540

aagaaaaacc cacaggctgt tctggatgtg ttggaatttt ataactctaa gaagacctcc    600

aatagtaaga agtacatgag ttttacagat aagtcagctg aagattataa ttcttctaac    660

actttgaatg tgaagactgt gtctgagacc ccagcagtac caccagtgtc agaagatgat    720

gaagatgatg atgacgatgc taccccacct ccagtgattg ctccacgccc agaacacaca    780

aaatctgtat atacacgatc tgtgattgaa ccacttcctg ttactccaac tcgggatgtg    840

gctacatctc ctatttctcc tactgagaat aacaccactc cgccagatgc tttgacccgg    900

aacacggaaa aacagaagaa gaagcctaaa atgtctgatg aggagatctt agagaaatta    960

cggagcatag tgagtgtggg tgaccccaag aagaagtaca caccgttcga gaagattgga   1020

caaggtgctt caggcacagt gtatactgca atggatgtag ccacagggca ggaggtggcc   1080

attaaacaga tgaatcttca gcagcagccg aagaaagagc tgattattaa tgagatcctg   1140

gtcatgaggg aaaacaaaaa cccaaatatt gtcaactacc tggacagtta ccttgtggga   1200

gatgagctgt gggttgttat ggaatacttg gctggaggct ccttgacaga tgtggtgaca   1260
```

-continued

```
gaaacctgta tggatgaagg ccagatagca gctgtgtgcc gagagtgtct acaagctttg   1320

gagtttctac attcaaacca agtcattcac agggacatca agagtgacaa tattctgctg   1380

ggaatggatg gctctgtcaa gttaactgac tttggattct gtgcacagat aactccagag   1440

cagagcaaaa ggagcaccat ggtgggaact ccatattgga tggcacctga agttgtgaca   1500

cgcaaggctt atggacccaa ggttgacatc tggtccctgg gcattatggc aattgaaatg   1560

attgaggggg agcccccata cctcaatgaa aaccctttga gagccttgta cctcattgct   1620

accaatggga cgccagagct tcagaaccca gagaagttgt cagctatctt ccgggacttt   1680

ctgcaatgct gtcttgagat ggatgtggag aagagaggct cagctaaaga gctgctgcag   1740

catcagttcc tgaagattgc caagcccctc tctagcctga ctccactgat gcatgctgca   1800

aaagaggcaa ccaagaacaa tcactgaaac catgctcatc ccagcctcat gtgccaagcc   1860

ttctatgaaa taaacacttg tttcgggaac tccgacacct catgtcctct tctcctttcc   1920

ttgcttctcc catttcctga tctagtgctc ccaagacttt gatccttgga aactgtc      1977
```

```
<210> SEQ ID NO 12
<211> LENGTH: 545
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse p21 (CDKN1A)-activated kinase 1B (PAK1B)

<400> SEQUENCE: 12

Met Ser Asn Asn Gly Val Asp Ile Gln Asp Lys Pro Pro Ala Pro Pro
  1               5                  10                  15

Met Arg Asn Thr Ser Thr Met Ile Gly Ala Gly Ser Lys Asp Thr Gly
             20                  25                  30

Thr Leu Asn His Gly Ser Lys Pro Leu Pro Pro Asn Pro Glu Glu Lys
         35                  40                  45

Lys Lys Lys Asp Arg Phe Tyr Arg Ser Ile Leu Pro Gly Asp Lys Thr
     50                  55                  60

Asn Lys Lys Arg Glu Lys Glu Arg Pro Glu Ile Ser Leu Pro Ser Asp
 65                  70                  75                  80

Phe Glu His Thr Ile His Val Gly Phe Asp Ala Val Thr Gly Glu Phe
                 85                  90                  95

Thr Gly Met Pro Glu Gln Trp Ala Arg Leu Leu Gln Thr Ser Asn Ile
            100                 105                 110

Thr Lys Ser Glu Gln Lys Lys Asn Pro Gln Ala Val Leu Asp Val Leu
        115                 120                 125

Glu Phe Tyr Asn Ser Lys Lys Thr Ser Asn Ser Lys Lys Tyr Met Ser
    130                 135                 140

Phe Thr Asp Lys Ser Ala Glu Asp Tyr Asn Ser Ser Asn Thr Leu Asn
145                 150                 155                 160

Val Lys Thr Val Ser Glu Thr Pro Ala Val Pro Pro Val Ser Glu Asp
                165                 170                 175

Asp Glu Asp Asp Asp Asp Asp Ala Thr Pro Pro Pro Val Ile Ala Pro
            180                 185                 190

Arg Pro Glu His Thr Lys Ser Val Tyr Thr Arg Ser Val Ile Glu Pro
        195                 200                 205

Leu Pro Val Thr Pro Thr Arg Asp Val Ala Thr Ser Pro Ile Ser Pro
    210                 215                 220
```

-continued

```
Thr Glu Asn Asn Thr Thr Pro Pro Asp Ala Leu Thr Arg Asn Thr Glu
225                 230                 235                 240

Lys Gln Lys Lys Lys Pro Lys Met Ser Asp Glu Glu Ile Leu Glu Lys
                245                 250                 255

Leu Arg Ser Ile Val Ser Val Gly Asp Pro Lys Lys Lys Tyr Thr Pro
            260                 265                 270

Phe Glu Lys Ile Gly Gln Gly Ala Ser Gly Thr Val Tyr Thr Ala Met
        275                 280                 285

Asp Val Ala Thr Gly Gln Glu Val Ala Ile Lys Gln Met Asn Leu Gln
    290                 295                 300

Gln Gln Pro Lys Lys Glu Leu Ile Ile Asn Glu Ile Leu Val Met Arg
305                 310                 315                 320

Glu Asn Lys Asn Pro Asn Ile Val Asn Tyr Leu Asp Ser Tyr Leu Val
                325                 330                 335

Gly Asp Glu Leu Trp Val Val Met Glu Tyr Leu Ala Gly Gly Ser Leu
            340                 345                 350

Thr Asp Val Val Thr Glu Thr Cys Met Asp Glu Gly Gln Ile Ala Ala
        355                 360                 365

Val Cys Arg Glu Cys Leu Gln Ala Leu Glu Phe Leu His Ser Asn Gln
    370                 375                 380

Val Ile His Arg Asp Ile Lys Ser Asp Asn Ile Leu Leu Gly Met Asp
385                 390                 395                 400

Gly Ser Val Lys Leu Thr Asp Phe Gly Phe Cys Ala Gln Ile Thr Pro
                405                 410                 415

Glu Gln Ser Lys Arg Ser Thr Met Val Gly Thr Pro Tyr Trp Met Ala
            420                 425                 430

Pro Glu Val Val Thr Arg Lys Ala Tyr Gly Pro Lys Val Asp Ile Trp
        435                 440                 445

Ser Leu Gly Ile Met Ala Ile Glu Met Ile Glu Gly Glu Pro Pro Tyr
    450                 455                 460

Leu Asn Glu Asn Pro Leu Arg Ala Leu Tyr Leu Ile Ala Thr Asn Gly
465                 470                 475                 480

Thr Pro Glu Leu Gln Asn Pro Glu Lys Leu Ser Ala Ile Phe Arg Asp
                485                 490                 495

Phe Leu Gln Cys Cys Leu Glu Met Asp Val Glu Lys Arg Gly Ser Ala
            500                 505                 510

Lys Glu Leu Leu Gln His Gln Phe Leu Lys Ile Ala Lys Pro Leu Ser
        515                 520                 525

Ser Leu Thr Pro Leu Met His Ala Ala Lys Glu Ala Thr Lys Asn Asn
    530                 535                 540

His
545


<210> SEQ ID NO 13
<211> LENGTH: 2539
<212> TYPE: DNA
<213> ORGANISM: Rattus norvegicus
<220> FEATURE:
<223> OTHER INFORMATION: rat p21 (CDKN1A)-activated kinase 1B (PAK1B)
      cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (389)..(2023)
<223> OTHER INFORMATION: PAK1B
```

-continued

<400> SEQUENCE: 13

```
cccgggcgga aggtaccgcc gcacggtggg cgctcctcac cgcttcccca cacctcccag     60
tcccactgca cctctaggct gcgcggagtc ctctcggcta ttaccggctc tttggagccg    120
cctgtgcccc ctcagagccc gcgctctcca cagtctcctc agagaaggac cccctcagag    180
agcgcacccc gccccttcgc agtccgggcg cgcctcccgc tccccgcggc ccgcagcgtc    240
ccgccctcca gagcccggct cgagcggcga gaggagaacg cggggctgcc gcccgcgcac    300
actcccgcgc tctctctggg tttggctgcc gccaggagga ggaggagcgg agcgagcgca    360
gaagtagctg ctgctgctgg tggtgacaat gtcaaataac ggcttagacg tccaggacaa    420
acccccagcc cctccgatga gaaacaccag cactatgatt ggagctggca gcaaagaccc    480
tggaacccta aaccacggtt ccaaacctct gcctccaaac ccagaggaga agaaaaagaa    540
ggaccggttc tatcgatcca tcttagctgg agataaaaca aataaaaaga aggagaagga    600
gcggccagag atttctcttc cttcagattt tgagcataca attcatgttg gttttgatgc    660
tgtcacaggg gagtttacgg ggatgccaga acagtgggcc cgcttgcttc aaacatcaaa    720
tatcaccaag tcagagcaga agaaaaaccc acaggctgtt ctggatgtgt tggaatttta    780
taactccaag aagacctcca atagtcagaa gtacatgagt tttacagata agtcagctga    840
agattataat tcttctaata ctttgaatgt gaagactgtg tctgagaccc cagcagtgcc    900
accagtgtca gaagatgaag atgatgatga cgatgctacc ccacctccag tgattgctcc    960
acgcccagaa cacacaaaat ctgtatatac acggtctgtg attgaaccac ttcctgttac   1020
tccaactcgg gatgtggcta catctcctat ttctcctact gagaataaca ccactccgcc   1080
agatgctttg acccggaata ctgaaaagca gaagaagaag cctaaaatgt ctgatgagga   1140
gatcttagag aaattacgga gcatagtgag tgtgggcgat cctaagaaga agtacacacg   1200
cttcgagaag attggacaag gtgcttcagg cacagtgtac actgcaatgg atgtagccac   1260
agggcaggag gtggccatta aacagatgaa ccttcagcag cagccgaaga aagaactcat   1320
tattaatgag atcctggtca tgagggaaaa caaaaaccca aacattgtga actatctgga   1380
cagttacctt gtgggagatg agctatgggt tgtcatggaa tacttggccg gaggctcctt   1440
aacagatgtg gtgacagaaa cctgtatgga tgaaggccag atagcagctg tgtgccgaga   1500
gtgtctacaa gctttggagt tcctgcattc aaaccaagtc attcacagag acatcaagag   1560
tgacaatatt ctgctgggaa tggatggctc tgtcaaatta actgactttg gattctgtgc   1620
acagataact ccagagcaga gcaaaaggag caccatggtg ggaactccat attggatggc   1680
acctgaagtt gtgacacgca aggcctatgg acccaaggtt gacatctggt ccctgggtat   1740
tatggcaatt gaaatgattg agggggagcc cccatacctc aatgaaaacc ctttgagagc   1800
cttgtacctc attgctacca atgggacgcc agagcttcag aacccagaga agttgtcagc   1860
tattttccgg gactttttga accgctgtct tgagatggat gtggagaaga gaggttcagc   1920
taaagagcta ctgcagcatc aattcctgaa gattgccaag cctctctcca gcctgactcc   1980
actgattgct gcagcaaaag aggcaaccaa gaacaatcac tgaaaccacg ctcaccccag   2040
cctcatgtgc caagccttct atgaaataaa cactcgtttc gggaactccg acccctcatg   2100
tcctcttctc ctttccttgc ttctcccatt tcctgatcta gtgctctcaa gactttgatc   2160
cttggaaacc gtctagcact gaagagaacc gcaaccggat gactaatcca gcagaggcca   2220
```

-continued

```
tttctaaata ggaattccct ttagcttgtg ggcatggagg ggactgatga gcaagggttt   2280

acctgaataa acctgtttct acgaaacaga aatctcaacc atcccattcc ttacccctca   2340

caatcagttc ttaactctat aaacttatgg tttgatagca ttatcaattt gctatcagtt   2400

gaaattgctt ttgtttttta tttctgtgac caaattgccc aaacacttca ttgtatttga   2460

aaaccagaac agctttgaaa tgccacgggg cctgataatc tgccagggac atgaagaggt   2520

cttgtttccc tgaacccac                                                2539


<210> SEQ ID NO 14
<211> LENGTH: 544
<212> TYPE: PRT
<213> ORGANISM: Rattus norvegicus
<220> FEATURE:
<223> OTHER INFORMATION: rat p21 (CDKN1A)-activated kinase 1B (PAK1B)

<400> SEQUENCE: 14

Met Ser Asn Asn Gly Leu Asp Val Gln Asp Lys Pro Pro Ala Pro Pro
  1               5                  10                  15

Met Arg Asn Thr Ser Thr Met Ile Gly Ala Gly Ser Lys Asp Pro Gly
             20                  25                  30

Thr Leu Asn His Gly Ser Lys Pro Leu Pro Pro Asn Pro Glu Glu Lys
         35                  40                  45

Lys Lys Lys Asp Arg Phe Tyr Arg Ser Ile Leu Ala Gly Asp Lys Thr
     50                  55                  60

Asn Lys Lys Lys Glu Lys Glu Arg Pro Glu Ile Ser Leu Pro Ser Asp
 65                  70                  75                  80

Phe Glu His Thr Ile His Val Gly Phe Asp Ala Val Thr Gly Glu Phe
                 85                  90                  95

Thr Gly Met Pro Glu Gln Trp Ala Arg Leu Leu Gln Thr Ser Asn Ile
            100                 105                 110

Thr Lys Ser Glu Gln Lys Lys Asn Pro Gln Ala Val Leu Asp Val Leu
        115                 120                 125

Glu Phe Tyr Asn Ser Lys Lys Thr Ser Asn Ser Gln Lys Tyr Met Ser
    130                 135                 140

Phe Thr Asp Lys Ser Ala Glu Asp Tyr Asn Ser Ser Asn Thr Leu Asn
145                 150                 155                 160

Val Lys Thr Val Ser Glu Thr Pro Ala Val Pro Pro Val Ser Glu Asp
                165                 170                 175

Glu Asp Asp Asp Asp Asp Ala Thr Pro Pro Pro Val Ile Ala Pro Arg
            180                 185                 190

Pro Glu His Thr Lys Ser Val Tyr Thr Arg Ser Val Ile Glu Pro Leu
        195                 200                 205

Pro Val Thr Pro Thr Arg Asp Val Ala Thr Ser Pro Ile Ser Pro Thr
    210                 215                 220

Glu Asn Asn Thr Thr Pro Pro Asp Ala Leu Thr Arg Asn Thr Glu Lys
225                 230                 235                 240

Gln Lys Lys Lys Pro Lys Met Ser Asp Glu Glu Ile Leu Glu Lys Leu
                245                 250                 255

Arg Ser Ile Val Ser Val Gly Asp Pro Lys Lys Lys Tyr Thr Arg Phe
            260                 265                 270

Glu Lys Ile Gly Gln Gly Ala Ser Gly Thr Val Tyr Thr Ala Met Asp
        275                 280                 285
```

-continued

```
Val Ala Thr Gly Gln Glu Val Ala Ile Lys Gln Met Asn Leu Gln Gln
    290                 295                 300

Gln Pro Lys Lys Glu Leu Ile Ile Asn Glu Ile Leu Val Met Arg Glu
305                 310                 315                 320

Asn Lys Asn Pro Asn Ile Val Asn Tyr Leu Asp Ser Tyr Leu Val Gly
                325                 330                 335

Asp Glu Leu Trp Val Val Met Glu Tyr Leu Ala Gly Gly Ser Leu Thr
            340                 345                 350

Asp Val Val Thr Glu Thr Cys Met Asp Glu Gly Gln Ile Ala Ala Val
        355                 360                 365

Cys Arg Glu Cys Leu Gln Ala Leu Glu Phe Leu His Ser Asn Gln Val
    370                 375                 380

Ile His Arg Asp Ile Lys Ser Asp Asn Ile Leu Leu Gly Met Asp Gly
385                 390                 395                 400

Ser Val Lys Leu Thr Asp Phe Gly Phe Cys Ala Gln Ile Thr Pro Glu
                405                 410                 415

Gln Ser Lys Arg Ser Thr Met Val Gly Thr Pro Tyr Trp Met Ala Pro
            420                 425                 430

Glu Val Val Thr Arg Lys Ala Tyr Gly Pro Lys Val Asp Ile Trp Ser
        435                 440                 445

Leu Gly Ile Met Ala Ile Glu Met Ile Glu Gly Glu Pro Pro Tyr Leu
    450                 455                 460

Asn Glu Asn Pro Leu Arg Ala Leu Tyr Leu Ile Ala Thr Asn Gly Thr
465                 470                 475                 480

Pro Glu Leu Gln Asn Pro Glu Lys Leu Ser Ala Ile Phe Arg Asp Phe
                485                 490                 495

Leu Asn Arg Cys Leu Glu Met Asp Val Glu Lys Arg Gly Ser Ala Lys
            500                 505                 510

Glu Leu Leu Gln His Gln Phe Leu Lys Ile Ala Lys Pro Leu Ser Ser
        515                 520                 525

Leu Thr Pro Leu Ile Ala Ala Ala Lys Glu Ala Thr Lys Asn Asn His
    530                 535                 540


<210> SEQ ID NO 15
<211> LENGTH: 1662
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human SPUVE serine protease 23 cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (121)..(1272)
<223> OTHER INFORMATION: SPUVE

<400> SEQUENCE: 15

gacccacgcg tccgggggca ggcatgggag ccgcgcgctc tctcccggcg cccacacctg      60

tctgagcggc gcagcgagcc gcggcccggg cgggctgctc ggcgcggaac agtgctcggc     120

atggcaggga ttccagggct cctcttcctt ctcttctttc tgctctgtgc tgttgggcaa     180

gtgagccctt acagtgcccc ctggaaaccc acttggcctg cataccgcct ccctgtcgtc     240

ttgccccagt ctaccctcaa tttagccaag ccagactttg gagccgaagc caaattagaa     300

gtatcttctt catgtggacc ccagtgtcat aagggaactc cactgcccac ttacgaagag     360

gccaagcaat atctgtctta tgaaacgctc tatgccaatg gcagccgcac agagacgcag     420
```

-continued

```
gtgggcatct acatcctcag cagtagtgga gatggggccc aacaccgaga ctcagggtct    480

tcaggaaagt ctcgaaggaa gcggcagatt tatggctatg acagcaggtt cagcattttt    540

gggaaggact tcctgctcaa ctaccctttc tcaacatcag tgaagttatc cacgggctgc    600

accggcaccc tggtggcaga gaagcatgtc ctcacagctg cccactgcat acacgatgga    660

aaaacctatg tgaaaggaac ccagaagctt cgagtgggct tcctaaagcc caagtttaaa    720

gatggtggtc gaggggccaa cgactccact tcagccatgc ccgagcagat gaaatttcag    780

tggatccggg tgaaacgcac ccatgtgccc aagggttgga tcaagggcaa tgccaatgac    840

atcggcatgg attatgatta tgccctcctg gaactcaaaa agccccacaa gagaaaattt    900

atgaagattg gggtgagccc tcctgctaag cagctgccag gggcagaat tcacttctct    960

ggttatgaca atgaccgacc aggcaatttg gtgtatcgct tctgtgacgt caaagacgag   1020

acctatgact tgctctacca gcaatgcgat gcccagccag gggccagcgg gtctggggtc   1080

tatgtgagga tgtggaagag acagcagcag aagtgggagc gaaaaattat tggcattttt   1140

tcagggcacc agtgggtgga catgaatggt tccccacagg atttcaacgt ggctgtcaga   1200

atcactcctc tcaaatatgc ccagatttgc tattggatta aaggaaacta cctggattgt   1260

agggaggggt gacacagtgt tccctcctgg cagcaattaa gggtcttcat gttcttattt   1320

taggagaggc caaattgttt tttgtcattg gcgtgcacac gtgtgtgtgt gtgtgtgtgt   1380

gtgtgtaagg tgtcttataa tcttttacct atttcttaca attgcaagat gactggcttt   1440

actatttgaa aactggtttg tgtatcatat catatatcat ttaagcagtt tgaaggcata   1500

cttttgcata gaaataaaaa aaatactgat ttggggcaat gaggaatatt tgacaattaa   1560

gttaatcttc acgtttttgc aaactttgat ttttatttca tctgaacttg tttcaaagat   1620

ttatattaaa tatttggcat acaagagaaa aaaaaaaaaa aa                      1662
```

```
<210> SEQ ID NO 16
<211> LENGTH: 383
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human SPUVE serine protease 23

