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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 May 2003 (22.05.2003)

PCT

(10) International Publication Number
WO 03/042358 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/34622

(22) International Filing Date: 29 October 2002 (29.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/335,084 31 October 2001 (31.10.2001) US

(71) Applicant *(for all designated States except US)*: **MILLENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **CHUN, Miyoung** [US/US]; 88 Country Club Lane, Belmont, MA 02478 (US).

(74) Agent: **BOSSONE, Steven, A.**; Millennium Pharmaceuticals, Inc., 75 Sidney Street, Cambridge, MA 02139 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

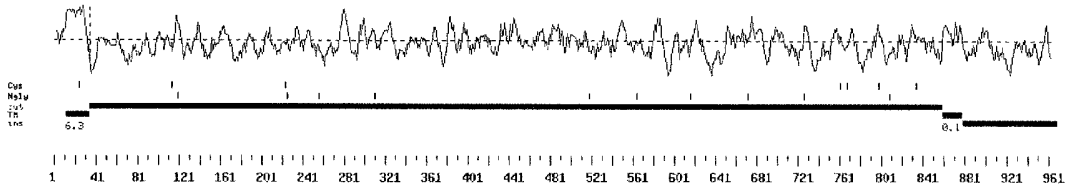
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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

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

WO 03/042358 A2

(54) Title: METHODS OF USING 48149, A HUMAN AMINOPEPTIDASE FAMILY MEMBER



(57) Abstract: Isolated nucleic acids molecules, designated 48149 nucleic acid molecules, which encode human Aminopeptidase N, are disclosed. The invention provides methods of modulating 48149 activity, which is associated with the formation of atherosclerotic lesions in blood vessels. The invention further provides methods of treating, preventing and diagnosing cardiovascular disorders such as atherosclerosis, as well as disorders associated with the metabolism of lipids, for example, in the liver.

METHODS OF USING 48149, A HUMAN AMINOPEPTIDASE FAMILY MEMBER

This applications claims priority to U.S. provisional application number 60/335,084, filed on October 31, 2001, the entire contents of which are incorporated herein by reference.

Aminopeptidase N (APN/CD13) is a human metalloprotease that catalyzes the removal of N-terminal, preferentially neutral, amino acid residues from small peptides (for review see Sanderink et al. (1988) *J. Clin. Chem. Clin. Biochem.* 26:795-807; Shipp and Look (1993) *Blood* 82:1052-1070). This peptidase is a transmembrane ectoenzyme expressed on several tissues, including brush border membranes of kidney proximal tubules, intestine and placenta, as well as hematopoietic cells. Expression of APN/CD13 in hematopoeitic cells has been considered specific for cells of the myeloid lineage because granulocytes and monocytes/macrophages, but not lymphocytes of peripheral blood, show a surface expression of the CD13 antigen (Kehlen, A. et al. (2000) *J Cellular Biochemistry* 80:115-123). However, expression of APN/CD13 mRNA is upregulated on lymphocytes upon cell-cell contact with endothelial cells, monocytes, and fibroblast-like synoviocytes (SFCs).

Atherosclerosis is a chronic inflammatory disease resulting from the interaction between activated monocyte/macrophages and endothelial and smooth muscle cells, involving humoral and cell-mediated immunological responses stimulated by the arterial walls. The role of hypercholesterolemia in the etiology of atheroscelrosis is well established. Targeting the genes that are involved in lipid metabolism is beneficial for the prevention of atherosclerosis. Presently, cholesterol-lowering drugs alone are not sufficient to fully prevent the progression of atherosclerosis in many susceptible individuals.

Accordingly, there is a need for new therapies that prevent and/or revert lesion formation in atherosclerotic vessels, thereby attenuating the progression of atherogenesis.

Nucleotide and corresponding amino acid sequences for an aminopeptidase family member, referred to herein as "48149", are disclosed. The nucleotide sequence of a cDNA encoding human 48149 is depicted in SEQ ID NO:1, and the amino acid sequence of a human 48149 polypeptide is depicted in SEQ ID NO:2. In addition, the nucleotide sequences of the human 48149 coding region are depicted in SEQ ID NO:3. Applicants have shown upregulated expression of 48149 mRNA in atherosclerotic plaques, as

compared to non-lesioned arterial wall tissue, as well as in macrophages during the process of lipid loading. Furthermore, expression of 48149 mRNA was decreased in monkeys that were fed a diet low in saturated fat, while increased in marmosets treated with cholesteyramin. Thus, the expression of 48149 mRNA positively correlates with atherosclerotic plaque formation or conditions that favor atherosclerotic plaque formation. Accordingly, modulators of 48149 nucleic acid expression or 48149 polypeptide activity may be used to treat or prevent metabolic and cardiovascular disorders such as atherosclerosis.

Accordingly, in one aspect, the invention features a nucleic acid molecule that encodes a 48149 protein or polypeptide, e.g., a biologically active portion of the 48149 protein. In a preferred embodiment the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence of SEQ ID NO:2. In other embodiments, the invention provides isolated 48149 nucleic acid molecules having the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3. In still other embodiments, the invention provides nucleic acid molecules that are substantially identical (e.g., naturally occurring allelic variants) to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3. In other embodiments, the invention provides a nucleic acid molecule which hybridizes under a stringency condition described herein to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, wherein the nucleic acid encodes a full length 48149 protein or an active fragment thereof.