<400> SEQUENCE: 16

Met Ala Gly Ile Pro Gly Leu Leu Phe Leu Leu Phe Phe Leu Leu Cys
  1               5                  10                  15

Ala Val Gly Gln Val Ser Pro Tyr Ser Ala Pro Trp Lys Pro Thr Trp
             20                  25                  30

Pro Ala Tyr Arg Leu Pro Val Val Leu Pro Gln Ser Thr Leu Asn Leu
         35                  40                  45

Ala Lys Pro Asp Phe Gly Ala Glu Ala Lys Leu Glu Val Ser Ser Ser
     50                  55                  60

Cys Gly Pro Gln Cys His Lys Gly Thr Pro Leu Pro Thr Tyr Glu Glu
 65                  70                  75                  80

Ala Lys Gln Tyr Leu Ser Tyr Glu Thr Leu Tyr Ala Asn Gly Ser Arg
                 85                  90                  95

Thr Glu Thr Gln Val Gly Ile Tyr Ile Leu Ser Ser Ser Gly Asp Gly
            100                 105                 110

Ala Gln His Arg Asp Ser Gly Ser Ser Gly Lys Ser Arg Arg Lys Arg
        115                 120                 125
```

-continued

```
Gln Ile Tyr Gly Tyr Asp Ser Arg Phe Ser Ile Phe Gly Lys Asp Phe
    130                 135                 140

Leu Leu Asn Tyr Pro Phe Ser Thr Ser Val Lys Leu Ser Thr Gly Cys
145                 150                 155                 160

Thr Gly Thr Leu Val Ala Glu Lys His Val Leu Thr Ala Ala His Cys
                165                 170                 175

Ile His Asp Gly Lys Thr Tyr Val Lys Gly Thr Gln Lys Leu Arg Val
            180                 185                 190

Gly Phe Leu Lys Pro Lys Phe Lys Asp Gly Gly Arg Gly Ala Asn Asp
        195                 200                 205

Ser Thr Ser Ala Met Pro Glu Gln Met Lys Phe Gln Trp Ile Arg Val
    210                 215                 220

Lys Arg Thr His Val Pro Lys Gly Trp Ile Lys Gly Asn Ala Asn Asp
225                 230                 235                 240

Ile Gly Met Asp Tyr Asp Tyr Ala Leu Leu Glu Leu Lys Lys Pro His
                245                 250                 255

Lys Arg Lys Phe Met Lys Ile Gly Val Ser Pro Pro Ala Lys Gln Leu
            260                 265                 270

Pro Gly Gly Arg Ile His Phe Ser Gly Tyr Asp Asn Asp Arg Pro Gly
        275                 280                 285

Asn Leu Val Tyr Arg Phe Cys Asp Val Lys Asp Glu Thr Tyr Asp Leu
    290                 295                 300

Leu Tyr Gln Gln Cys Asp Ala Gln Pro Gly Ala Ser Gly Ser Gly Val
305                 310                 315                 320

Tyr Val Arg Met Trp Lys Arg Gln Gln Gln Lys Trp Glu Arg Lys Ile
                325                 330                 335

Ile Gly Ile Phe Ser Gly His Gln Trp Val Asp Met Asn Gly Ser Pro
            340                 345                 350

Gln Asp Phe Asn Val Ala Val Arg Ile Thr Pro Leu Lys Tyr Ala Gln
        355                 360                 365

Ile Cys Tyr Trp Ile Lys Gly Asn Tyr Leu Asp Cys Arg Glu Gly
    370                 375                 380


<210> SEQ ID NO 17
<211> LENGTH: 1936
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse SPUVE serine protease 23 cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (170)..(1318)
<223> OTHER INFORMATION: SPUVE

<400> SEQUENCE: 17

gtgctggtgt cagtactaga ctcttctggc agggtgacct tcaatgactc tgccttatga     60

gaagattctc agaactgaca ccctcccgga gcccagggac agagctgcag gagaggcctg    120

cctgcctgcc tgcctgcctg catgcctgcc ttcctatgta gcactcagca tggctggaat    180

cccggggctc ttcatccttc ttgtcctgct ctgtgtgttc atgcaggtga gtccctacac    240

cgttccgtgg aaacccacat ggccggctta tcgcctccct gtagtcttgc ctcagtctac    300

cctcaactta gctaaggcag acttcgacgc caaagcgaaa ttggaggtgt cctcctcatg    360

tggacctcag tgtcacaagg gaacaccact gcccacctac gaagaggcca agcagtacct    420
```

-continued

```
ttcctatgaa accctttatg ccaatggcag ccgcacagag actcgggtgg gcatctacat    480

cctcagcaat ggtgaaggca gggcacgagg cagagactcg gaggccacag ggagatctcg    540

caggaagagg cagatttatg gctacgatgg caggtttagc atttttggga aggacttcct    600

gctcaattat cctttctcaa catcggtgaa gttgtctact ggctgcactg gcaccctggt    660

ggcagagaag cacgtcctca ctgctgccca ctgcatacac gatgggaaaa cctatgtgaa    720

agggacacag aaactccgag tgggcttcct gaagcccaag tataaagatg gtgccggagg    780

ggacaacagc tcgagctcag ccatgccaga caagatgaag tttcagtgga tccgcgtgaa    840

acgcacccat gtgcccaagg ggtggatcaa gggcaatgcc aatgacatcg gcatggatta    900

tgactacgcc ctgctggaac tcaagaaacc ccacaaaaga cagttcatga agattggtgt    960

gagtcctcca gcgaagcagc tcccaggggg caggatccac ttctctggtt atgacaatga   1020

ccggcccggc aatttggtgt accgcttctg tgatgtcaaa gatgagacct acgaccttct   1080

ctaccagcag tgtgacgccc agcccggggc cagtggttca ggggtctatg tgaggatgtg   1140

gaagagacca cagcagaaat gggaaagaaa aattatcggc atcttttcag ggcaccagtg   1200

ggtggacatg aatggctctc cacaggattt caacgtggca gttagaatca cgcctcttaa   1260

atatgcccag atttgctatt ggattaaagg aaactaccta gattgcaggg aggggtgaca   1320

tgcgtcttct tgccagcacc aatggtcttt ttgcactcat tgtaggagag gctagctttt   1380

tatcattgac tcttgtggtg tgagtcacat agtatctttt acctagtatt cttcaaatgg   1440

caaaaattat tggctatatt attttaaaac tgttgtgtgc gttatagcat ttaagcagtc   1500

tgaaagcata cttttgcata gagactttaa agtattcggg taatagggcc tatttgacaa   1560

ggaagttaaa ctttcagttt ttggagaatt ctaatttttg tctgatccaa acttgcttca   1620

gaggtttata tcaaatacgt gacacacagg gaatatgaat tcttatgttt gtatatgtat   1680

atgttttctt ctgagagtca tatattgata tttttgtaat gtgtggttat tatgcttcca   1740

gataatgata gcaaagtctt caataggcaa tttataatgt tttggattca aacatttacg   1800

tagtagtcct tgaagagaac aataatttat tggctatatt gatacccata taagactgta   1860

tcttacagtg cacagaattc ccacgctgct tttagttttg aaaataaaac tttcccttgt   1920

aaaaaaaaaa aaaaaa                                                    1936
```

<210> SEQ ID NO 18
<211> LENGTH: 382
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse SPUVE serine protease 23

<400> SEQUENCE: 18

```
Met Ala Gly Ile Pro Gly Leu Phe Ile Leu Leu Val Leu Leu Cys Val
  1               5                  10                  15

Phe Met Gln Val Ser Pro Tyr Thr Val Pro Trp Lys Pro Thr Trp Pro
             20                  25                  30

Ala Tyr Arg Leu Pro Val Val Leu Pro Gln Ser Thr Leu Asn Leu Ala
         35                  40                  45

Lys Ala Asp Phe Asp Ala Lys Ala Lys Leu Glu Val Ser Ser Ser Cys
     50                  55                  60

Gly Pro Gln Cys His Lys Gly Thr Pro Leu Pro Thr Tyr Glu Glu Ala
 65                  70                  75                  80
```

-continued

```
Lys Gln Tyr Leu Ser Tyr Glu Thr Leu Tyr Ala Asn Gly Ser Arg Thr
                85                  90                  95

Glu Thr Arg Val Gly Ile Tyr Ile Leu Ser Asn Gly Glu Gly Arg Ala
            100                 105                 110

Arg Gly Arg Asp Ser Glu Ala Thr Gly Arg Ser Arg Arg Lys Arg Gln
        115                 120                 125

Ile Tyr Gly Tyr Asp Gly Arg Phe Ser Ile Phe Gly Lys Asp Phe Leu
    130                 135                 140

Leu Asn Tyr Pro Phe Ser Thr Ser Val Lys Leu Ser Thr Gly Cys Thr
145                 150                 155                 160

Gly Thr Leu Val Ala Glu Lys His Val Leu Thr Ala Ala His Cys Ile
                165                 170                 175

His Asp Gly Lys Thr Tyr Val Lys Gly Thr Gln Lys Leu Arg Val Gly
            180                 185                 190

Phe Leu Lys Pro Lys Tyr Lys Asp Gly Ala Gly Gly Asp Asn Ser Ser
        195                 200                 205

Ser Ser Ala Met Pro Asp Lys Met Lys Phe Gln Trp Ile Arg Val Lys
    210                 215                 220

Arg Thr His Val Pro Lys Gly Trp Ile Lys Gly Asn Ala Asn Asp Ile
225                 230                 235                 240

Gly Met Asp Tyr Asp Tyr Ala Leu Leu Glu Leu Lys Lys Pro His Lys
                245                 250                 255

Arg Gln Phe Met Lys Ile Gly Val Ser Pro Pro Ala Lys Gln Leu Pro
            260                 265                 270

Gly Gly Arg Ile His Phe Ser Gly Tyr Asp Asn Asp Arg Pro Gly Asn
        275                 280                 285

Leu Val Tyr Arg Phe Cys Asp Val Lys Asp Glu Thr Tyr Asp Leu Leu
    290                 295                 300

Tyr Gln Gln Cys Asp Ala Gln Pro Gly Ala Ser Gly Ser Gly Val Tyr
305                 310                 315                 320

Val Arg Met Trp Lys Arg Pro Gln Gln Lys Trp Glu Arg Lys Ile Ile
                325                 330                 335

Gly Ile Phe Ser Gly His Gln Trp Val Asp Met Asn Gly Ser Pro Gln
            340                 345                 350

Asp Phe Asn Val Ala Val Arg Ile Thr Pro Leu Lys Tyr Ala Gln Ile
        355                 360                 365

Cys Tyr Trp Ile Lys Gly Asn Tyr Leu Asp Cys Arg Glu Gly
    370                 375                 380

<210> SEQ ID NO 19
<211> LENGTH: 827
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human similar to natural killer cell transcript
      4 (NK4) cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (59)..(625)
<223> OTHER INFORMATION: NK4

<400> SEQUENCE: 19

gactgtctca gtggagctgg gtcatctcag gccttggctc cttgaacttt tggccgccat      60

gtgcttcccg aaggtcctct ctgatgacat gaagaagctg aaggcccgaa tgcaccaggc     120
```

-continued

```
catagaaaga ttttatgata aaatgcaaaa tgcagaatca ggacgtggac aggtgatgtc    180

gagcctggca gagctggagg acgacttcaa agagggctac ctggagacag tggcggctta    240

ttatgaggag cagcacccag agctcactcc tctacttgaa aaagaaagag atggattacg    300

gtgccgaggc aacagatccc ctgtcccgga tgttgaggat cccgcaaccg aggagcctgg    360

ggagagcttt tgtgacaagg tcatgagatg gttccaggcc atgctgcagc ggctgcagac    420

ctggtggcac ggggttctgg cctgggtgaa ggagaaggtg gtggccctgg tccatgcagt    480

gcaggccctc tggaaacagt tccagagttt ctgctgctct ctgtcagagc tcttcatgtc    540

ctctttccag tcctacggag ccccacgggg ggacaaggag gagctgacac cccagaagtg    600

ctctgaaccc caatcctcaa aatgaagata ctgacaccac ctttgccctc cccgtcaccg    660

cgcacccacc ctgacccctc cctcagctgt cctgtgcccc gccctctccc gcacactcag    720

tccccctgcc tggcgttcct gccgcagctc tgacctggtg ctgtcgccct ggcatcttaa    780

taaaacctgc ttatacttcc ctggaaaaaa aaaaaaaaaa aaaaaaa                  827
```

<210> SEQ ID NO 20
<211> LENGTH: 188
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human similar to natural killer cell transcript
      4 (NK4)

<400> SEQUENCE: 20

```
Met Cys Phe Pro Lys Val Leu Ser Asp Asp Met Lys Lys Leu Lys Ala
  1               5                  10                  15

Arg Met His Gln Ala Ile Glu Arg Phe Tyr Asp Lys Met Gln Asn Ala
             20                  25                  30

Glu Ser Gly Arg Gly Gln Val Met Ser Ser Leu Ala Glu Leu Glu Asp
         35                  40                  45

Asp Phe Lys Glu Gly Tyr Leu Glu Thr Val Ala Ala Tyr Tyr Glu Glu
     50                  55                  60

Gln His Pro Glu Leu Thr Pro Leu Leu Glu Lys Glu Arg Asp Gly Leu
 65                  70                  75                  80

Arg Cys Arg Gly Asn Arg Ser Pro Val Pro Asp Val Glu Asp Pro Ala
                 85                  90                  95

Thr Glu Glu Pro Gly Glu Ser Phe Cys Asp Lys Val Met Arg Trp Phe
            100                 105                 110

Gln Ala Met Leu Gln Arg Leu Gln Thr Trp Trp His Gly Val Leu Ala
        115                 120                 125

Trp Val Lys Glu Lys Val Val Ala Leu Val His Ala Val Gln Ala Leu
    130                 135                 140

Trp Lys Gln Phe Gln Ser Phe Cys Cys Ser Leu Ser Glu Leu Phe Met
145                 150                 155                 160

Ser Ser Phe Gln Ser Tyr Gly Ala Pro Arg Gly Asp Lys Glu Glu Leu
                165                 170                 175

Thr Pro Gln Lys Cys Ser Glu Pro Gln Ser Ser Lys
            180                 185
```

<210> SEQ ID NO 21
<211> LENGTH: 2254
<212> TYPE: DNA

-continued

<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human Protein C inhibitor (PCI) cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (140)..(1360)
<223> OTHER INFORMATION: PCI

<400> SEQUENCE: 21

ggcacgaggc tcagctacgg gacacatttc aggtatccaa ggcagcagag gtgagtgggt 60

cccccgagct ctgtgacctt atgctccaca ctaactctgg cagagcctcc gtttcctcat 120

agaacaaaga acagccacca tgcagctctt cctcctcttg tgcctggtgc ttctcagccc 180

tcaggggggcc tcccttcacc gccaccaccc ccgggagatg aagaagagag tcgaggacct 240

ccatgtaggt gccacggtgg cccccagcag cagaagggac tttacctttg acctctacag 300

ggccttggct tccgctgccc ccagccagaa catcttcttc tcccctgtga gcatctccat 360

gagcctggcc atgctctccc tgggggctgg gtccagcaca aagatgcaga tcctggaggg 420

cctgggcctc aacctccaga aaagctcaga gaaggagctg cacagaggct ttcagcagct 480

ccttcaggaa ctcaaccagc ccagagatgg cttccagctg agcctcggca atgccctttt 540

caccgacctg gtggtagacc tgcaggacac cttcgtaagt gccatgaaga cgctgtacct 600

ggcagacact ttccccacca actttaggga ctctgcaggg gccatgaagc agatcaatga 660

ttatgtggca aagcaaacga agggcaagat tgtggacttg cttaagaacc tcgatagcaa 720

tgcggtcgtg atcatggtga attacatctt ctttaaagct aagtgggaga caagcttcaa 780

ccacaaaggc acccaagagc aagacttcta cgtgacctcg gagactgtgg tgcgggtacc 840

catgatgagc cgcgaggatc agtatcacta cctcctggac cggaacctct cctgcagggt 900

ggtgggggtc ccctaccaag gcaatgccac ggctttgttc attctcccca gtgagggaaa 960

gatgcagcag gtggagaatg gactgagtga gaaaacgctg aggaagtggc ttaagatgtt 1020

caaaaagagg cagctcgagc tttaccttcc caaattctcc attgagggct cctatcagct 1080

ggagaaagtc ctccccagtc tggggatcag taacgtcttc acctcccatg ctgatctgtc 1140

cggcatcagc aaccactcaa atatccaggt gtctgagatg gtgcacaaag ctgtggtgga 1200

ggtggacgag tcgggaacca gagcagcggc agccacgggg acaatcttca ctttcaggtc 1260

ggcccgcctg aactctcaga ggctagtgtt caacaggccc tttctgatgt tcattgtgga 1320

taacaacatc ctcttccttg gcaaagtgaa ccgcccctga ggtggggctt ctcctgaaat 1380

ctacaggcct cagggtggga gatgaagggg gctatgctat ggcccatctg tatgctggta 1440

gctagtgatt tacacaggtt tagttgacta atgaggcatt acaaataata ttactctatg 1500

atgattgctt ccacccacac gactgcaaca tacaggtgcc ttggggaaat gtggagaaca 1560

ttcaatcttg ccgtcactat tcatcaatga agattagcac tgagatccag agaggctgga 1620

tgacttgctc aagttcacca gcatggtagt ggcaaagaga ggtccagagt cctggccctt 1680

gatgcccagc tcagtgccac aaagctcagt aggagggatg ttccagtgga tgagggccac 1740

caggaagcac aggtccaagg ctggtcccac acttatcagc agcaacaact gtcagttcat 1800

cctgcatggg aaaaatgttg gaatgggagt ctgaaatggg gctactgttt cagtcctaac 1860

gtgctgtgtg acattgggac aacactttcc ctctctggac ctcagtttcc ctctgtatac 1920

aaggatcaga ttcttgctgt gacccaagaa ctcctgaaat catatagaaa ggctggggtg 1980

-continued

```
ggccctgtca ttcgtggttg atttcaatac actcaagtgc cattcatcct ttaagaaaaa   2040

catctggata tcaaggtgga aatggcccat ttaatgattg attatatcat tttgtggata   2100

tagttataat ctgatgggcc tggctgggag tggaagaagg gaagcctttt gcaaatagta   2160

gagtgtcagt tgcaggtgcc aatgactaac tttttgaatt ctatgttggc attaacaata   2220

aagcattttg caaacaaaaa aaaaaaaaaa aaaa                               2254


<210> SEQ ID NO 22
<211> LENGTH: 406
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human Protein C inhibitor (PCI)

<400> SEQUENCE: 22

Met Gln Leu Phe Leu Leu Leu Cys Leu Val Leu Leu Ser Pro Gln Gly
1               5                   10                  15

Ala Ser Leu His Arg His His Pro Arg Glu Met Lys Lys Arg Val Glu
            20                  25                  30

Asp Leu His Val Gly Ala Thr Val Ala Pro Ser Ser Arg Arg Asp Phe
        35                  40                  45

Thr Phe Asp Leu Tyr Arg Ala Leu Ala Ser Ala Ala Pro Ser Gln Asn
    50                  55                  60

Ile Phe Phe Ser Pro Val Ser Ile Ser Met Ser Leu Ala Met Leu Ser
65                  70                  75                  80

Leu Gly Ala Gly Ser Ser Thr Lys Met Gln Ile Leu Glu Gly Leu Gly
                85                  90                  95

Leu Asn Leu Gln Lys Ser Ser Glu Lys Glu Leu His Arg Gly Phe Gln
            100                 105                 110

Gln Leu Leu Gln Glu Leu Asn Gln Pro Arg Asp Gly Phe Gln Leu Ser
        115                 120                 125

Leu Gly Asn Ala Leu Phe Thr Asp Leu Val Val Asp Leu Gln Asp Thr
    130                 135                 140

Phe Val Ser Ala Met Lys Thr Leu Tyr Leu Ala Asp Thr Phe Pro Thr
145                 150                 155                 160

Asn Phe Arg Asp Ser Ala Gly Ala Met Lys Gln Ile Asn Asp Tyr Val
                165                 170                 175

Ala Lys Gln Thr Lys Gly Lys Ile Val Asp Leu Leu Lys Asn Leu Asp
            180                 185                 190

Ser Asn Ala Val Val Ile Met Val Asn Tyr Ile Phe Phe Lys Ala Lys
        195                 200                 205

Trp Glu Thr Ser Phe Asn His Lys Gly Thr Gln Glu Gln Asp Phe Tyr
    210                 215                 220

Val Thr Ser Glu Thr Val Val Arg Val Pro Met Met Ser Arg Glu Asp
225                 230                 235                 240

Gln Tyr His Tyr Leu Leu Asp Arg Asn Leu Ser Cys Arg Val Val Gly
                245                 250                 255

Val Pro Tyr Gln Gly Asn Ala Thr Ala Leu Phe Ile Leu Pro Ser Glu
            260                 265                 270

Gly Lys Met Gln Gln Val Glu Asn Gly Leu Ser Glu Lys Thr Leu Arg
        275                 280                 285

Lys Trp Leu Lys Met Phe Lys Lys Arg Gln Leu Glu Leu Tyr Leu Pro
    290                 295                 300
```

-continued

```
Lys Phe Ser Ile Glu Gly Ser Tyr Gln Leu Glu Lys Val Leu Pro Ser
305                 310                 315                 320

Leu Gly Ile Ser Asn Val Phe Thr Ser His Ala Asp Leu Ser Gly Ile
                325                 330                 335

Ser Asn His Ser Asn Ile Gln Val Ser Glu Met Val His Lys Ala Val
            340                 345                 350

Val Glu Val Asp Glu Ser Gly Thr Arg Ala Ala Ala Ala Thr Gly Thr
        355                 360                 365

Ile Phe Thr Phe Arg Ser Ala Arg Leu Asn Ser Gln Arg Leu Val Phe
    370                 375                 380

Asn Arg Pro Phe Leu Met Phe Ile Val Asp Asn Asn Ile Leu Phe Leu
385                 390                 395                 400

Gly Lys Val Asn Arg Pro
                405


<210> SEQ ID NO 23
<211> LENGTH: 2125
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse Protein C inhibitor (PCI), serine (or
      cysteine) proteinase inhibitor, clade A, member 5
      (Serpina5) cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (125)..(1342)
<223> OTHER INFORMATION: PCI

<400> SEQUENCE: 23

cagagtctct gtggtcagtg acatttgccg gctgcaattc ctctgctctg ctctctgtcc     60

tcttgtaagc caggtggact ataccttgac actcagccac cacgatacag aagctgtcac    120

cacaatgagg ttcttcccca ttctgtgcct ggtgctgttc atcagccatg ggtggcttc     180

ccgccgacac tcccattcca agaagaagaa ggctaaagag tcctcggtgg gtgctgtggg    240

acctcccagt agcaaagact ttgctttcag actctacagg gccttggctt ctgaatcccc    300

tggtcagaat gtcttcttct ccccccttgag cgtgtctatg agtttgggca tgctctccct    360

gggggctggc ttgaagacga agacccagat cctagatggc ctaggcctca gcctccagca    420

aggccaagaa gacaagctcc acaagggctt ccaacagctg ctacagagat tcaggcagcc    480

tagtgatggc ctgcagctga gcctgggcag tgcccttttt aaagacccag cagtacatat    540

ccgggacgac ttcctgagtg ccatgaagac actgtacatg tcagacactt tctctaccaa    600

ctttgggaac cctgaaattg ccaagaagca gatcaacaac tatgtagcca agcagaccaa    660

gggcaagatt gtagacttga tcaaggacct tgacagcacc catgtcatga tagtggtaaa    720

ttacatcttc ttcaaagcca agtggcagac ggccttcagt gagacaaaca cccacaagat    780

ggatttccat gtgaccccca aaaagaccat acgggtgccc atgatgaacc gtgaagatga    840

gtattcctac tacctggacc aaaacatctc ctgcacggtg gtggggatcc cttatcaagg    900

caatgccatt gctctattta ttctccccag cgagggcaag atgaagcagg tggaggatgg    960

cctggacgag agaacattga ggaactggct taagatgttc accaagagac gcttagatct   1020

ttacctcccc aagttctcca ttgaggctac ctacaaactg gaaaatgtcc tccccaagct   1080

gggcatccag gacgtcttca ccacccatgc tgacttgtct ggcattactg accataccaa   1140
```

-continued

```
tatcaagttg tctgagatgg tgcacaaatc catgatggag gttgaagagt caggaaccac   1200

agcagctgcc atcacaggag ccatcttcac attcagatct gctcggccga gctccctgaa   1260

gatagaattc accagaccct ttctgctgac ccttatggag gattcacata tacttttcgt   1320

tggcaaggtg acccggccct gaggtgggac tccttctgaa actcatgggc ttacacagag   1380

ggagccaggc atgttcttca gcccatccat ctcttggttg ctagtgattt tcatagagta   1440

ggttgactag taaggtgttg taagtgataa gatcaatatc cctatggctg cttccttgat   1500

cacagttgca agatgtgttc ttgctctcat ccttcatcat tgacactgac cctaaagagg   1560

tcacttgatt tctccaaggt cctaccatct tatatgtgac atagtcgtgc caagactcca   1620

ggcctctcct gctcagtcta gctctacaaa gttccctaag agactggtcc gcatagataa   1680

aggtcaccag gaagcacagt cataaacctg gtcccccacc tatcagcacc aactctcagt   1740

tcatcagggt tgagaaagga cactggattg agcttctgaa atgccgctgg tagctctgta   1800

acatgggaat ggaaatgtcc ctctctgggc ctgtgtttct ttgtgtgtcc agtgaaggaa   1860

cagaactaga agcccattca gcaacctgtg agattcttac acaggggttg gcttctctat   1920

gttagccagg gttgctgtaa gaactcaggt tagtagcata cttgagtaat cagggtcaaa   1980

gtgtctcacc acaaacaaag gtatcaaatc acaagcctgg ctggctgagg tggtgtggta   2040

gaggtctgtg gtaaacagca gaacttaggg tgcagactga cttattgatt tccgccttgg   2100

ctggcaataa aggcattttg caaag                                         2125
```

```
<210> SEQ ID NO 24
<211> LENGTH: 405
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse Protein C inhibitor (PCI), serine (or
      cysteine) proteinase inhibitor, clade A, member 5
      (Serpina5)

<400> SEQUENCE: 24

Met Arg Phe Phe Pro Ile Leu Cys Leu Val Leu Phe Ile Ser His Gly
  1               5                  10                  15

Val Ala Ser Arg Arg His Ser His Ser Lys Lys Lys Lys Ala Lys Glu
             20                  25                  30

Ser Ser Val Gly Ala Val Gly Pro Pro Ser Ser Lys Asp Phe Ala Phe
         35                  40                  45

Arg Leu Tyr Arg Ala Leu Ala Ser Glu Ser Pro Gly Gln Asn Val Phe
     50                  55                  60

Phe Ser Pro Leu Ser Val Ser Met Ser Leu Gly Met Leu Ser Leu Gly
 65                  70                  75                  80

Ala Gly Leu Lys Thr Lys Thr Gln Ile Leu Asp Gly Leu Gly Leu Ser
                 85                  90                  95

Leu Gln Gln Gly Gln Glu Asp Lys Leu His Lys Gly Phe Gln Gln Leu
            100                 105                 110

Leu Gln Arg Phe Arg Gln Pro Ser Asp Gly Leu Gln Leu Ser Leu Gly
        115                 120                 125

Ser Ala Leu Phe Lys Asp Pro Ala Val His Ile Arg Asp Asp Phe Leu
    130                 135                 140

Ser Ala Met Lys Thr Leu Tyr Met Ser Asp Thr Phe Ser Thr Asn Phe
145                 150                 155                 160
```

-continued

```
Gly Asn Pro Glu Ile Ala Lys Lys Gln Ile Asn Asn Tyr Val Ala Lys
                165                 170                 175

Gln Thr Lys Gly Lys Ile Val Asp Leu Ile Lys Asp Leu Asp Ser Thr
            180                 185                 190

His Val Met Ile Val Val Asn Tyr Ile Phe Phe Lys Ala Lys Trp Gln
        195                 200                 205

Thr Ala Phe Ser Glu Thr Asn Thr His Lys Met Asp Phe His Val Thr
    210                 215                 220

Pro Lys Lys Thr Ile Arg Val Pro Met Met Asn Arg Glu Asp Glu Tyr
225                 230                 235                 240

Ser Tyr Tyr Leu Asp Gln Asn Ile Ser Cys Thr Val Val Gly Ile Pro
                245                 250                 255

Tyr Gln Gly Asn Ala Ile Ala Leu Phe Ile Leu Pro Ser Glu Gly Lys
            260                 265                 270

Met Lys Gln Val Glu Asp Gly Leu Asp Glu Arg Thr Leu Arg Asn Trp
        275                 280                 285

Leu Lys Met Phe Thr Lys Arg Arg Leu Asp Leu Tyr Leu Pro Lys Phe
    290                 295                 300

Ser Ile Glu Ala Thr Tyr Lys Leu Glu Asn Val Leu Pro Lys Leu Gly
305                 310                 315                 320

Ile Gln Asp Val Phe Thr Thr His Ala Asp Leu Ser Gly Ile Thr Asp
                325                 330                 335

His Thr Asn Ile Lys Leu Ser Glu Met Val His Lys Ser Met Met Glu
            340                 345                 350

Val Glu Glu Ser Gly Thr Thr Ala Ala Ala Ile Thr Gly Ala Ile Phe
        355                 360                 365

Thr Phe Arg Ser Ala Arg Pro Ser Ser Leu Lys Ile Glu Phe Thr Arg
    370                 375                 380

Pro Phe Leu Leu Thr Leu Met Glu Asp Ser His Ile Leu Phe Val Gly
385                 390                 395                 400

Lys Val Thr Arg Pro
                405


<210> SEQ ID NO 25
<211> LENGTH: 2035
<212> TYPE: DNA
<213> ORGANISM: Rattus norvegicus
<220> FEATURE:
<223> OTHER INFORMATION: rat Protein C inhibitor (PCI), serine (or
      cysteine) proteinase inhibitor, clade A, member 5
      (Serpina5) cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (48)..(1268)
<223> OTHER INFORMATION: PCI

<400> SEQUENCE: 25

tacccgtcaa gacactcagc caccacaata cagaaatggc agccacaatg aggttcttcc      60

ccattctgtg cctggtgctg ttcttcagtc acggggtggc ttcccgccaa cgctcccatt     120

ctaaggagaa gaagaagtct aaagagtcct ctgtgggtgc tgtggggact tcccgaagca     180

gagactttgc cttcaggctc tacagggcct tggcttctga agcccctggt cagaatgttt     240

tcttctcccc catgagcgtg tctatgagct tgggtatgct ctccctgggg tctggcttga     300

agactaaggc acagatcctg gagggcctgg gcctcagcct ccagcaaggc caagaggata     360
```

-continued

```
tgctccacaa gggcttccaa cagctgctgc agcagttcag ccagcctagt gatggcctcc    420

agctgagcct aggcagtgcc ctttttacag acccagcagt acatatccgg gaccactttc    480

tgagtgccat gaagacattg tacatgtcag acatgttctc taccaacttt gggaaccctg    540

aaagtgctaa gaagcagatc aatgactatg tagccaagaa gaccaacggc aagattgtag    600

acttgatcaa ggatcttgac agcacccacg tcatggtggt ggtaaattat atcttcttca    660

aagccaagtg gcagacggcc ttcagtagca ccaacaccca caagatggat ttccatgtga    720

ctcccaagaa gaccattcag gtgccaatga tgaaccgcga agatatatat tcctacatcc    780

tcgaccaaaa catttcctgc acggtggtgg ggatccctta ccaaggcaat acatttgctc    840

tgtttattct tcccagcgag ggcaagatga agcgggtaga ggatggcctg gatgagagaa    900

cgttgaggaa ctggcttaag atgttcacaa agagacagtt agatctttac ctccccaagt    960

tctccattga gggtacctat aaactggaaa aatcctcccc caagctgggc atccaggaca   1020

tcttcaccac ccatgctgac ttgtctggcc ttactgacca caccaatatt aagttgtctg   1080

agatggtaca caaatccatg gtggaggtag acgagtcagg aacaacagca gctgcctcca   1140

caggaattct cttcacactc agatctgctc gaccgagctc tctgaaggta gaattcacca   1200

gaccctttct ggtggtcatt atggatggta caaacctcta tttcattggc aaagtgattc   1260

agccctgagg tgggactcct tctgaaactc acaggcctat acagaaggag ccaggcatat   1320

tctacagact atccacctct tggtagctag tgattttcat agagtagatt gactaataag   1380

gtgttgtaag tgagaagatc aatatcctgg tggaagcttc cttgatcaca gttgcaagat   1440

gtgctcttgc tctatttttt cattattgac actgatgcta aagaggccac ttgattgctc   1500

caaggtccta tcaccatatt tgtgacatag ttatgcccag acctcagacc tctcttgctc   1560

agtctagcgc tacaaagctc tgcaagaggc tggtcagcat agataaaggc caccaggaag   1620

cacggttata aacctggtct cacacctatt agcaccatgg ttgagaaagg acattggact   1680

gagattctga aatggtgcta atagctatgt aactcgaaaa tgaaaatgtt cctctctgcc   1740

tgtgtttctt tgagtatcca gtcaatgatc agaacaagta gcccattcag tgatctgtga   1800

aattgttaca caagggttgg ctactctatg ctagccatgg ttgatgtaag aactcagttt   1860

aatagcatac ttgagcaatc agggtaaaag tgttccacca tgaacaaagg tatcaaatca   1920

caaacctggc tgactgcagt tggatgttag aggtctgttg taaatagtag aacatagatt   1980

gcagactgac ttgttgattt ctactttggc aatagcaata aagcactttg caaag        2035
```

```
<210> SEQ ID NO 26
<211> LENGTH: 406
<212> TYPE: PRT
<213> ORGANISM: Rattus norvegicus
<220> FEATURE:
<223> OTHER INFORMATION: rat Protein C inhibitor (PCI), serine (or
      cysteine) proteinase inhibitor, clade A, member 5
      (Serpina5)

<400> SEQUENCE: 26

Met Arg Phe Phe Pro Ile Leu Cys Leu Val Leu Phe Phe Ser His Gly
1               5                   10                  15

Val Ala Ser Arg Gln Arg Ser His Ser Lys Glu Lys Lys Lys Ser Lys
            20                  25                  30

Glu Ser Ser Val Gly Ala Val Gly Thr Ser Arg Ser Arg Asp Phe Ala
        35                  40                  45
```

-continued

```
Phe Arg Leu Tyr Arg Ala Leu Ala Ser Glu Ala Pro Gly Gln Asn Val
     50                  55                  60

Phe Phe Ser Pro Met Ser Val Ser Met Ser Leu Gly Met Leu Ser Leu
 65                  70                  75                  80

Gly Ser Gly Leu Lys Thr Lys Ala Gln Ile Leu Glu Gly Leu Gly Leu
                 85                  90                  95

Ser Leu Gln Gln Gly Gln Glu Asp Met Leu His Lys Gly Phe Gln Gln
            100                 105                 110

Leu Leu Gln Gln Phe Ser Gln Pro Ser Asp Gly Leu Gln Leu Ser Leu
        115                 120                 125

Gly Ser Ala Leu Phe Thr Asp Pro Ala Val His Ile Arg Asp His Phe
    130                 135                 140

Leu Ser Ala Met Lys Thr Leu Tyr Met Ser Asp Met Phe Ser Thr Asn
145                 150                 155                 160

Phe Gly Asn Pro Glu Ser Ala Lys Lys Gln Ile Asn Asp Tyr Val Ala
                165                 170                 175

Lys Lys Thr Asn Gly Lys Ile Val Asp Leu Ile Lys Asp Leu Asp Ser
            180                 185                 190

Thr His Val Met Val Val Val Asn Tyr Ile Phe Phe Lys Ala Lys Trp
        195                 200                 205

Gln Thr Ala Phe Ser Ser Thr Asn Thr His Lys Met Asp Phe His Val
    210                 215                 220

Thr Pro Lys Lys Thr Ile Gln Val Pro Met Met Asn Arg Glu Asp Ile
225                 230                 235                 240

Tyr Ser Tyr Ile Leu Asp Gln Asn Ile Ser Cys Thr Val Val Gly Ile
                245                 250                 255

Pro Tyr Gln Gly Asn Thr Phe Ala Leu Phe Ile Leu Pro Ser Glu Gly
            260                 265                 270

Lys Met Lys Arg Val Glu Asp Gly Leu Asp Glu Arg Thr Leu Arg Asn
        275                 280                 285

Trp Leu Lys Met Phe Thr Lys Arg Gln Leu Asp Leu Tyr Leu Pro Lys
    290                 295                 300

Phe Ser Ile Glu Gly Thr Tyr Lys Leu Glu Lys Ile Leu Pro Lys Leu
305                 310                 315                 320

Gly Ile Gln Asp Ile Phe Thr Thr His Ala Asp Leu Ser Gly Leu Thr
                325                 330                 335

Asp His Thr Asn Ile Lys Leu Ser Glu Met Val His Lys Ser Met Val
            340                 345                 350

Glu Val Asp Glu Ser Gly Thr Thr Ala Ala Ala Ser Thr Gly Ile Leu
        355                 360                 365

Phe Thr Leu Arg Ser Ala Arg Pro Ser Ser Leu Lys Val Glu Phe Thr
    370                 375                 380

Arg Pro Phe Leu Val Val Ile Met Asp Gly Thr Asn Leu Tyr Phe Ile
385                 390                 395                 400

Gly Lys Val Ile Gln Pro
                405


<210> SEQ ID NO 27
<211> LENGTH: 5073
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
```

-continued

```
<223> OTHER INFORMATION: human MAST205b novel variant
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(5073)
<223> OTHER INFORMATION: MAST205b novel variant

<400> SEQUENCE: 27

atgttttcac ccacatctgc tccagccctc ttcctcacta aagtcccatt tagtgctgat     60

tgtgctttgg ctacttctcc tcttgccatt ttcctgaacc cacgagccca cagcagtcct    120

ggcactcctt gttccagccg cccactgccg tggagttgtc ggacaagtaa ccgcaagagc    180

ttgattgtga cctctagcac atcacctaca ctaccacggc cacactcacc actccatggc    240

cacacaggta acagtccttt ggacagcccc cggaatttct ctccaaatgc acctgctcac    300

ttttcttttg ttcctgcccg tagccatagc cacagagctg acaggactga tgggcggcgc    360

tggtctttgg cctctttgcc ctcttcagga tatggaacta acactcctag ctccactgtc    420

tcatcatcat gctcctcaca ggaaaagctg catcagttgc ttttccagcc tacagctgat    480

gagctgcact ttttgacgaa gcatttcagc acagagagcg taccagatga ggaaggacgg    540

cagtccccag ccatgcggcc tcgctcccgg agcctcagtc ccggacgatc cccagtatcc    600

tttgacagtg aaataataat gatgaatcat gtttacaaag aaagattccc aaaggccacc    660

gcacaaatgg aagagcgact agcagagttt atttcctcca acactccaga cagcgtgctg    720

cccttggcag atggagccct gagctttatt catcatcagg tgattgagat ggcccgagac    780

tgcctggata aatctcggag tggcctcatt acatcacaat acttctacga acttcaagag    840

aatttggaga aacttttaca agatgctcat gagcgctcag agagctcaga agtggccttt    900

gtgatgcagc tggtgaaaaa gctgatgatt atcattgccc gcccagcacg tctcctggaa    960

tgcctggagt ttgaccctga agagttctac cacctttttag aagcagctga gggccacgcc   1020

aaagagggac aagggattaa atgtgacatt ccccgctaca tcgttagcca gctgggcctc   1080

acccgggatc cctagaaga aatggcccag ttgagcagct gtgacagtcc tgacactcca   1140

gagacagatg attctattga gggccatggg gcatctctgc catctaaaaa gacaccctct   1200

gaagaggact tcgagaccat taagctcatc agcaatggcg cctatggggc tgtatttctg   1260

gtgcggcaca agtccacccg gcagcgcttt gccatgaaga agatcaacaa gcagaacctg   1320

atcctacgga accagatcca gcaggccttc gtggagcgtg acatactgac tttcgctgag   1380

aacccctttg tggtcagcat gttctgctcc tttgatacca agcgccactt gtgcatggtg   1440

atggagtacg ttgaaggggg agactgtgcc actctgctga agaatattgg ggccctgcct   1500

gtggacatgg tgcgtctata ctttgcggaa actgtgctgg ccctggagta cttacacaac   1560

tatggcatcg tgcaccgtga cctcaagcct gacaacctcc taattacatc catggggcac   1620

atcaagctca cggactttgg actgtccaaa atgggcctca tgagtctgac aacgaacttg   1680

tatgagggtc atattgaaaa ggatgcccgg gaattcctgg acaagcaggt atgcgggacc   1740

ccagaataca ttgcgcctga ggtgatcctg cgccagggct atgggaagcc agtggactgg   1800

tgggccatgg gcattatcct gtatgagttc ctggtgggct gcgtcccttt ttttggagat   1860

actccggagg agctctttgg gcaggtgatc agtgatgaga ttgtgtggcc tgagggtgat   1920

gaggcactgc ccccagacgc ccaggacctc acctccaaac tgctccacca gaaccctctg   1980

gagagacttg gcacaggcag tgcctatgag gtgaagcagc acccattctt tactggtctg   2040
```

-continued

```
gactggacag gacttctccg ccagaaggct gaatttattc ctcagttgga gtcagaggat   2100
gatactagct attttgacac ccgctcagag cgataccacc acatggactc ggaggatgag   2160
gaagaagtga gtgaggatgg ctgccttgag atccgccagt tctcttcctg ctctccaagg   2220
ttcaacaagg tgtacagcag catggagcgg ctctcactgc tcgaggagcg ccggacacca   2280
cccccgacca agcgcagcct gagtgaggag aaggaggacc attcagatgg cctggcaggg   2340
ctcaaaggcc gagaccggag ctgggtgatt ggctcccctg agatattacg gaagcggctg   2400
tcggtgtctg agtcgtccca cacagagagt gactcaagcc ctccaatgac agtgcgacgc   2460
cgctgctcag gcctcctgga tgcgcctcgg ttcccggagg gccctgagga ggccagcagc   2520
accctcagga ggcaaccaca ggagggtata tgggtcctga cacccccatc tggagagggg   2580
gtatctgggc ctgtcactga acactcaggg gagcagcggc caaagctgga tgaggaagct   2640
gttggccgga gcagtggttc cagtccagct atggagaccc gaggccgtgg gacctcacag   2700
ctggctgagg gagccacagc caaggccatc agtgacctgg ctgtgcgtag ggcccgccac   2760
cggctgctct ctggggactc aacagagaag cgcactgctc gccctgtcaa caaagtgatc   2820
aagtccgcct cagccacagc cctctcactc ctcattcctt cggaacacca cacctgctcc   2880
ccgttggcca gccccatgtc cccacattct cagtcgtcca acccatcatc ccgggactct   2940
tctccaagca gggacttctt gccagccctt ggcagcatga ggcctcccat catcatccac   3000
cgagctggca agaagtatgg cttcaccctg cgggccattc gcgtctacat gggtgactcc   3060
gatgtctaca ccgtgcacca tatggtgtgg cacgtggagg atggaggtcc ggccagtgag   3120
gcagggcttc gtcaaggtga cctcatcacc catgtcaatg gggaacctgt gcatggcctg   3180
gtgcacacgg aggtggtgga gctgatcctg aagagtggaa acaaggtggc catttcaaca   3240
actcccctgg agaacacatc cattaaagtg gggccagctc ggaagggcag ctacaaggcc   3300
aagatggccc gaaggagcaa gaggagccgc ggcaaggatg ggcaagaaag cagaaaaagg   3360
agctccctgt tccgcaagat caccaagcaa gcatccctgc tccacaccag ccgcagcctt   3420
tcttccctta accgctcctt gtcatcaggg gagagtgggc caggctctcc cacacacagc   3480
cacagccttt ccccccgatc tcccactcaa ggctaccggg tgacccccga tgctgtgcat   3540
tcagtgggag ggaattcatc acagagcagc tcccccagct ccagcgtgcc cagttcccca   3600
gccggctctg ggcacacacg gcccagctcc ctccacggtc tggcacccaa gctccaacgc   3660
cagtaccgct ctccacggcg caagtcagca ggcagcatcc cactgtcacc actggcccac   3720
accccttctc ccccaccccc aacagcttca cctcagcggt ccccatcgcc cctgtctggc   3780
catgtagccc aggcctttcc cacaaagctt cacttgtcac ctcccctggg caggcaactc   3840
tcacggccca agagtgcgga gccacccgt tcaccactac tcaagagggt gcagtcggct   3900
gagaaactgg cagcagcact tgccgcctct gagaagaagc tagccacttc tcgcaagcac   3960
agccttgacc tgccccactc tgaactaaag aaggaactgc cgcccaggga agtgagccct   4020
ctggaggtag ttggagccag gagtgtgctg tctggcaagg gggccctgcc agggaagggg   4080
gtgctgcagc ctgctccctc acgggcccta ggcaccctcc ggcaggaccg agccgaacga   4140
cgggagtcgc tgcagaagca agaagccatt cgtgaggtgg actcctcaga ggacgacacc   4200
gaggaagggc ctgagaacag ccagggtgca caggagctga gcttggcacc tcacccagaa   4260
gtgagccaga gtgtggcccc taaaggagca ggagagagtg gggaagagga tcctttcccg   4320
```

-continued

```
tccagaggcc ctaggagcct gggcccaatg gtcccaagcc tattgacagg gatcacactg   4380

gggcctccca gaatggaaag tcccagtggt ccccacagga ggctcgggag cccacaagcc   4440

attgaggagg ctgccagctc ctcctcagca ggccccaacc taggtcagtc tggagccaca   4500

gaccccatcc ctcctgaagg ttgctggaag gcccagcacc tccacaccca ggcactaaca   4560

gcactttctc ccagcacttc gggactcacc cccaccagca gttgctctcc tcccagctcc   4620

acctctggga agctgagcat gtggtcctgg aaatccctta ttgagggccc agacagggca   4680

tccccaagca gaaaggcaac catggcaggt gggctagcca acctccagga tttggaaaac   4740

acaactccag cccagcctaa gaacctgtct cccagggagc aggggaagac acagccacct   4800

agtgccccca gactggccca tccatcttat gaggatccca gccagggctg gctatgggag   4860

tctgagtgtg cacaagcagt gaaagaggat ccagccctga gcatcaccca agtgcctgat   4920

gcctcaggtg acagaaggca ggacgttcca tgccgaggct gccccctcac ccagaagtct   4980

gagcccagcc tcaggagggg ccaagaacca gggggccatc aaaagcatcg ggatttggca   5040

ttggttccag atgagctttt aaagcaaaca tag                                5073
```

<210> SEQ ID NO 28
<211> LENGTH: 1690
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human MAST205b novel variant

<400> SEQUENCE: 28

```
Met Phe Ser Pro Thr Ser Ala Pro Ala Leu Phe Leu Thr Lys Val Pro
1               5                   10                  15

Phe Ser Ala Asp Cys Ala Leu Ala Thr Ser Pro Leu Ala Ile Phe Leu
            20                  25                  30

Asn Pro Arg Ala His Ser Ser Pro Gly Thr Pro Cys Ser Ser Arg Pro
        35                  40                  45

Leu Pro Trp Ser Cys Arg Thr Ser Asn Arg Lys Ser Leu Ile Val Thr
    50                  55                  60

Ser Ser Thr Ser Pro Thr Leu Pro Arg Pro His Ser Pro Leu His Gly
65                  70                  75                  80

His Thr Gly Asn Ser Pro Leu Asp Ser Pro Arg Asn Phe Ser Pro Asn
                85                  90                  95

Ala Pro Ala His Phe Ser Phe Val Pro Ala Arg Ser His Ser His Arg
            100                 105                 110

Ala Asp Arg Thr Asp Gly Arg Arg Trp Ser Leu Ala Ser Leu Pro Ser
        115                 120                 125

Ser Gly Tyr Gly Thr Asn Thr Pro Ser Ser Thr Val Ser Ser Ser Cys
    130                 135                 140

Ser Ser Gln Glu Lys Leu His Gln Leu Leu Phe Gln Pro Thr Ala Asp
145                 150                 155                 160

Glu Leu His Phe Leu Thr Lys His Phe Ser Thr Glu Ser Val Pro Asp
                165                 170                 175

Glu Glu Gly Arg Gln Ser Pro Ala Met Arg Pro Arg Ser Arg Ser Leu
            180                 185                 190

Ser Pro Gly Arg Ser Pro Val Ser Phe Asp Ser Glu Ile Ile Met Met
        195                 200                 205

Asn His Val Tyr Lys Glu Arg Phe Pro Lys Ala Thr Ala Gln Met Glu
```

-continued

```
    210                 215                 220

Glu Arg Leu Ala Glu Phe Ile Ser Ser Asn Thr Pro Asp Ser Val Leu
225                 230                 235                 240

Pro Leu Ala Asp Gly Ala Leu Ser Phe Ile His His Gln Val Ile Glu
                245                 250                 255

Met Ala Arg Asp Cys Leu Asp Lys Ser Arg Ser Gly Leu Ile Thr Ser
            260                 265                 270

Gln Tyr Phe Tyr Glu Leu Gln Glu Asn Leu Glu Lys Leu Leu Gln Asp
        275                 280                 285

Ala His Glu Arg Ser Glu Ser Ser Glu Val Ala Phe Val Met Gln Leu
    290                 295                 300

Val Lys Lys Leu Met Ile Ile Ile Ala Arg Pro Ala Arg Leu Leu Glu
305                 310                 315                 320

Cys Leu Glu Phe Asp Pro Glu Glu Phe Tyr His Leu Leu Glu Ala Ala
                325                 330                 335

Glu Gly His Ala Lys Glu Gly Gln Gly Ile Lys Cys Asp Ile Pro Arg
            340                 345                 350

Tyr Ile Val Ser Gln Leu Gly Leu Thr Arg Asp Pro Leu Glu Glu Met
        355                 360                 365

Ala Gln Leu Ser Ser Cys Asp Ser Pro Asp Thr Pro Glu Thr Asp Asp
    370                 375                 380

Ser Ile Glu Gly His Gly Ala Ser Leu Pro Ser Lys Lys Thr Pro Ser
385                 390                 395                 400

Glu Glu Asp Phe Glu Thr Ile Lys Leu Ile Ser Asn Gly Ala Tyr Gly
                405                 410                 415

Ala Val Phe Leu Val Arg His Lys Ser Thr Arg Gln Arg Phe Ala Met
            420                 425                 430

Lys Lys Ile Asn Lys Gln Asn Leu Ile Leu Arg Asn Gln Ile Gln Gln
        435                 440                 445

Ala Phe Val Glu Arg Asp Ile Leu Thr Phe Ala Glu Asn Pro Phe Val
    450                 455                 460

Val Ser Met Phe Cys Ser Phe Asp Thr Lys Arg His Leu Cys Met Val
465                 470                 475                 480

Met Glu Tyr Val Glu Gly Gly Asp Cys Ala Thr Leu Leu Lys Asn Ile
                485                 490                 495

Gly Ala Leu Pro Val Asp Met Val Arg Leu Tyr Phe Ala Glu Thr Val
            500                 505                 510

Leu Ala Leu Glu Tyr Leu His Asn Tyr Gly Ile Val His Arg Asp Leu
        515                 520                 525

Lys Pro Asp Asn Leu Leu Ile Thr Ser Met Gly His Ile Lys Leu Thr
    530                 535                 540

Asp Phe Gly Leu Ser Lys Met Gly Leu Met Ser Leu Thr Thr Asn Leu
545                 550                 555                 560

Tyr Glu Gly His Ile Glu Lys Asp Ala Arg Glu Phe Leu Asp Lys Gln
                565                 570                 575

Val Cys Gly Thr Pro Glu Tyr Ile Ala Pro Glu Val Ile Leu Arg Gln
            580                 585                 590

Gly Tyr Gly Lys Pro Val Asp Trp Trp Ala Met Gly Ile Ile Leu Tyr
        595                 600                 605

Glu Phe Leu Val Gly Cys Val Pro Phe Phe Gly Asp Thr Pro Glu Glu
    610                 615                 620
```

-continued

```
Leu Phe Gly Gln Val Ile Ser Asp Glu Ile Val Trp Pro Glu Gly Asp
625                 630                 635                 640

Glu Ala Leu Pro Pro Asp Ala Gln Asp Leu Thr Ser Lys Leu Leu His
                645                 650                 655

Gln Asn Pro Leu Glu Arg Leu Gly Thr Gly Ser Ala Tyr Glu Val Lys
            660                 665                 670

Gln His Pro Phe Phe Thr Gly Leu Asp Trp Thr Gly Leu Leu Arg Gln
        675                 680                 685

Lys Ala Glu Phe Ile Pro Gln Leu Glu Ser Glu Asp Asp Thr Ser Tyr
    690                 695                 700

Phe Asp Thr Arg Ser Glu Arg Tyr His His Met Asp Ser Glu Asp Glu
705                 710                 715                 720

Glu Glu Val Ser Glu Asp Gly Cys Leu Glu Ile Arg Gln Phe Ser Ser
                725                 730                 735

Cys Ser Pro Arg Phe Asn Lys Val Tyr Ser Ser Met Glu Arg Leu Ser
            740                 745                 750

Leu Leu Glu Glu Arg Arg Thr Pro Pro Pro Thr Lys Arg Ser Leu Ser
        755                 760                 765

Glu Glu Lys Glu Asp His Ser Asp Gly Leu Ala Gly Leu Lys Gly Arg
    770                 775                 780

Asp Arg Ser Trp Val Ile Gly Ser Pro Glu Ile Leu Arg Lys Arg Leu
785                 790                 795                 800

Ser Val Ser Glu Ser Ser His Thr Glu Ser Asp Ser Ser Pro Pro Met
                805                 810                 815

Thr Val Arg Arg Arg Cys Ser Gly Leu Leu Asp Ala Pro Arg Phe Pro
            820                 825                 830

Glu Gly Pro Glu Glu Ala Ser Ser Thr Leu Arg Arg Gln Pro Gln Glu
        835                 840                 845

Gly Ile Trp Val Leu Thr Pro Pro Ser Gly Glu Gly Val Ser Gly Pro
    850                 855                 860

Val Thr Glu His Ser Gly Glu Gln Arg Pro Lys Leu Asp Glu Glu Ala
865                 870                 875                 880

Val Gly Arg Ser Ser Gly Ser Ser Pro Ala Met Glu Thr Arg Gly Arg
                885                 890                 895

Gly Thr Ser Gln Leu Ala Glu Gly Ala Thr Ala Lys Ala Ile Ser Asp
            900                 905                 910

Leu Ala Val Arg Arg Ala Arg His Arg Leu Leu Ser Gly Asp Ser Thr
        915                 920                 925

Glu Lys Arg Thr Ala Arg Pro Val Asn Lys Val Ile Lys Ser Ala Ser
    930                 935                 940

Ala Thr Ala Leu Ser Leu Leu Ile Pro Ser Glu His His Thr Cys Ser
945                 950                 955                 960

Pro Leu Ala Ser Pro Met Ser Pro His Ser Gln Ser Ser Asn Pro Ser
                965                 970                 975

Ser Arg Asp Ser Ser Pro Ser Arg Asp Phe Leu Pro Ala Leu Gly Ser
            980                 985                 990

Met Arg Pro Pro Ile Ile Ile His Arg Ala Gly Lys Lys Tyr Gly Phe
        995                 1000                1005

Thr Leu Arg Ala Ile Arg Val Tyr Met Gly Asp Ser Asp Val Tyr Thr
   1010                 1015                1020
```

-continued

```
Val His His Met Val Trp His Val Glu Asp Gly Gly Pro Ala Ser Glu
1025                1030                1035                1040

Ala Gly Leu Arg Gln Gly Asp Leu Ile Thr His Val Asn Gly Glu Pro
                1045                1050                1055

Val His Gly Leu Val His Thr Glu Val Val Glu Leu Ile Leu Lys Ser
            1060                1065                1070

Gly Asn Lys Val Ala Ile Ser Thr Thr Pro Leu Glu Asn Thr Ser Ile
        1075                1080                1085

Lys Val Gly Pro Ala Arg Lys Gly Ser Tyr Lys Ala Lys Met Ala Arg
    1090                1095                1100

Arg Ser Lys Arg Ser Arg Gly Lys Asp Gly Gln Glu Ser Arg Lys Arg
1105                1110                1115                1120

Ser Ser Leu Phe Arg Lys Ile Thr Lys Gln Ala Ser Leu Leu His Thr
                1125                1130                1135

Ser Arg Ser Leu Ser Ser Leu Asn Arg Ser Leu Ser Ser Gly Glu Ser
            1140                1145                1150

Gly Pro Gly Ser Pro Thr His Ser His Ser Leu Ser Pro Arg Ser Pro
        1155                1160                1165

Thr Gln Gly Tyr Arg Val Thr Pro Asp Ala Val His Ser Val Gly Gly
    1170                1175                1180

Asn Ser Ser Gln Ser Ser Ser Pro Ser Ser Ser Val Pro Ser Ser Pro
1185                1190                1195                1200

Ala Gly Ser Gly His Thr Arg Pro Ser Ser Leu His Gly Leu Ala Pro
                1205                1210                1215

Lys Leu Gln Arg Gln Tyr Arg Ser Pro Arg Arg Lys Ser Ala Gly Ser
            1220                1225                1230

Ile Pro Leu Ser Pro Leu Ala His Thr Pro Ser Pro Pro Pro Pro Thr
        1235                1240                1245

Ala Ser Pro Gln Arg Ser Pro Ser Pro Leu Ser Gly His Val Ala Gln
    1250                1255                1260

Ala Phe Pro Thr Lys Leu His Leu Ser Pro Pro Leu Gly Arg Gln Leu
1265                1270                1275                1280

Ser Arg Pro Lys Ser Ala Glu Pro Pro Arg Ser Pro Leu Leu Lys Arg
                1285                1290                1295

Val Gln Ser Ala Glu Lys Leu Ala Ala Ala Leu Ala Ala Ser Glu Lys
            1300                1305                1310

Lys Leu Ala Thr Ser Arg Lys His Ser Leu Asp Leu Pro His Ser Glu
        1315                1320                1325

Leu Lys Lys Glu Leu Pro Pro Arg Glu Val Ser Pro Leu Glu Val Val
    1330                1335                1340

Gly Ala Arg Ser Val Leu Ser Gly Lys Gly Ala Leu Pro Gly Lys Gly
1345                1350                1355                1360

Val Leu Gln Pro Ala Pro Ser Arg Ala Leu Gly Thr Leu Arg Gln Asp
                1365                1370                1375

Arg Ala Glu Arg Arg Glu Ser Leu Gln Lys Gln Glu Ala Ile Arg Glu
            1380                1385                1390

Val Asp Ser Ser Glu Asp Asp Thr Glu Glu Gly Pro Glu Asn Ser Gln
        1395                1400                1405

Gly Ala Gln Glu Leu Ser Leu Ala Pro His Pro Glu Val Ser Gln Ser
    1410                1415                1420

Val Ala Pro Lys Gly Ala Gly Glu Ser Gly Glu Glu Asp Pro Phe Pro
```

-continued

```
1425                1430                1435                1440

Ser Arg Gly Pro Arg Ser Leu Gly Pro Met Val Pro Ser Leu Leu Thr
               1445                1450                1455

Gly Ile Thr Leu Gly Pro Pro Arg Met Glu Ser Pro Ser Gly Pro His
           1460                1465                1470

Arg Arg Leu Gly Ser Pro Gln Ala Ile Glu Glu Ala Ala Ser Ser Ser
       1475                1480                1485

Ser Ala Gly Pro Asn Leu Gly Gln Ser Gly Ala Thr Asp Pro Ile Pro
   1490                1495                1500

Pro Glu Gly Cys Trp Lys Ala Gln His Leu His Thr Gln Ala Leu Thr
1505                1510                1515                1520

Ala Leu Ser Pro Ser Thr Ser Gly Leu Thr Pro Thr Ser Ser Cys Ser
               1525                1530                1535

Pro Pro Ser Ser Thr Ser Gly Lys Leu Ser Met Trp Ser Trp Lys Ser
           1540                1545                1550

Leu Ile Glu Gly Pro Asp Arg Ala Ser Pro Ser Arg Lys Ala Thr Met
       1555                1560                1565

Ala Gly Gly Leu Ala Asn Leu Gln Asp Leu Glu Asn Thr Thr Pro Ala
   1570                1575                1580

Gln Pro Lys Asn Leu Ser Pro Arg Glu Gln Gly Lys Thr Gln Pro Pro
1585                1590                1595                1600

Ser Ala Pro Arg Leu Ala His Pro Ser Tyr Glu Asp Pro Ser Gln Gly
               1605                1610                1615

Trp Leu Trp Glu Ser Glu Cys Ala Gln Ala Val Lys Glu Asp Pro Ala
           1620                1625                1630

Leu Ser Ile Thr Gln Val Pro Asp Ala Ser Gly Asp Arg Arg Gln Asp
       1635                1640                1645

Val Pro Cys Arg Gly Cys Pro Leu Thr Gln Lys Ser Glu Pro Ser Leu
   1650                1655                1660

Arg Arg Gly Gln Glu Pro Gly Gly His Gln Lys His Arg Asp Leu Ala
1665                1670                1675                1680

Leu Val Pro Asp Glu Leu Leu Lys Gln Thr
               1685                1690
```

<210> SEQ ID NO 29
<211> LENGTH: 5737
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human MAST205 cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (284)..(5488)
<223> OTHER INFORMATION: MAST205

<400> SEQUENCE: 29

```
taggcaggcg gctgagccgg cggcgggtgg cctgcccaac gtgtgctggg tgggagaagg    60

cgaggcggca gcgatgctgt ctcttccgtg aggagcgcag aggaggtcgc ggcgccggag   120

gccccagaag gctcgaaggc gccgcgggct ggggtcggtg gcttagggag cccgtccggc   180

catggtggcc gcgggtggtg gttggcgcgg ctgcgctgcg gcccggggca gtgcggagcc   240

gggacagtcg cggcgctgac gcccgcgggc cccagctgca gatatgaagc ggagccgctg   300

ccgcgaccga ccgcagccgc cgccgcccga ccgccgggag gatggagttc agcgggcagc   360
```

-continued

```
ggagctgtct cagtctttgc cgccgcgccg gcgagcgccg cccgggaggc agcggctgga    420
ggagcggacg ggccccgcgg ggcccgaggg caaggagcag gatgtagtaa ctggagttag    480
tcccctgctc ttcaggaaac tcagtaatcc tgacatattt tcatccactg gaaaagttaa    540
acttcagcga caactgagtc aggatgattg taagttatgg agaggaaacc tggccagctc    600
tctatcgggt aagcagctgc tccctttgtc cagcagtgta catagcagtg tgggacaggt    660
gacttggcag tcgtcaggag aagcatcaaa cctggttcga atgagaaacc agtcccttgg    720
acagtctgca ccttctctta ctgctggcct gaaggagttg agccttccaa gaagaggcag    780
cttttgtcgg acaagtaacc gcaagagctt gattgtgacc tctagcacat cacctacact    840
accacggcca cactcaccac tccatggcca cacaggtaac agtcctttgg acagcccccg    900
gaatttctct ccaaatgcac ctgctcactt ttcttttgtt cctgcccgta ggactgatgg    960
gcggcgctgg tctttggcct ctttgccctc ttcaggatat ggaactaaca ctcctagctc   1020
cactgtctca tcatcatgct cctcacagga aaagctgcat cagttgcctt tccagcctac   1080
agctgatgag ctgcactttt tgacgaagca tttcagcaca gagagcgtac cagatgagga   1140
aggacggcag tccccagcca tgcggcctcg ctcccggagc ctcagtcccg gacgatcccc   1200
agtatccttt gacagtgaaa taataatgat gaatcacgtt tacaaagaaa gattcccaaa   1260
ggccaccgca caaatggaag agcgactagc agagtttatt tcctccaaca ctccagacag   1320
cgtgctgccc ttggcagatg gagccctgag ctttattcat catcaggtga ttgagatggc   1380
ccgagactgc ctggataaat ctcggagtgg cctcattaca tcacaatact tctacgaact   1440
tcaagagaat ttggagaaac ttttacaaga tgctcatgag cgctcagaga gctcagaagt   1500
ggcttttgtg atgcagctgg tgaaaaagct gatgattatc attgcccgcc cagcacgtct   1560
cctggaatgc ctggagtttg accctgaaga gttctaccac cttttagaag cagctgaggg   1620
ccacgccaaa gagggacaag ggattaaatg tgacattccc cgctacatcg ttagccagct   1680
gggcctcacc cgggatcccc tagaagaaat ggcccagttg agcagctgtg acagtcctga   1740
cactccagag acagatgatt ctattgaggg ccatggggca tctctgccat ctaaaaagac   1800
accctctgaa gaggacttcg agaccattaa gctcatcagc aatggcgcct atggggctgt   1860
atttctggtg cggcacaagt ccacccggca gcgctttgcc atgaagaaga tcaacaagca   1920
gaacctgatc ctacggaacc agatccagca ggccttcgtg gagcgtgaca tactgacttt   1980
cgctgagaac ccctttgtgg tcagcatgtt ctgctccttt gataccaagc gccacttgtg   2040
catggtgatg gagtacgttg aagggggaga ctgtgccact ctgctgaaga atattggggc   2100
cctgcctgtg gacatggtgc gtctatactt tgcggaaact gtgctggccc tggagtactt   2160
acacaactat ggcatcgtgc accgtgacct caagcctgac aacctcctaa ttacatccat   2220
ggggcacatc aagctcacgg actttggact gtccaaaatg ggcctcatga gtctgacaac   2280
gaacttgtat gagggtcata ttgaaaagga tgcccgggaa ttcctggaca agcaggtatg   2340
cgggacccca gaatacattg cgcctgaggt gatcctgcgc cagggctatg ggaagccagt   2400
ggactggtgg gccatgggca ttatcctgta tgagttcctg gtgggctgcg tccctttttt   2460
tggagatact ccggaggagc tctttgggca ggtgatcagt gatgagattg tgtggcctga   2520
gggtgatgag gcactgcccc cagacgccca ggacctcacc tccaaactgc tccaccagaa   2580
ccctctggag agacttggca caggcagtgc ctatgaggtg aagcagcacc cattctttac   2640
```

-continued

```
tggtctggac tggacaggac ttctccgcca gaaggctgaa tttattcctc agttggagtc   2700
agaggatgat actagctatt ttgacacccg ctcagagcga taccaccaca tggactcgga   2760
ggatgaggaa gaagtgagtg aggatggctg ccttgagatc cgccagttct cttcctgctc   2820
tccaaggttc aacaaggtgt acagcagcat ggagcggctc tcactgctcg aggagcgccg   2880
gacaccaccc ccgaccaagc gcagcctgag tgaggagaag gaggaccatt cagatggcct   2940
ggcagggctc aaaggccgag accggagctg ggtgattggc tcccctgaga tattacggaa   3000
gcggctgtcg gtgtctgagt cgtcccacac agagagtgac tcaagccctc caatgacagt   3060
gcgacgccgc tgctcaggcc tcctggatgc gcctcggttc ccggagggcc ctgaggaggc   3120
cagcagcacc ctcaggaggc aaccacagga gggtatatgg gtcctgacac ccccatctgg   3180
agaggggggta tctgggcctg tcactgaaca ctcaggggag cagcggccaa agctggatga   3240
ggaagctgtt ggccggagca gtggttccag tccagctatg gagacccgag gccgtgggac   3300
ctcacagctg gctgagggag ccacagccaa ggccatcagt gacctggctg tgcgtagggc   3360
ccgccaccgg ctgctctctg ggactcaac agagaagcgc actgctcgcc ctgtcaacaa   3420
agtgatcaag tccgcctcag ccacagccct ctcactcctc attccttcgg aacaccacac   3480
ctgctccccg ttggccagcc ccatgtcccc acattctcag tcgtccaacc catcatcccg   3540
ggactcttct ccaagcaggg acttcttgcc agcccttggc agcatgaggc ctcccatcat   3600
catccaccga gctggcaaga agtatggctt caccctgcgg gccattcgcg tctacatggg   3660
tgactccgat gtctacaccg tgcaccatat ggtgtggcac gtggaggatg gaggtccggc   3720
cagtgaggca gggcttcgtc aaggtgacct catcacccat gtcaatgggg aacctgtgca   3780
tggcctggtg cacacggagg tggtggagct gatcctgaag agtggaaaca aggtggccat   3840
ttcaacaact cccctggaga acacatccat taaagtgggg ccagctcgga agggcagcta   3900
caaggccaag atggcccgaa ggagcaagag gagccgcggc aaggatgggc aagaaagcag   3960
aaaaaggagc tccctgttcc gcaagatcac caagcaagca tccctgctcc acaccagccg   4020
cagcctttct tcccttaacc gctccttgtc atcaggggag agtgggccag gctctcccac   4080
acacagccac agcctttccc cccgatctcc cactcaaggc taccgggtga cccccgatgc   4140
tgtgcattca gtgggaggga attcatcaca gagcagctcc cccagctcca gcgtgcccag   4200
ttccccagcc ggctctgggc acacacggcc cagctccctc cacggtctgg cacccaagct   4260
ccaacgccag taccgctctc cacggcgcaa gtcagcaggc agcatcccac tgtcaccact   4320
ggcccacacc ccttctcccc cacccccaac agcttcacct cagcggtccc catcgcccct   4380
gtctggccat gtagcccagg cctttcccac aaagcttcac ttgtcacctc ccctgggcag   4440
gcaactctca cggcccaaga gtgcggagcc accccgttca ccactactca agagggtgca   4500
gtcggctgag aaactggcag cagcacttgc cgcctctgag aagaagctag ccacttctcg   4560
caagcacagc cttgacctgc cccactctga actaaagaag gaactgccgc ccagggaagt   4620
gagccctctg gaggtagttg gagccaggag tgtgctgtct ggcaaggggg ccctgccagg   4680
gaagggggtg ctgcagcctg ctccctcacg ggccctaggc accctccggc aggaccgagc   4740
cgaacgacgg gagtcgctgc agaagcaaga agccattcgt gaggtggact cctcagagga   4800
cgacaccgag gaagggcctg agaacagcca gggtgcacag gagctgagct tggcacctca   4860
cccagaagtg agccagagtg tggcccctaa aggagcagga gagagtgggg aagaggatcc   4920
```

-continued

```
tttcccgtcc agaggcccta ggagcctggg cccaatggtc ccaagcctat tgacagggat   4980
cacactgggg cctcccagaa tggaaagtcc cagtggtccc cacaggaggc tcgggagccc   5040
acaagccatt gaggaggctg ccagctcctc ctcagcaggc cccaacctag gtcagtctgg   5100
agccacagac cccatccctc ctgaaggttg ctggaaggcc cagcacctcc acacccaggc   5160
actaacagca ctttctccca gcacttcggg actcaccccc accagcagtt gctctcctcc   5220
cagctccacc tctgggaagc tgagcatgtg gtcctggaaa tcccttattg agggcccaga   5280
cagggcatcc ccaagcagaa aggcaaccat ggcaggtggg ctagccaacc tccaggattt   5340
ggaaacacaa ctccagccca gcctaagaac ctgtctccca gggagcaggg gaagacacag   5400
ccacctagtg cccccagact ggcccatcca tcttatgagg atcccagcca gggctggcta   5460
tgggagtctg agtgtgcaca agcagtgaaa gaggatccag ccctgagcat cacccaagtg   5520
cctgatgcct caggtgacag aaggcaggac gttccatgcc gaggctgccc cctcacccag   5580
aagtctgagc ccagcctcag gaggggccaa gaaccagggg gccatcaaaa gcatcgggat   5640
ttggcattgg ttccagatga gcttttaaag caaacatagc agttgtttgc catttcttgc   5700
actcagacct gtgtaatata tgctcctgga aaccatc                            5737
```

<210> SEQ ID NO 30
<211> LENGTH: 1734
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human MAST205

<400> SEQUENCE: 30

```
Met Lys Arg Ser Arg Cys Arg Asp Arg Pro Gln Pro Pro Pro Pro Asp
  1               5                  10                  15

Arg Arg Glu Asp Gly Val Gln Arg Ala Ala Glu Leu Ser Gln Ser Leu
             20                  25                  30

Pro Pro Arg Arg Arg Ala Pro Pro Gly Arg Gln Arg Leu Glu Glu Arg
         35                  40                  45

Thr Gly Pro Ala Gly Pro Glu Gly Lys Glu Gln Asp Val Val Thr Gly
     50                  55                  60

Val Ser Pro Leu Leu Phe Arg Lys Leu Ser Asn Pro Asp Ile Phe Ser
 65                  70                  75                  80

Ser Thr Gly Lys Val Lys Leu Gln Arg Gln Leu Ser Gln Asp Asp Cys
                 85                  90                  95

Lys Leu Trp Arg Gly Asn Leu Ala Ser Ser Leu Ser Gly Lys Gln Leu
            100                 105                 110

Leu Pro Leu Ser Ser Ser Val His Ser Ser Val Gly Gln Val Thr Trp
        115                 120                 125

Gln Ser Ser Gly Glu Ala Ser Asn Leu Val Arg Met Arg Asn Gln Ser
    130                 135                 140

Leu Gly Gln Ser Ala Pro Ser Leu Thr Ala Gly Leu Lys Glu Leu Ser
145                 150                 155                 160

Leu Pro Arg Arg Gly Ser Phe Cys Arg Thr Ser Asn Arg Lys Ser Leu
                165                 170                 175

Ile Val Thr Ser Ser Thr Ser Pro Thr Leu Pro Arg Pro His Ser Pro
            180                 185                 190

Leu His Gly His Thr Gly Asn Ser Pro Leu Asp Ser Pro Arg Asn Phe
        195                 200                 205
```

-continued

```
Ser Pro Asn Ala Pro Ala His Phe Ser Phe Val Pro Ala Arg Arg Thr
    210                 215                 220

Asp Gly Arg Arg Trp Ser Leu Ala Ser Leu Pro Ser Ser Gly Tyr Gly
225                 230                 235                 240

Thr Asn Thr Pro Ser Ser Thr Val Ser Ser Ser Cys Ser Ser Gln Glu
                245                 250                 255

Lys Leu His Gln Leu Pro Phe Gln Pro Thr Ala Asp Glu Leu His Phe
            260                 265                 270

Leu Thr Lys His Phe Ser Thr Glu Ser Val Pro Asp Glu Glu Gly Arg
        275                 280                 285

Gln Ser Pro Ala Met Arg Pro Arg Ser Arg Ser Leu Ser Pro Gly Arg
    290                 295                 300

Ser Pro Val Ser Phe Asp Ser Glu Ile Ile Met Met Asn His Val Tyr
305                 310                 315                 320

Lys Glu Arg Phe Pro Lys Ala Thr Ala Gln Met Glu Glu Arg Leu Ala
                325                 330                 335

Glu Phe Ile Ser Ser Asn Thr Pro Asp Ser Val Leu Pro Leu Ala Asp
            340                 345                 350

Gly Ala Leu Ser Phe Ile His His Gln Val Ile Glu Met Ala Arg Asp
        355                 360                 365

Cys Leu Asp Lys Ser Arg Ser Gly Leu Ile Thr Ser Gln Tyr Phe Tyr
    370                 375                 380

Glu Leu Gln Glu Asn Leu Glu Lys Leu Leu Gln Asp Ala His Glu Arg
385                 390                 395                 400

Ser Glu Ser Ser Glu Val Ala Phe Val Met Gln Leu Val Lys Lys Leu
                405                 410                 415

Met Ile Ile Ile Ala Arg Pro Ala Arg Leu Leu Glu Cys Leu Glu Phe
            420                 425                 430

Asp Pro Glu Glu Phe Tyr His Leu Leu Glu Ala Ala Glu Gly His Ala
        435                 440                 445

Lys Glu Gly Gln Gly Ile Lys Cys Asp Ile Pro Arg Tyr Ile Val Ser
    450                 455                 460

Gln Leu Gly Leu Thr Arg Asp Pro Leu Glu Glu Met Ala Gln Leu Ser
465                 470                 475                 480

Ser Cys Asp Ser Pro Asp Thr Pro Glu Thr Asp Asp Ser Ile Glu Gly
                485                 490                 495

His Gly Ala Ser Leu Pro Ser Lys Lys Thr Pro Ser Glu Glu Asp Phe
            500                 505                 510

Glu Thr Ile Lys Leu Ile Ser Asn Gly Ala Tyr Gly Ala Val Phe Leu
        515                 520                 525

Val Arg His Lys Ser Thr Arg Gln Arg Phe Ala Met Lys Lys Ile Asn
    530                 535                 540

Lys Gln Asn Leu Ile Leu Arg Asn Gln Ile Gln Gln Ala Phe Val Glu
545                 550                 555                 560

Arg Asp Ile Leu Thr Phe Ala Glu Asn Pro Phe Val Val Ser Met Phe
                565                 570                 575

Cys Ser Phe Asp Thr Lys Arg His Leu Cys Met Val Met Glu Tyr Val
            580                 585                 590

Glu Gly Gly Asp Cys Ala Thr Leu Leu Lys Asn Ile Gly Ala Leu Pro
        595                 600                 605
```

-continued

```
Val Asp Met Val Arg Leu Tyr Phe Ala Glu Thr Val Leu Ala Leu Glu
    610                 615                 620

Tyr Leu His Asn Tyr Gly Ile Val His Arg Asp Leu Lys Pro Asp Asn
625                 630                 635                 640

Leu Leu Ile Thr Ser Met Gly His Ile Lys Leu Thr Asp Phe Gly Leu
                645                 650                 655

Ser Lys Met Gly Leu Met Ser Leu Thr Thr Asn Leu Tyr Glu Gly His
            660                 665                 670

Ile Glu Lys Asp Ala Arg Glu Phe Leu Asp Lys Gln Val Cys Gly Thr
        675                 680                 685

Pro Glu Tyr Ile Ala Pro Glu Val Ile Leu Arg Gln Gly Tyr Gly Lys
    690                 695                 700

Pro Val Asp Trp Trp Ala Met Gly Ile Ile Leu Tyr Glu Phe Leu Val
705                 710                 715                 720

Gly Cys Val Pro Phe Phe Gly Asp Thr Pro Glu Glu Leu Phe Gly Gln
                725                 730                 735

Val Ile Ser Asp Glu Ile Val Trp Pro Glu Gly Asp Glu Ala Leu Pro
            740                 745                 750

Pro Asp Ala Gln Asp Leu Thr Ser Lys Leu Leu His Gln Asn Pro Leu
        755                 760                 765

Glu Arg Leu Gly Thr Gly Ser Ala Tyr Glu Val Lys Gln His Pro Phe
    770                 775                 780

Phe Thr Gly Leu Asp Trp Thr Gly Leu Leu Arg Gln Lys Ala Glu Phe
785                 790                 795                 800

Ile Pro Gln Leu Glu Ser Glu Asp Asp Thr Ser Tyr Phe Asp Thr Arg
                805                 810                 815

Ser Glu Arg Tyr His His Met Asp Ser Glu Asp Glu Glu Glu Val Ser
            820                 825                 830

Glu Asp Gly Cys Leu Glu Ile Arg Gln Phe Ser Ser Cys Ser Pro Arg
        835                 840                 845

Phe Asn Lys Val Tyr Ser Ser Met Glu Arg Leu Ser Leu Leu Glu Glu
    850                 855                 860

Arg Arg Thr Pro Pro Pro Thr Lys Arg Ser Leu Ser Glu Glu Lys Glu
865                 870                 875                 880

Asp His Ser Asp Gly Leu Ala Gly Leu Lys Gly Arg Asp Arg Ser Trp
                885                 890                 895

Val Ile Gly Ser Pro Glu Ile Leu Arg Lys Arg Leu Ser Val Ser Glu
            900                 905                 910

Ser Ser His Thr Glu Ser Asp Ser Ser Pro Pro Met Thr Val Arg Arg
        915                 920                 925

Arg Cys Ser Gly Leu Leu Asp Ala Pro Arg Phe Pro Glu Gly Pro Glu
    930                 935                 940

Glu Ala Ser Ser Thr Leu Arg Arg Gln Pro Gln Glu Gly Ile Trp Val
945                 950                 955                 960

Leu Thr Pro Pro Ser Gly Glu Gly Val Ser Gly Pro Val Thr Glu His
                965                 970                 975

Ser Gly Glu Gln Arg Pro Lys Leu Asp Glu Glu Ala Val Gly Arg Ser
            980                 985                 990

Ser Gly Ser Ser Pro Ala Met Glu Thr Arg Gly Arg Gly Thr Ser Gln
        995                 1000                1005

Leu Ala Glu Gly Ala Thr Ala Lys Ala Ile Ser Asp Leu Ala Val Arg
```

-continued

```
   1010                1015                1020

Arg Ala Arg His Arg Leu Leu Ser Gly Asp Ser Thr Glu Lys Arg Thr
1025                1030                1035                1040

Ala Arg Pro Val Asn Lys Val Ile Lys Ser Ala Ser Ala Thr Ala Leu
                1045                1050                1055

Ser Leu Leu Ile Pro Ser Glu His His Thr Cys Ser Pro Leu Ala Ser
            1060                1065                1070

Pro Met Ser Pro His Ser Gln Ser Ser Asn Pro Ser Ser Arg Asp Ser
        1075                1080                1085

Ser Pro Ser Arg Asp Phe Leu Pro Ala Leu Gly Ser Met Arg Pro Pro
    1090                1095                1100

Ile Ile Ile His Arg Ala Gly Lys Lys Tyr Gly Phe Thr Leu Arg Ala
1105                1110                1115                1120

Ile Arg Val Tyr Met Gly Asp Ser Asp Val Tyr Thr Val His His Met
                1125                1130                1135

Val Trp His Val Glu Asp Gly Gly Pro Ala Ser Glu Ala Gly Leu Arg
            1140                1145                1150

Gln Gly Asp Leu Ile Thr His Val Asn Gly Glu Pro Val His Gly Leu
        1155                1160                1165

Val His Thr Glu Val Val Glu Leu Ile Leu Lys Ser Gly Asn Lys Val
    1170                1175                1180

Ala Ile Ser Thr Thr Pro Leu Glu Asn Thr Ser Ile Lys Val Gly Pro
1185                1190                1195                1200

Ala Arg Lys Gly Ser Tyr Lys Ala Lys Met Ala Arg Arg Ser Lys Arg
                1205                1210                1215

Ser Arg Gly Lys Asp Gly Gln Glu Ser Arg Lys Arg Ser Ser Leu Phe
            1220                1225                1230

Arg Lys Ile Thr Lys Gln Ala Ser Leu Leu His Thr Ser Arg Ser Leu
        1235                1240                1245

Ser Ser Leu Asn Arg Ser Leu Ser Ser Gly Glu Ser Gly Pro Gly Ser
    1250                1255                1260

Pro Thr His Ser His Ser Leu Ser Pro Arg Ser Pro Thr Gln Gly Tyr
1265                1270                1275                1280

Arg Val Thr Pro Asp Ala Val His Ser Val Gly Gly Asn Ser Ser Gln
                1285                1290                1295

Ser Ser Ser Pro Ser Ser Ser Val Pro Ser Ser Pro Ala Gly Ser Gly
            1300                1305                1310

His Thr Arg Pro Ser Ser Leu His Gly Leu Ala Pro Lys Leu Gln Arg
        1315                1320                1325

Gln Tyr Arg Ser Pro Arg Arg Lys Ser Ala Gly Ser Ile Pro Leu Ser
    1330                1335                1340

Pro Leu Ala His Thr Pro Ser Pro Pro Pro Pro Thr Ala Ser Pro Gln
1345                1350                1355                1360

Arg Ser Pro Ser Pro Leu Ser Gly His Val Ala Gln Ala Phe Pro Thr
                1365                1370                1375

Lys Leu His Leu Ser Pro Pro Leu Gly Arg Gln Leu Ser Arg Pro Lys
            1380                1385                1390

Ser Ala Glu Pro Pro Arg Ser Pro Leu Leu Lys Arg Val Gln Ser Ala
        1395                1400                1405

Glu Lys Leu Ala Ala Ala Leu Ala Ala Ser Glu Lys Lys Leu Ala Thr
    1410                1415                1420
```

-continued

```
Ser Arg Lys His Ser Leu Asp Leu Pro His Ser Glu Leu Lys Lys Glu
1425                1430                1435                1440

Leu Pro Pro Arg Glu Val Ser Pro Leu Glu Val Val Gly Ala Arg Ser
                1445                1450                1455

Val Leu Ser Gly Lys Gly Ala Leu Pro Gly Lys Gly Val Leu Gln Pro
            1460                1465                1470

Ala Pro Ser Arg Ala Leu Gly Thr Leu Arg Gln Asp Arg Ala Glu Arg
        1475                1480                1485

Arg Glu Ser Leu Gln Lys Gln Glu Ala Ile Arg Glu Val Asp Ser Ser
    1490                1495                1500

Glu Asp Asp Thr Glu Glu Gly Pro Glu Asn Ser Gln Gly Ala Gln Glu
1505                1510                1515                1520

Leu Ser Leu Ala Pro His Pro Glu Val Ser Gln Ser Val Ala Pro Lys
                1525                1530                1535

Gly Ala Gly Glu Ser Gly Glu Glu Asp Pro Phe Pro Ser Arg Gly Pro
            1540                1545                1550

Arg Ser Leu Gly Pro Met Val Pro Ser Leu Leu Thr Gly Ile Thr Leu
        1555                1560                1565

Gly Pro Pro Arg Met Glu Ser Pro Ser Gly Pro His Arg Arg Leu Gly
    1570                1575                1580

Ser Pro Gln Ala Ile Glu Glu Ala Ala Ser Ser Ser Ser Ala Gly Pro
1585                1590                1595                1600

Asn Leu Gly Gln Ser Gly Ala Thr Asp Pro Ile Pro Pro Glu Gly Cys
                1605                1610                1615

Trp Lys Ala Gln His Leu His Thr Gln Ala Leu Thr Ala Leu Ser Pro
            1620                1625                1630

Ser Thr Ser Gly Leu Thr Pro Thr Ser Ser Cys Ser Pro Pro Ser Ser
        1635                1640                1645

Thr Ser Gly Lys Leu Ser Met Trp Ser Trp Lys Ser Leu Ile Glu Gly
    1650                1655                1660

Pro Asp Arg Ala Ser Pro Ser Arg Lys Ala Thr Met Ala Gly Gly Leu
1665                1670                1675                1680

Ala Asn Leu Gln Asp Leu Glu Thr Gln Leu Gln Pro Ser Leu Arg Thr
                1685                1690                1695

Cys Leu Pro Gly Ser Arg Gly Arg His Ser His Leu Val Pro Pro Asp
            1700                1705                1710

Trp Pro Ile His Leu Met Arg Ile Pro Ala Arg Ala Gly Tyr Gly Ser
        1715                1720                1725

Leu Ser Val His Lys Gln
    1730


<210> SEQ ID NO 31
<211> LENGTH: 5374
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse microtubule associated testis specific
      serine/threonine protein kinase (Mtssk, MAST205)
      cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (117)..(5321)
<223> OTHER INFORMATION: MAST205

<400> SEQUENCE: 31
```

-continued

```
ggacggccga gttgccccag ccgcagtctt tgcccccgcg ccggcgagcg ccgcccggga     60
ggcagctact ggaggagcgg agcgggcctt tggggcacga cagcagggag caggatatgg    120
ttactggact tagtcctctg ctcttcagga agcttagtaa tcctgacata tttgcaccca    180
ctggaaaagt taaactccag cgacaactta gtcaggatga ctgtaagtta cggagaggaa    240
gcctggcaag ttctctgtcg ggtaagcagc tgctcccttt gtccagcagt gtacacagca    300
gtgtgggaca ggtaacttgg cagtctacag gagaagcatc aaacctggtt cgaatgagaa    360
accaatccct tggacagtct gcaccttccc ttacggctgg cttgaaggaa ttgagccttc    420
caaggagagg cagcttttgt cggacaagta accgcaagag cttgattgta acctccagca    480
catcacctac gctaccaagg ccacactccc cactccatgg ccacacaggt aacagtccct    540
tggacagccc ccggaatttc tctccaaatg ctcctgctca cttttccttc gttcctgccc    600
gtaggactga tggacggcgc tggtctttgg cctcgttgcc ttcttcaggc tatggaacca    660
acactcctag ctctacagtc tcttcatcat gttcctcaca agaaaagctt catcagttac    720
ctttccagcc aacagctgat gaactacact ttctgacgaa gcattttagc acagaaaacg    780
taccagatga ggagggacgt cggtccccac ggatgcggcc ccgttcccgc agcctcagtc    840
ctggacggtc cccagtttcc tttgacagtg aaataataat gatgaatcat gtgtacaaag    900
aaagattccc caaggccact gcacagatgg aagagcgacc tagcctgacc ttcatttcct    960
ctaacactcc agatagtgtg ttgcccttgg cagatggagc actaagcttt attcatcatc   1020
aggtgattga gatggcccga gactgcctgg ataaatctcg gagtggcctc attacgtcac   1080
actatttcta tgaacttcaa gagaatttgg aaaagcttct gcaagatgct cacgaacgct   1140
cagaaagctc agatgtagcc tttgtgatac agctggtaaa aaagttgatg atcatcattg   1200
ctcgcccagc tcgcctcctg gaatgcctgg agtttgaccc tgaagaattt taccacctgt   1260
tagaagcagc tgaaggtcat gccaaagagg gacatggaat taaatgtgac attccccgct   1320
acatcgttag ccagctgggc ctaactcggg atcccttgga ggaaatggcc cagttgagca   1380
gctatgacag tccagatact ccagagacag atgattcagt tgagggtcgt ggggtatctc   1440
agccatctca gaagaccccc tctgaagagg actttgaaac cattaagctc atcagcaatg   1500
gcgcctatgg ggctgtcttt ctggtgcggc acaagtccac gcggcagcgc tttgcaatga   1560
agaagattaa taagcaaaac ctaatcctac ggaaccagat ccagcaagca tttgtggaac   1620
gcgacatact gactttcgct gaaaacccct ttgtggtcag catgttctgc tcctttgaga   1680
ccaagcgtca cttatgcatg gtgatggaat acgtagaagg gggagactgt gccactctgc   1740
tcaagaacat cggggcccta cctgtggaca tggtacgcct gtactttgcg gaaactgtgc   1800
tggctttgga atacttacac aactacggca tcgtgcaccg tgacctcaag cctgacaacc   1860
ttctgattac atccatggga cacatcaaac tcactgactt tggactttct aaaatcggcc   1920
tcatgagttt gacaaccaac ttgtatgagg gtcatattga aaaggatgcc cgggagttcc   1980
tagacaagca ggtatgcggg accccggaat acatagcacc tgaggtgatc ctgcgtcagg   2040
gatatgggaa gccagtggac tggtgggcca tgggcataat cctgtatgag ttcctggtgg   2100
gttgcgtccc tttctttgga gacactccgg aggagctctt tgggcaagtg atcagtgatg   2160
agatcgtgtg gcccgaaggt gatgacgcgc ttcccccaga tgcccaagac ctcacttcca   2220
aactgcttca tcagaatcca ctagaaagac tgggcacaag tagtgcctat gaagtgaagc   2280
```

-continued

```
agcacccatt cttcatgggt ctggactgga caggacttct acggcagaag gctgaattta   2340
tacctcagct ggagtcagag gatgacacca gctactttga tacccgttca gaacgatacc   2400
accatgtgga ctctgaggat gaggaggaag tgagtgagga tggctgcctt gagattcgcc   2460
agttttcttc ctgctctcca aggttcagca aggtttacag tagcatggaa cggctttccc   2520
tgcttgagga acgccggaca ccacctccaa ccaagcgcag cctcagtgag gaaaaggaag   2580
atcactcaga cggcttggca ggactgaagg gccgagatcg cagctgggtg attgggtccc   2640
ctgagatatt acggaagcgg ttatctgtgt ccgagtcatc ccacacagag agtgattcga   2700
gtcctccaat gacagtgcga catcggtgtt caggcctccc agatgggcct cattgccctg   2760
aggagactag tagcacccct cggaagcaac aacaggaagg tatatgggtc ctcattcccc   2820
catctggaga ggggtcatct aggcctgttc ctgaacgacc gttggagagg caactgaagc   2880
tggatgagga gcctcctggc caaagcagtc ggtgttgccc agccctggag acacgaggcc   2940
gtgggacccc tcagctagct gaggaagcta cagccaaagc catcagcgac ctagctgtgc   3000
gtagggcccg tcaccggctg ctctctgggg actctataga gaagcgcacc actcgccctg   3060
tcaacaaagt aatcaagtca gcctcagcta cggccctttc cctcctcatt ccttcagaac   3120
accatgcctg ctcaccattg gccagcccta tgtccccaca ttcccagtca tccaatccat   3180
catccaggga ctcttctcca agcagggact tcttgccagc ccttggcagc ttgaggcctc   3240
ccatcatcat ccaccgagct ggcaagaagt atggcttcac cctgcgggcc attcgagtct   3300
acatgggtga cactgatgtc tacaccgtac accacatggt gtggcatgtg gaggatggtg   3360
gtccagccag tgaagcaggg cttcgtcagg gtgacctcat cacccatgtc aatggtgagc   3420
ctgtgcatgg gctagtccac acagaagtgg tggagctggt tctgaagagt ggaaacaagg   3480
tatcaatttc aacaactccc ttggagaaca cgtcaatcaa agtggggcca gctaggaaag   3540
gcagctataa agccaagatg gcccgaagga gcaaacggag caaaggcaag gatgggcaag   3600
aaagccgaaa aagaagctcc ctattccgga aaatcacaaa gcaggcctcc ttgctccaca   3660
ccagccgcag cctttcttcc cttaaccgct ccttgtcatc aggggagagt ggtccaggct   3720
ctcccacaca cagccacagc ctctctccca gatctcctcc tcagggatac cgggtagccc   3780
cagatgctgt gcactcagta ggagggaatt cctcgcagag cagctctccc agctccagtg   3840
tgcccagttc tcctgctggc tctggacata cacggcccag ctctcttcac ggtctggcac   3900
ccaagctcca acgccaatac cgctcaccac ggcgcaagtc agcaggcagc atcccactgt   3960
caccgttggc ccacacccct tccccaccag caacggcagc ttcacctcag cgttccccat   4020
cacccttgtc tggccatggg tctcagtcct ttcctaccaa acttcacttg tctcctccgc   4080
taggtaggca gctctcacgg cccaagagtg cagagccacc ccgctctccc ctacttaaga   4140
gggtgcagtc tgctgagaag ctggcggctg cactggcagc tgctgagaag aagttagcac   4200
cttcccgcaa acatagtctt gacctgcccc atggtgaact aaagaaggaa ctgacaccca   4260
gggaagccag ccctctggag gtagttggaa ccagaagtgt gctatccggg aaagggccac   4320
ttccaggaaa gggggtactg cagcctgctc cttcacgggc ccttgggacc ctacggcagg   4380
atcgagctga acgccgtgag tcactgcaaa aacaagaagc aatccgggaa gtagactcct   4440
cagaagatga cactgatgag gagcctgaga acagccaggc cacacaggag ccaagattgt   4500
ccccccaccc agaagcaagc cacaatctac tccctaaagg ttcaggagag ggtacagaag   4560
```

-continued

```
aggacacttt cttgcacagg gatctaaaga agcagggccc tgtactctca ggtctagtga   4620

caggggccac actaggctcc ccccgagtag acgttcctgg gctctcccca aggaaggtca   4680

gcaggccaca agcctttgag gaagctacca acccсttaca agtccctagc ctgagcaggt   4740

ctggacccac aagccccacc ccctctgaag gctgctggaa ggcccagcac ctccacacac   4800

aggcactaac tgcactttgt cccagctttt cagaacttac ccctaccggt tgttctgctg   4860

ccacctccac ctctggaaag ccagggacat ggtcctggaa attccttatt gagggtccag   4920

acagagcatc cacgaacaag accataacaa ggaaaggtga accagctaac tcccaagata   4980

cgaataccac ggtcccaaat cttctgaaga acctgtctcc tgaggaggag aagccacagc   5040

caccaagtgt gcctgggctg acccatccgc ttcttgaggt ccccagccag aactggccat   5100

gggagtctga atgtgaacaa atggagaaag aagaaccatc cctgagcatc accgaagtgc   5160

ctgattcctc aggcgacagg aggcaggaca ttccatgcag agcccacccc ctgagcccag   5220

aaacccggcc cagcctgctc tggaaaagcc aagaacttgg gggccagcaa gatcatcagg   5280

acttagcact gacatcagat gagctcttaa agcaaaccta gcacttgttt gcttccctta   5340

cattcacctg tgtaatacac cctcctggaa acca                               5374


<210> SEQ ID NO 32
<211> LENGTH: 1734
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse microtubule associated testis specific
      serine/threonine protein kinase (Mtssk, MAST205)

<400> SEQUENCE: 32

Met Val Thr Gly Leu Ser Pro Leu Leu Phe Arg Lys Leu Ser Asn Pro
1               5                   10                  15

Asp Ile Phe Ala Pro Thr Gly Lys Val Lys Leu Gln Arg Gln Leu Ser
            20                  25                  30

Gln Asp Asp Cys Lys Leu Arg Arg Gly Ser Leu Ala Ser Ser Leu Ser
        35                  40                  45

Gly Lys Gln Leu Leu Pro Leu Ser Ser Ser Val His Ser Ser Val Gly
    50                  55                  60

Gln Val Thr Trp Gln Ser Thr Gly Glu Ala Ser Asn Leu Val Arg Met
65                  70                  75                  80

Arg Asn Gln Ser Leu Gly Gln Ser Ala Pro Ser Leu Thr Ala Gly Leu
                85                  90                  95

Lys Glu Leu Ser Leu Pro Arg Arg Gly Ser Phe Cys Arg Thr Ser Asn
            100                 105                 110

Arg Lys Ser Leu Ile Val Thr Ser Ser Thr Ser Pro Thr Leu Pro Arg
        115                 120                 125

Pro His Ser Pro Leu His Gly His Thr Gly Asn Ser Pro Leu Asp Ser
    130                 135                 140

Pro Arg Asn Phe Ser Pro Asn Ala Pro Ala His Phe Ser Phe Val Pro
145                 150                 155                 160

Ala Arg Arg Thr Asp Gly Arg Arg Trp Ser Leu Ala Ser Leu Pro Ser
                165                 170                 175

Ser Gly Tyr Gly Thr Asn Thr Pro Ser Ser Thr Val Ser Ser Ser Cys
            180                 185                 190
```

-continued

```
Ser Ser Gln Glu Lys Leu His Gln Leu Pro Phe Gln Pro Thr Ala Asp
        195                 200                 205

Glu Leu His Phe Leu Thr Lys His Phe Ser Thr Glu Asn Val Pro Asp
    210                 215                 220

Glu Glu Gly Arg Arg Ser Pro Arg Met Arg Pro Arg Ser Arg Ser Leu
225                 230                 235                 240

Ser Pro Gly Arg Ser Pro Val Ser Phe Asp Ser Glu Ile Ile Met Met
                245                 250                 255

Asn His Val Tyr Lys Glu Arg Phe Pro Lys Ala Thr Ala Gln Met Glu
            260                 265                 270

Glu Arg Pro Ser Leu Thr Phe Ile Ser Ser Asn Thr Pro Asp Ser Val
        275                 280                 285

Leu Pro Leu Ala Asp Gly Ala Leu Ser Phe Ile His His Gln Val Ile
    290                 295                 300

Glu Met Ala Arg Asp Cys Leu Asp Lys Ser Arg Ser Gly Leu Ile Thr
305                 310                 315                 320

Ser His Tyr Phe Tyr Glu Leu Gln Glu Asn Leu Glu Lys Leu Leu Gln
                325                 330                 335

Asp Ala His Glu Arg Ser Glu Ser Ser Asp Val Ala Phe Val Ile Gln
            340                 345                 350

Leu Val Lys Lys Leu Met Ile Ile Ile Ala Arg Pro Ala Arg Leu Leu
        355                 360                 365

Glu Cys Leu Glu Phe Asp Pro Glu Glu Phe Tyr His Leu Leu Glu Ala
    370                 375                 380

Ala Glu Gly His Ala Lys Glu Gly His Gly Ile Lys Cys Asp Ile Pro
385                 390                 395                 400

Arg Tyr Ile Val Ser Gln Leu Gly Leu Thr Arg Asp Pro Leu Glu Glu
                405                 410                 415

Met Ala Gln Leu Ser Ser Tyr Asp Ser Pro Asp Thr Pro Glu Thr Asp
            420                 425                 430

Asp Ser Val Glu Gly Arg Gly Val Ser Gln Pro Ser Gln Lys Thr Pro
        435                 440                 445

Ser Glu Glu Asp Phe Glu Thr Ile Lys Leu Ile Ser Asn Gly Ala Tyr
    450                 455                 460

Gly Ala Val Phe Leu Val Arg His Lys Ser Thr Arg Gln Arg Phe Ala
465                 470                 475                 480

Met Lys Lys Ile Asn Lys Gln Asn Leu Ile Leu Arg Asn Gln Ile Gln
                485                 490                 495

Gln Ala Phe Val Glu Arg Asp Ile Leu Thr Phe Ala Glu Asn Pro Phe
            500                 505                 510

Val Val Ser Met Phe Cys Ser Phe Glu Thr Lys Arg His Leu Cys Met
        515                 520                 525

Val Met Glu Tyr Val Glu Gly Gly Asp Cys Ala Thr Leu Leu Lys Asn
    530                 535                 540

Ile Gly Ala Leu Pro Val Asp Met Val Arg Leu Tyr Phe Ala Glu Thr
545                 550                 555                 560

Val Leu Ala Leu Glu Tyr Leu His Asn Tyr Gly Ile Val His Arg Asp
                565                 570                 575

Leu Lys Pro Asp Asn Leu Leu Ile Thr Ser Met Gly His Ile Lys Leu
            580                 585                 590

Thr Asp Phe Gly Leu Ser Lys Ile Gly Leu Met Ser Leu Thr Thr Asn
```

-continued

```
        595                 600                 605

Leu Tyr Glu Gly His Ile Glu Lys Asp Ala Arg Glu Phe Leu Asp Lys
    610                 615                 620

Gln Val Cys Gly Thr Pro Glu Tyr Ile Ala Pro Glu Val Ile Leu Arg
625                 630                 635                 640

Gln Gly Tyr Gly Lys Pro Val Asp Trp Trp Ala Met Gly Ile Ile Leu
                645                 650                 655

Tyr Glu Phe Leu Val Gly Cys Val Pro Phe Phe Gly Asp Thr Pro Glu
            660                 665                 670

Glu Leu Phe Gly Gln Val Ile Ser Asp Glu Ile Val Trp Pro Glu Gly
        675                 680                 685

Asp Asp Ala Leu Pro Pro Asp Ala Gln Asp Leu Thr Ser Lys Leu Leu
    690                 695                 700

His Gln Asn Pro Leu Glu Arg Leu Gly Thr Ser Ser Ala Tyr Glu Val
705                 710                 715                 720

Lys Gln His Pro Phe Phe Met Gly Leu Asp Trp Thr Gly Leu Leu Arg
                725                 730                 735

Gln Lys Ala Glu Phe Ile Pro Gln Leu Glu Ser Glu Asp Asp Thr Ser
            740                 745                 750

Tyr Phe Asp Thr Arg Ser Glu Arg Tyr His His Val Asp Ser Glu Asp
        755                 760                 765

Glu Glu Glu Val Ser Glu Asp Gly Cys Leu Glu Ile Arg Gln Phe Ser
    770                 775                 780

Ser Cys Ser Pro Arg Phe Ser Lys Val Tyr Ser Ser Met Glu Arg Leu
785                 790                 795                 800

Ser Leu Leu Glu Glu Arg Arg Thr Pro Pro Pro Thr Lys Arg Ser Leu
                805                 810                 815

Ser Glu Glu Lys Glu Asp His Ser Asp Gly Leu Ala Gly Leu Lys Gly
            820                 825                 830

Arg Asp Arg Ser Trp Val Ile Gly Ser Pro Glu Ile Leu Arg Lys Arg
        835                 840                 845

Leu Ser Val Ser Glu Ser Ser His Thr Glu Ser Asp Ser Ser Pro Pro
    850                 855                 860

Met Thr Val Arg His Arg Cys Ser Gly Leu Pro Asp Gly Pro His Cys
865                 870                 875                 880

Pro Glu Glu Thr Ser Ser Thr Pro Arg Lys Gln Gln Gln Glu Gly Ile
                885                 890                 895

Trp Val Leu Ile Pro Pro Ser Gly Glu Gly Ser Ser Arg Pro Val Pro
            900                 905                 910

Glu Arg Pro Leu Glu Arg Gln Leu Lys Leu Asp Glu Glu Pro Pro Gly
        915                 920                 925

Gln Ser Ser Arg Cys Cys Pro Ala Leu Glu Thr Arg Gly Arg Gly Thr
    930                 935                 940

Pro Gln Leu Ala Glu Glu Ala Thr Ala Lys Ala Ile Ser Asp Leu Ala
945                 950                 955                 960

Val Arg Arg Ala Arg His Arg Leu Leu Ser Gly Asp Ser Ile Glu Lys
                965                 970                 975

Arg Thr Thr Arg Pro Val Asn Lys Val Ile Lys Ser Ala Ser Ala Thr
            980                 985                 990

Ala Leu Ser Leu Leu Ile Pro Ser Glu His His Ala Cys Ser Pro Leu
        995                 1000                1005
```

-continued

```
Ala Ser Pro Met Ser Pro His Ser Gln Ser Ser Asn Pro Ser Ser Arg
   1010                1015                1020

Asp Ser Ser Pro Ser Arg Asp Phe Leu Pro Ala Leu Gly Ser Leu Arg
1025                1030                1035                1040

Pro Pro Ile Ile Ile His Arg Ala Gly Lys Lys Tyr Gly Phe Thr Leu
               1045                1050                1055

Arg Ala Ile Arg Val Tyr Met Gly Asp Thr Asp Val Tyr Thr Val His
           1060                1065                1070

His Met Val Trp His Val Glu Asp Gly Gly Pro Ala Ser Glu Ala Gly
       1075                1080                1085

Leu Arg Gln Gly Asp Leu Ile Thr His Val Asn Gly Glu Pro Val His
   1090                1095                1100

Gly Leu Val His Thr Glu Val Val Glu Leu Val Leu Lys Ser Gly Asn
1105                1110                1115                1120

Lys Val Ser Ile Ser Thr Thr Pro Leu Glu Asn Thr Ser Ile Lys Val
               1125                1130                1135

Gly Pro Ala Arg Lys Gly Ser Tyr Lys Ala Lys Met Ala Arg Arg Ser
           1140                1145                1150

Lys Arg Ser Lys Gly Lys Asp Gly Gln Glu Ser Arg Lys Arg Ser Ser
       1155                1160                1165

Leu Phe Arg Lys Ile Thr Lys Gln Ala Ser Leu Leu His Thr Ser Arg
   1170                1175                1180

Ser Leu Ser Ser Leu Asn Arg Ser Leu Ser Ser Gly Glu Ser Gly Pro
1185                1190                1195                1200

Gly Ser Pro Thr His Ser His Ser Leu Ser Pro Arg Ser Pro Pro Gln
               1205                1210                1215

Gly Tyr Arg Val Ala Pro Asp Ala Val His Ser Val Gly Gly Asn Ser
           1220                1225                1230

Ser Gln Ser Ser Ser Pro Ser Ser Ser Val Pro Ser Ser Pro Ala Gly
       1235                1240                1245

Ser Gly His Thr Arg Pro Ser Ser Leu His Gly Leu Ala Pro Lys Leu
   1250                1255                1260

Gln Arg Gln Tyr Arg Ser Pro Arg Arg Lys Ser Ala Gly Ser Ile Pro
1265                1270                1275                1280

Leu Ser Pro Leu Ala His Thr Pro Ser Pro Pro Ala Thr Ala Ala Ser
               1285                1290                1295

Pro Gln Arg Ser Pro Ser Pro Leu Ser Gly His Gly Ser Gln Ser Phe
           1300                1305                1310

Pro Thr Lys Leu His Leu Ser Pro Pro Leu Gly Arg Gln Leu Ser Arg
       1315                1320                1325

Pro Lys Ser Ala Glu Pro Pro Arg Ser Pro Leu Leu Lys Arg Val Gln
   1330                1335                1340

Ser Ala Glu Lys Leu Ala Ala Ala Leu Ala Ala Ala Glu Lys Lys Leu
1345                1350                1355                1360

Ala Pro Ser Arg Lys His Ser Leu Asp Leu Pro His Gly Glu Leu Lys
               1365                1370                1375

Lys Glu Leu Thr Pro Arg Glu Ala Ser Pro Leu Glu Val Val Gly Thr
           1380                1385                1390

Arg Ser Val Leu Ser Gly Lys Gly Pro Leu Pro Gly Lys Gly Val Leu
       1395                1400                1405
```

-continued

```
Gln Pro Ala Pro Ser Arg Ala Leu Gly Thr Leu Arg Gln Asp Arg Ala
   1410                1415                1420

Glu Arg Arg Glu Ser Leu Gln Lys Gln Glu Ala Ile Arg Glu Val Asp
1425                1430                1435                1440

Ser Ser Glu Asp Asp Thr Asp Glu Glu Pro Glu Asn Ser Gln Ala Thr
               1445                1450                1455

Gln Glu Pro Arg Leu Ser Pro His Pro Glu Ala Ser His Asn Leu Leu
           1460                1465                1470

Pro Lys Gly Ser Gly Glu Gly Thr Glu Glu Asp Thr Phe Leu His Arg
       1475                1480                1485

Asp Leu Lys Lys Gln Gly Pro Val Leu Ser Gly Leu Val Thr Gly Ala
   1490                1495                1500

Thr Leu Gly Ser Pro Arg Val Asp Val Pro Gly Leu Ser Pro Arg Lys
1505                1510                1515                1520

Val Ser Arg Pro Gln Ala Phe Glu Glu Ala Thr Asn Pro Leu Gln Val
               1525                1530                1535

Pro Ser Leu Ser Arg Ser Gly Pro Thr Ser Pro Thr Pro Ser Glu Gly
           1540                1545                1550

Cys Trp Lys Ala Gln His Leu His Thr Gln Ala Leu Thr Ala Leu Cys
       1555                1560                1565

Pro Ser Phe Ser Glu Leu Thr Pro Thr Gly Cys Ser Ala Ala Thr Ser
   1570                1575                1580

Thr Ser Gly Lys Pro Gly Thr Trp Ser Trp Lys Phe Leu Ile Glu Gly
1585                1590                1595                1600

Pro Asp Arg Ala Ser Thr Asn Lys Thr Ile Thr Arg Lys Gly Glu Pro
               1605                1610                1615

Ala Asn Ser Gln Asp Thr Asn Thr Thr Val Pro Asn Leu Leu Lys Asn
           1620                1625                1630

Leu Ser Pro Glu Glu Glu Lys Pro Gln Pro Pro Ser Val Pro Gly Leu
       1635                1640                1645

Thr His Pro Leu Leu Glu Val Pro Ser Gln Asn Trp Pro Trp Glu Ser
   1650                1655                1660

Glu Cys Glu Gln Met Glu Lys Glu Glu Pro Ser Leu Ser Ile Thr Glu
1665                1670                1675                1680

Val Pro Asp Ser Ser Gly Asp Arg Arg Gln Asp Ile Pro Cys Arg Ala
               1685                1690                1695

His Pro Leu Ser Pro Glu Thr Arg Pro Ser Leu Leu Trp Lys Ser Gln
           1700                1705                1710

Glu Leu Gly Gly Gln Gln Asp His Gln Asp Leu Ala Leu Thr Ser Asp
       1715                1720                1725

Glu Leu Leu Lys Gln Thr
   1730
```

<210> SEQ ID NO 33
<211> LENGTH: 3568
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human colon Kruppel-like factor (CKLF) cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (537)..(1910)
<223> OTHER INFORMATION: CKLF

<400> SEQUENCE: 33

-continued

```
gcggccgccc tgcgcgcgaa gctcgtggcc cgagaggggt gcggtcgggc cgacggaggc     60
ggggccctgg ctgcctctct ccctgctcat aggctggccg ctcaggcctg gccggcctcg    120
gggcctcggg attcgcggcg gcgctgccaa tcaggcgatc gggccccgcc cccccggagt    180
tgggtgaaat agaggcgggc gtcaagtgtc agtagtcgcg gggcaggtac gtgcgctcgc    240
ggttctctcg cggaggtcgg cggtggcggg agcgggctcc ggagagcctg agagcacggt    300
ggggcggggc gggagaaagt ggccgcccgg aggacgttgg cgtttacgtg tggaagagcg    360
gaagagtttt gcttttcgtg cgcgccttcg aaaactgcct gccgctgtct gaggagtcca    420
cccgaaacct cccctcctcc gccggcagcc ccgcgctgag ctcgccgacc caagccagcg    480
tgggcgaggt gggaagtgcg cccgacccgc gcctggagct gcgcccccga gtgcccatgg    540
ctacaagggt gctgagcatg agcgcccgcc tgggacccgt gccccagccg ccggcgccgc    600
aggacgagcc ggtgttcgcg cagctcaagc cggtgctggg cgccgcgaat ccggcccgcg    660
acgcggcgct cttccccggc gaggagctga agcacgcgca ccaccgcccg caggcgcagc    720
ccgcgcccgc gcaggccccg cagccggccc agccgcccgc caccggcccg cggctgcctc    780
cagaggacct ggtccaaaca agatgtgaaa tggagaagta tctgacacct cagcttcctc    840
cagttcctat aattccagag cataaaaagt atagacgaga cagtgcctca gtcgtagacc    900
agttcttcac tgacactgaa gggttacctt acagtatcaa catgaacgtc ttcctccctg    960
acatcactca cctgagaact ggcctctaca aatcccagag accgtgcgta acacacatca   1020
agacagaacc tgttgccatt ttcagccacc agagtgaaac gactgcccct cctccggccc   1080
cgacccaggc cctccctgag ttcaccagta tattcagctc acaccagacc gcagctccag   1140
aggtgaacaa tattttcatc aaacaagaac ttcctacacc agatcttcat ctttctgtcc   1200
ctacccagca gggccacctg taccagctac tgaatacacc ggatctagat atgcccagtt   1260
ctacaaatca gacagcagca atggacactc ttaatgtttc tatgtcagct gccatggcag   1320
gccttaacac acacacctct gctgttccgc agactgcagt gaaacaattc cagggcatgc   1380
cccccttgcac atacacaatg ccaagtcagt ttcttccaca acaggccact tactttcccc   1440
cgtcaccacc aagctcagag cctggaagtc cagatagaca agcagagatg ctccagaatt   1500
taaccccacc tccatcctat gctgctacaa ttgcttctaa actggcaatt cacaatccaa   1560
atttacccac caccctgcca gttaactcac aaaacatcca acctgtcaga tacaatagaa   1620
ggagtaaccc cgatttggag aaacgacgca tccactactg cgattaccct ggttgcacaa   1680
aagtttatac caagtcttct catttaaaag ctcacctgag gactcacact ggtgaaaagc   1740
catacaagtg tacctgggaa ggctgcgact ggaggttcgc gcgatcggat gagctgaccc   1800
gccactaccg gaagcacaca ggcgccaagc ccttccagtg cggggtgtgc aaccgcagct   1860
tctcgcgctc tgaccacctg gccctgcata tgaagaggca ccagaactga gcactgcccg   1920
tgtgacccgt tccaggtccc ctgggctccc tcaaatgaca gacctaacta ttcctgtgta   1980
aaaacaacaa aaaacaaaca aaagcaagaa aaccacaact aaaactggaa atgtatattt   2040
tgtatatttg agaaaacagg gaatacattg tattaatacc aaagtgtttg gtcattttaa   2100
gaatctggaa tgcttgctgt aatgtatatg gctttactca agcagatctc atctcatgac   2160
aggcagccac gtctcaacat gggtaagggg tgggggtgga ggggagtgtg tgcagcgttt   2220
ttacctaggc accatcattt aatgtgacag tgttcagtaa acaaatcagt tggcaggcac   2280
```

-continued

```
cagaagaaga atggattgta tgtcaagatt ttacttggca ttgagtagtt tttttcaata   2340

gtaggtaatt ccttagagat acagtatacc tggcaattca caaatagcca ttgaacaaat   2400

gtgtgggttt ttaaaaatta tatactatat gagttgccta tatttgctat tcaaaatttt   2460

gtaaatatgc aaatcagctt tataggttta ttacaagttt tttaggattc ttttggggaa   2520

gagtcataat tcttttgaaa ataaccatga atacacttac agttaggatt tgtggtaagg   2580

tacctctcaa cattaccaaa atcatttctt tagagggaag gaataatcat tcaaatgaac   2640

tttaaaaaag caaatttcat gcactgatta aaataggatt attttaaata caaaaggcat   2700

tttatatgaa ttataaactg aagagcttaa agatagttac aaatacaaaa gttcaacctc   2760

ttacaataag ctaaacgcaa tgtcttttta aaaagaggac ttagggtgtc gtttttcaca   2820

tatgacaatg ttgcatttat gatgcagttt caagtaccaa aacgttgaat tgatgatgca   2880

gttttcatat atcgagatgt tcgctcgtgc agtactgttg gttaaatgac aatttatgtg   2940

gattttgcat gtaatacaca gtgagacaca gtaattttat ctaaattaca gtgcagttta   3000

gttaatctat taatactgac tcagtgtctg cctttaaata taaatgatat gttgaaaact   3060

taaggaagca aatgctacat atatgcaata taaaatagta atgtgatgct gatgctgtta   3120

accaaagggc agaataaata agcaaaatgc caaaaggggt cttaattgaa atgaaaattt   3180

aattttgttt ttaaaatatt gtttatcttt atttatttgg gggtaatatt gtaagttttt   3240

tagaagacaa ttttcataac ttgataaatt atagttttgt ttgttagaaa agtagctctt   3300

aaaagatgta aatagatgac aaacgatgta aataattttg taagaggctt caaaatgttt   3360

atacgtggaa acacacctac atgaaaagca gaaatcggtt gctgttttgc ttctttttcc   3420

ctcttatttt tgtattgtgg tcatttccta tgcaaataat ggagcaaaca gctgtatagt   3480

tgtagaattt tttgagagaa tgagatgttt atatattaac gacaattttt ttttggaaaa   3540

taaaaagtgc cctaaaagaa aaaaaaaa                                      3568
```

<210> SEQ ID NO 34
<211> LENGTH: 457
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human colon Kruppel-like factor (CKLF)

<400> SEQUENCE: 34

```
Met Ala Thr Arg Val Leu Ser Met Ser Ala Arg Leu Gly Pro Val Pro
  1               5                  10                  15

Gln Pro Pro Ala Pro Gln Asp Glu Pro Val Phe Ala Gln Leu Lys Pro
             20                  25                  30

Val Leu Gly Ala Ala Asn Pro Ala Arg Asp Ala Ala Leu Phe Pro Gly
         35                  40                  45

Glu Glu Leu Lys His Ala His His Arg Pro Gln Ala Gln Pro Ala Pro
     50                  55                  60

Ala Gln Ala Pro Gln Pro Ala Gln Pro Pro Ala Thr Gly Pro Arg Leu
 65                  70                  75                  80

Pro Pro Glu Asp Leu Val Gln Thr Arg Cys Glu Met Glu Lys Tyr Leu
                 85                  90                  95

Thr Pro Gln Leu Pro Pro Val Pro Ile Ile Pro Glu His Lys Lys Tyr
            100                 105                 110
```

-continued

```
Arg Arg Asp Ser Ala Ser Val Val Asp Gln Phe Phe Thr Asp Thr Glu
        115                 120                 125

Gly Leu Pro Tyr Ser Ile Asn Met Asn Val Phe Leu Pro Asp Ile Thr
    130                 135                 140

His Leu Arg Thr Gly Leu Tyr Lys Ser Gln Arg Pro Cys Val Thr His
145                 150                 155                 160

Ile Lys Thr Glu Pro Val Ala Ile Phe Ser His Gln Ser Glu Thr Thr
                165                 170                 175

Ala Pro Pro Pro Ala Pro Thr Gln Ala Leu Pro Glu Phe Thr Ser Ile
            180                 185                 190

Phe Ser Ser His Gln Thr Ala Ala Pro Glu Val Asn Asn Ile Phe Ile
        195                 200                 205

Lys Gln Glu Leu Pro Thr Pro Asp Leu His Leu Ser Val Pro Thr Gln
    210                 215                 220

Gln Gly His Leu Tyr Gln Leu Leu Asn Thr Pro Asp Leu Asp Met Pro
225                 230                 235                 240

Ser Ser Thr Asn Gln Thr Ala Ala Met Asp Thr Leu Asn Val Ser Met
                245                 250                 255

Ser Ala Ala Met Ala Gly Leu Asn Thr His Thr Ser Ala Val Pro Gln
            260                 265                 270

Thr Ala Val Lys Gln Phe Gln Gly Met Pro Pro Cys Thr Tyr Thr Met
        275                 280                 285

Pro Ser Gln Phe Leu Pro Gln Gln Ala Thr Tyr Phe Pro Pro Ser Pro
    290                 295                 300

Pro Ser Ser Glu Pro Gly Ser Pro Asp Arg Gln Ala Glu Met Leu Gln
305                 310                 315                 320

Asn Leu Thr Pro Pro Pro Ser Tyr Ala Ala Thr Ile Ala Ser Lys Leu
                325                 330                 335

Ala Ile His Asn Pro Asn Leu Pro Thr Thr Leu Pro Val Asn Ser Gln
            340                 345                 350

Asn Ile Gln Pro Val Arg Tyr Asn Arg Arg Ser Asn Pro Asp Leu Glu
        355                 360                 365

Lys Arg Arg Ile His Tyr Cys Asp Tyr Pro Gly Cys Thr Lys Val Tyr
    370                 375                 380

Thr Lys Ser Ser His Leu Lys Ala His Leu Arg Thr His Thr Gly Glu
385                 390                 395                 400

Lys Pro Tyr Lys Cys Thr Trp Glu Gly Cys Asp Trp Arg Phe Ala Arg
                405                 410                 415

Ser Asp Glu Leu Thr Arg His Tyr Arg Lys His Thr Gly Ala Lys Pro
            420                 425                 430

Phe Gln Cys Gly Val Cys Asn Arg Ser Phe Ser Arg Ser Asp His Leu
        435                 440                 445

Ala Leu His Met Lys Arg His Gln Asn
    450                 455
```

<210> SEQ ID NO 35
<211> LENGTH: 1591
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse intestinal-enriched Kruppel-like factor
      (IKLF, CKLF) cDNA
<220> FEATURE:
<221> NAME/KEY: CDS

-continued

```
<222> LOCATION: (167)..(1507)
<223> OTHER INFORMATION: CKLF

<400> SEQUENCE: 35

ccgagcccag gagccccgat ctccgtgccc gccttcgtga gcgtctggct gccggcccag     60

gggtcccccg ccgcggcccc ccgccgagtc cgccgtcccg tgccagcccg agcgaggtgg    120

gatcgcgatc gctccgtgtc ccgctcccgt aatccccaga ccgtccatgc ccacgcgggt    180

gctgaccatg agcgcccgcc tgggaccact gccccagccg ccggccgcgc aggccgagcc    240

cgtgttcgcg cagctcaagc cggtgctggg cgctgcgaac ccggcccgcg acgcggcgct    300

cttctccgga gacgatctga aacacgcgca ccaccacccg cctgcgccgc cgccagccgc    360

tggcccgcga ctgccctcgg aggagctggt ccagacaaga tgtgaaatgg agaagtatct    420

gacccctcag ctccctccag ttccgataat ttcagagcat aaaaagtata gacgagacag    480

tgcctcagtg gtagaccagt tcttcactga cactgaaggc ataccttaca gcatcaacat    540

gaacgtcttc ctccctgaca tcactcacct gagaactggc ctctacaaat cccagagacc    600

atgcgtaaca cagatcaaga cagaacctgt taccattttc agccaccaga gcgagtcgac    660

ggcccctcct cctcctccgg cccccaccca ggctctcccc gagttcacta gtatcttcag    720

ctcccaccag accacagcgc caccacagga ggtgaacaat atcttcatca aacaagaact    780

tcctatacca gatcttcatc tctctgtccc ttcccagcag ggccacctgt accagctgtt    840

gaatacaccg gatctagaca tgcccagttc gacaaaccag acggcagtaa tggacaccct    900

taatgtttct atggcaggcc ttaacccaca cccctctgct gttccacaga cgtcaatgaa    960

acagttccag ggcatgcccc cttgcacgta caccatgcca agtcagtttc ttccacagca   1020

ggccacttat tttccccgt caccaccaag ctcagagcct ggaagtcccg atagacaagc   1080

tgagatgctg cagaatctca ccccacctcc gtcctatgcc gctacaattg cttccaaact   1140

ggcgattcac aacccaaatt tacctgccac tctgccagtt aattcgccaa ctctcccacc   1200

tgtcagatac aacagaagga gtaacccgga tctggagaag cgacgtatcc acttctgcga   1260

ttataatggt tgcacaaaag tttatacaaa gtcgtctcac ttaaaagctc acctgaggac   1320

tcatacgggc gagaagccct acaagtgcac ctgggagggc tgcgactgga ggtttgcccg   1380

gtcggatgag ctgacccgcc actacaggaa gcacacgggc gccaagccgt tccagtgcat   1440

ggtgtgccaa cgcagcttct cccgctccga ccacctcgcg ctgcacatga agcgccacca   1500

gaactgagcg agcgaacgct gcgcccaccc gcctgacgcc ttgcagtccg ctttgccatc   1560

ctttaaaccg cagacctaac ttcataaaaa g                                   1591

<210> SEQ ID NO 36
<211> LENGTH: 446
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse intestinal-enriched Kruppel-like factor
      (IKLF, CKLF)

<400> SEQUENCE: 36

Met Pro Thr Arg Val Leu Thr Met Ser Ala Arg Leu Gly Pro Leu Pro
1               5                   10                  15

Gln Pro Pro Ala Ala Gln Ala Glu Pro Val Phe Ala Gln Leu Lys Pro
            20                  25                  30
```

-continued

```
Val Leu Gly Ala Ala Asn Pro Ala Arg Asp Ala Ala Leu Phe Ser Gly
        35                  40                  45

Asp Asp Leu Lys His Ala His His His Pro Pro Ala Pro Pro Pro Ala
    50                  55                  60

Ala Gly Pro Arg Leu Pro Ser Glu Glu Leu Val Gln Thr Arg Cys Glu
65                  70                  75                  80

Met Glu Lys Tyr Leu Thr Pro Gln Leu Pro Pro Val Pro Ile Ile Ser
                85                  90                  95

Glu His Lys Lys Tyr Arg Arg Asp Ser Ala Ser Val Val Asp Gln Phe
            100                 105                 110

Phe Thr Asp Thr Glu Gly Ile Pro Tyr Ser Ile Asn Met Asn Val Phe
        115                 120                 125

Leu Pro Asp Ile Thr His Leu Arg Thr Gly Leu Tyr Lys Ser Gln Arg
    130                 135                 140

Pro Cys Val Thr Gln Ile Lys Thr Glu Pro Val Thr Ile Phe Ser His
145                 150                 155                 160

Gln Ser Glu Ser Thr Ala Pro Pro Pro Pro Pro Ala Pro Thr Gln Ala
                165                 170                 175

Leu Pro Glu Phe Thr Ser Ile Phe Ser Ser His Gln Thr Thr Ala Pro
            180                 185                 190

Pro Gln Glu Val Asn Asn Ile Phe Ile Lys Gln Glu Leu Pro Ile Pro
        195                 200                 205

Asp Leu His Leu Ser Val Pro Ser Gln Gln Gly His Leu Tyr Gln Leu
    210                 215                 220

Leu Asn Thr Pro Asp Leu Asp Met Pro Ser Ser Thr Asn Gln Thr Ala
225                 230                 235                 240

Val Met Asp Thr Leu Asn Val Ser Met Ala Gly Leu Asn Pro His Pro
                245                 250                 255

Ser Ala Val Pro Gln Thr Ser Met Lys Gln Phe Gln Gly Met Pro Pro
            260                 265                 270

Cys Thr Tyr Thr Met Pro Ser Gln Phe Leu Pro Gln Gln Ala Thr Tyr
        275                 280                 285

Phe Pro Pro Ser Pro Pro Ser Ser Glu Pro Gly Ser Pro Asp Arg Gln
    290                 295                 300

Ala Glu Met Leu Gln Asn Leu Thr Pro Pro Pro Ser Tyr Ala Ala Thr
305                 310                 315                 320

Ile Ala Ser Lys Leu Ala Ile His Asn Pro Asn Leu Pro Ala Thr Leu
                325                 330                 335

Pro Val Asn Ser Pro Thr Leu Pro Pro Val Arg Tyr Asn Arg Arg Ser
            340                 345                 350

Asn Pro Asp Leu Glu Lys Arg Arg Ile His Phe Cys Asp Tyr Asn Gly
        355                 360                 365

Cys Thr Lys Val Tyr Thr Lys Ser Ser His Leu Lys Ala His Leu Arg
    370                 375                 380

Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Thr Trp Glu Gly Cys Asp
385                 390                 395                 400

Trp Arg Phe Ala Arg Ser Asp Glu Leu Thr Arg His Tyr Arg Lys His
                405                 410                 415

Thr Gly Ala Lys Pro Phe Gln Cys Met Val Cys Gln Arg Ser Phe Ser
            420                 425                 430

Arg Ser Asp His Leu Ala Leu His Met Lys Arg His Gln Asn
```

-continued

```
        435                 440                 445

<210> SEQ ID NO 37
<211> LENGTH: 877
<212> TYPE: DNA
<213> ORGANISM: Rattus norvegicus
<220> FEATURE:
<223> OTHER INFORMATION: rat Kruppel-like factor 5, intestinal (KLF5,
      CKLF) cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (145)..(792)
<223> OTHER INFORMATION: CKLF

<400> SEQUENCE: 37

cggtatttca gctcccacca gaccacagcg ccagaggtga acaatatctt catcaaacaa      60

gaacttccta taccagatct tcatctctcg gtcccttccc agcagggcca cctgtaccag     120

ctgttgaata cacctgatct agacatgccc agttcgacaa accagacagc agtcatggac     180

acccttaatg tctctatggc tggccttaac tcacacccct ctgctgtgcc acagacgtcc     240

atgaaacagt tccagggcat gcctccttgc acgtacacca tgccgagtca gtttcttcca     300

cagcaggcca cctactttcc cccatcacca ccgagctcag agcctggaag tcctgataga     360

caagctgaga tgctccagaa tctgacccca cctccgtcct atgctgctac aattgcttcg     420

aaactggcaa ttcacaatcc aaatttacct gccactctgc cagttaattc gccaaatatc     480

caacctgtcc gatacaacag aaggagtaac ccggatctgg agaagcgacg catccatttc     540

tgtgattatg atggttgcac aaaagtttat acaaagtcgt ctcatttaaa agctcacctg     600

aggactcata cgggcgagaa gccctacaag tgcacctggg agggctgcga ctggaggttt     660

gcccggtcgg acgagctgac ccgccactac aggaagcaca cgggtgccaa gccgttccag     720

tgcgtggtgt gcaaccgcag cttctcccgc tccgaccacc tggcgctgca catgaagcgc     780

caccagaact gagcactgcg cacaaccggc tcgacgcctc gcagtccgct cgccatcctt     840

taaaccgcag acctaacttc atataaaaaa aaaaaaa                              877

<210> SEQ ID NO 38
<211> LENGTH: 215
<212> TYPE: PRT
<213> ORGANISM: Rattus norvegicus
<220> FEATURE:
<223> OTHER INFORMATION: rat Kruppel-like factor 5, intestinal (KLF5,
      CKLF)

<400> SEQUENCE: 38

Met Pro Ser Ser Thr Asn Gln Thr Ala Val Met Asp Thr Leu Asn Val
  1               5                  10                  15

Ser Met Ala Gly Leu Asn Ser His Pro Ser Ala Val Pro Gln Thr Ser
             20                  25                  30

Met Lys Gln Phe Gln Gly Met Pro Pro Cys Thr Tyr Thr Met Pro Ser
         35                  40                  45

Gln Phe Leu Pro Gln Gln Ala Thr Tyr Phe Pro Pro Ser Pro Pro Ser
     50                  55                  60

Ser Glu Pro Gly Ser Pro Asp Arg Gln Ala Glu Met Leu Gln Asn Leu
 65                  70                  75                  80

Thr Pro Pro Pro Ser Tyr Ala Ala Thr Ile Ala Ser Lys Leu Ala Ile
                 85                  90                  95
```

-continued

```
His Asn Pro Asn Leu Pro Ala Thr Leu Pro Val Asn Ser Pro Asn Ile
            100                 105                 110

Gln Pro Val Arg Tyr Asn Arg Arg Ser Asn Pro Asp Leu Glu Lys Arg
        115                 120                 125

Arg Ile His Phe Cys Asp Tyr Asp Gly Cys Thr Lys Val Tyr Thr Lys
    130                 135                 140

Ser Ser His Leu Lys Ala His Leu Arg Thr His Thr Gly Glu Lys Pro
145                 150                 155                 160

Tyr Lys Cys Thr Trp Glu Gly Cys Asp Trp Arg Phe Ala Arg Ser Asp
                165                 170                 175

Glu Leu Thr Arg His Tyr Arg Lys His Thr Gly Ala Lys Pro Phe Gln
            180                 185                 190

Cys Val Val Cys Asn Arg Ser Phe Ser Arg Ser Asp His Leu Ala Leu
        195                 200                 205

His Met Lys Arg His Gln Asn
    210                 215
```

```
<210> SEQ ID NO 39
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:MAST205b PCR
      Forward primer 110F

<400> SEQUENCE: 39

acagcagtcc tggcactcct t                                              21


<210> SEQ ID NO 40
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:MAST205b PCR
      Reverse primer 174R

<400> SEQUENCE: 40

gcggttactt gtccgacaac tc                                             22


<210> SEQ ID NO 41
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:MAST205b PCR
      Taqman Probe Probe133

<400> SEQUENCE: 41

tccagccgcc cactgccg                                                  18


<210> SEQ ID NO 42
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:MAST205 PCR
      Forward primer 717F

<400> SEQUENCE: 42

ttggacagtc tgcaccttct ctta                                           24
```

-continued

```
<210> SEQ ID NO 43
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:MAST205 PCR
      Reverse primer 801R

<400> SEQUENCE: 43

cggttacttg tccgacaaaa gc                                              22


<210> SEQ ID NO 44
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:MAST205 PCR
      Taqman Probe Probe745

<400> SEQUENCE: 44

tggcctgaag gacttgagcc ttccagccca ctgccg                               36


<210> SEQ ID NO 45
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:
      hexahistidine (His) affinity tag

<400> SEQUENCE: 45

His His His His His His
  1               5


<210> SEQ ID NO 46
<211> LENGTH: 200
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:poly-Gly
      flexible linker
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (6)..(200)
<223> OTHER INFORMATION: Gly residues from position 6 to 200 may be
      present or absent

<400> SEQUENCE: 46

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
  1               5                  10                  15

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
             20                  25                  30

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
         35                  40                  45

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
     50                  55                  60

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 65                  70                  75                  80

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
                 85                  90                  95

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
            100                 105                 110
```

-continued

```
Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
        115                 120                 125

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
    130                 135                 140

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
145                 150                 155                 160

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
                165                 170                 175

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
            180                 185                 190

Gly Gly Gly Gly Gly Gly Gly Gly
        195                 200
```

What is claimed is:

1. A method for identifying an agent for treating a diabetic or pre-diabetic individual, the method comprising the steps of:

(i) contacting an agent to a mixture comprising a polypeptide encoded by a nucleic acid that hybridizes under stringent conditions to a nucleic acid encoding SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34; and

(ii) selecting an agent that modulates the expression or activity of the polypeptide or that binds to the polypeptide, thereby identifying an agent for treating a diabetic or pre-diabetic individual.

2. The method of claim 1, the method further comprising selecting an agent that modulates insulin sensitivity.

3. The method of claim 1, wherein step (ii) comprises selecting an agent that modulates expression of the polypeptide.

4. The method of claim 1, wherein step (ii) comprises selecting an agent that modulates the activity of the polypeptide.

5. The method of claim 1, wherein step (ii) comprises selecting an agent that specifically binds to the polypeptide.

6. The method of claim 1, wherein the polypeptide is expressed in a cell and the cell is contacted with the agent.

7. The method of claim 1, wherein the polypeptide is SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

8. A method of treating a diabetic or pre-diabetic animal, the method comprising administering to the animal a therapeutically effective amount of an agent identified by the method of claim 1.

9. The method of claim 8, wherein the agent is an antibody.

10. The method of claim 9, wherein the antibody is a monoclonal antibody.

11. The method of claim 8, wherein the animal is a human.

12. A method of introducing an expression cassette into a cell, the method comprising,

introducing into the cell an expression cassette comprising a promoter operably linked to a polynucleotide encoding a polypeptide, wherein the polynucleotide hybridizes under stringent conditions to a nucleic acid encoding SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

13. The method of claim 12, wherein the polypeptide comprises SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

14. The method of claim 12, wherein the cell is selected from the group consisting of an adipocyte and a skeletal muscle cell.

15. The method of claim 12, the method further comprising introducing the cell into a human.

16. The method of claim 15, wherein the human is diabetic.

17. The method of claim 15, wherein the human is prediabetic.

18. The method of claim 15, wherein the cell is from the human.

19. A method of diagnosing an individual who has Type 2 diabetes or is prediabetic, the method comprising,

detecting in a sample from the individual the level of a polypeptide or the level of a polynucleotide encoding the polypeptide, wherein the polynucleotide hybridizes under stringent conditions to a nucleic acid encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:34;

wherein a modulated level of the polypeptide or polynucleotide in the sample compared to a level of the polypeptide or polynucleotide in either a lean individual or a previous sample from the individual indicates that the individual is diabetic or prediabetic.

20. The method of claim 19, wherein the detecting step comprises contacting the sample with an antibody that specifically binds to the polypeptide.

21. The method of claim 19, wherein the amino acid sequence comprises SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

22. The method of claim 19, wherein the detecting step comprises quantifying mRNA encoding the polypeptide.

23. The method of claim 22, wherein the mRNA is reverse transcribed and amplified in a polymerase chain reaction.

24. The method of claim 19, wherein the sample is a blood, urine or tissue sample.

25. An isolated nucleic acid encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:10 or SEQ ID NO:28.

26. The isolated nucleic acid of claim 25, wherein the nucleic acid comprises the sequence set forth in SEQ ID NO:9 or SEQ ID NO:27.

27. An expression cassette comprising the isolated nucleic acid of claim 25.

28. A host cell comprising the expression cassette of claim 27.

* * * * *

patsnap

专利名称(译)	诊断和治疗糖尿病和胰岛素抵抗的方法		
公开(公告)号	US20060234292A1	公开(公告)日	2006-10-19
申请号	US10/516780	申请日	2003-06-05
[标]申请(专利权)人(译)	ALLAN BERNARD GREGOIRE弗朗辛 拉旺BRIAN 穆迪SHONNA WATERS史蒂夫 黄志伟		
申请(专利权)人(译)	ALLAN BERNARD GREGOIRE弗朗辛 拉旺BRIAN 穆迪SHONNA WATERS史蒂夫 黄威志		
当前申请(专利权)人(译)	METABOLEX INC.		
[标]发明人	ALLAN BERNARD GREGOIRE FRANCINE LAVAN BRIAN MOODIE SHONNA WATERS STEVE WONG CHI WAI		
发明人	ALLAN, BERNARD GREGOIRE, FRANCINE LAVAN, BRIAN MOODIE, SHONNA WATERS, STEVE WONG, CHI-WAI		
IPC分类号	A61K38/22 G01N33/53 C07H21/04 C07K14/575 A61K48/00 C07K14/47 G01N33/68		
CPC分类号	A61K38/16 A61K48/00 C07K14/4713 G01N33/5008 G01N2800/042 G01N33/5091 G01N33/6893 G01N2500/04 G01N2500/10 G01N33/5023		
优先权	60/386521 2002-06-05 US 60/386551 2002-06-05 US 60/386954 2002-06-06 US 60/386936 2002-06-06 US 60/387301 2002-06-07 US 60/386429 2002-06-06 US 60/386527 2002-06-05 US		
外部链接	Espacenet USPTO		

摘要(译)

本发明提供了用于诊断和治疗糖尿病和胰岛素抵抗的组合物和方法。特别地，本发明提供了鉴定本发明的多核苷酸或多肽的调节剂并使用那些调节剂治疗糖尿病的方法，以及通过测量患者中本发明的多核苷酸或多肽的水平来诊断糖尿病的方法。

	Diabetic Pre-Trog			Diabetic Post-Trog				Stu-	
B/C	Mean Expr	SEM	N	Mean Expr	SEM	N	Fold Change	dents t test	Gene name
B	1234	411	9	919	325	8	0.74	0.01	MAST205