In a related aspect, the invention further provides nucleic acid constructs that include a 48149 nucleic acid molecule described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or heterologous regulatory sequences. Also included, are vectors and host cells containing the 48149 nucleic acid molecules of the invention e.g., vectors and host cells suitable for producing 48149 nucleic acid molecules and polypeptides.

In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for the detection of 48149-encoding nucleic acids.

In still another related aspect, isolated nucleic acid molecules that are antisense to a 48149 encoding nucleic acid molecule are provided.

In another aspect, the invention features, 48149 polypeptides, and biologically active or antigenic fragments thereof that are useful, e.g., as reagents or targets in assays applicable to treatment and diagnosis of 48149-mediated or -related disorders. In another embodiment, the invention provides 48149 polypeptides having a 48149 activity.

Preferred polypeptides are 48149 proteins including at least a peptidase domain and an extracellular pentapeptide consensus sequence required for zinc coordination and catalytic activity, and, preferably, having a 48149 activity, e.g., a 48149 activity as described herein.

In other embodiments, the invention provides 48149 polypeptides, e.g., a 48149 polypeptide having the amino acid sequence shown in SEQ ID NO:2; an amino acid sequence that is substantially identical to the amino acid sequence shown in SEQ ID NO:2; or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under a stringency condition described herein to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, wherein the nucleic acid encodes a full length 48149 protein or an active fragment thereof.

In a related aspect, the invention provides 48149 polypeptides or fragments operatively linked to non-48149 polypeptides to form fusion proteins.

In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably specifically bind 48149 polypeptides or fragments thereof, e.g., an extracellular domain of a 48149 polypeptide, e.g., a peptidase domain of a 48149 polypeptide.

In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the 48149 polypeptides or nucleic acids.

In yet another aspect, the invention features a method of modulating (e.g., inhibiting) the activity or expression of 48149 molecules. The method includes contacting one or more of: 48149, a 48149-expressing cell or tissue, or an activator of 48149, with an agent, e.g., an 48149 inhibitor, in an amount sufficient to modulate (e.g., inhibit) the activity or expression of 48149. The subject method can be used on cells in culture, e.g. in vitro or ex vivo, or in vivo in a subject, e.g., as part of an *in vivo* therapeutic or prophylactic protocol. For in vitro embodiments, 48149 can be contacted with the agent by, e.g., forming a mixture, e.g., a reconstituted system, which includes 48149 and the agent. In other embodiments, a 48149-expressing cell (e.g., a macrophage, an endothelial cell, a smooth muscle cell, or a liver cell), or a 48149-expressing tissue (e.g., a cardiovascular tissue, an atheroma-associated tissue, or a liver tissue) is contacted with the agent by, e.g., adding the agent to the culture medium.

The method can also be performed in vivo in a subject. Preferably, the agent, or a pharmaceutically acceptable composition thereof, is administered to the subject in an amount effective to inhibit the activity or expression of 48149. The method can be used for the treatment of, or prophylactic prevention of, e.g., a disorder involving aberrant

activity of macrophage cells, endothelial cells, smooth muscle cells, or liver cells (e.g., a cardiovascular disorder, such as atherosclerosis).

For *ex vivo* embodiments, the method further includes removing 48149 or 48149-expressing cells from the subject. For example, blood containing 48149 or 48149-expressing cells, e.g., 48149-expressing macrophages, can be obtained from the subject. 48149 or 48149-expressing cells can be treated with the agent in an amount effective to inhibit the activity or expression of 48149. Treated 48149-expressing cells can be introduced into the subject.

In a preferred embodiment, the method further includes evaluating 48149 nucleic acid or protein expression level or activity in the cell or subject before or after the administration or contacting step. For example, a subject, e.g., a patient having, or at risk of a cardiovascular disorder, can be evaluated before or after the agent is administered. If the subject has a level of 48149 above a predetermined level, therapy can begin or be continued.

In a preferred embodiment, the 48149 is human 48149. All forms of 48149 (i.e., active and latent forms) can be inhibited. Preferably the agent inhibits the active form of 48149.

In a preferred embodiment, the agent decreases the expression or activity of 48149, e.g., human 48149. In one embodiment, the agent can directly inhibit the activity or expression of 48149. For example, the agent can interact with, e.g., bind to, a 48149 protein and block or reduce the 48149 aminopeptidase activity, e.g., hydrolysis of 48149 substrate polypeptides. In other embodiments, the agent can block or reduce expression (e.g., transcription, translation, or mRNA or protein stability) of 48149.

In a preferred embodiment, the agent is a small molecule (e.g., a chemical agent having a molecular weight of less than 2500 Da, preferably, less than 1500 Da), a chemical, e.g., a small organic molecule, e.g., a product of a combinatorial or natural product library; a polypeptide (e.g., an antibody, such as an 48149 specific antibody); a peptide, a peptide fragment (e.g., a substrate fragment such as a collagen I fragment), or a peptidomimetic; a modulator (e.g., an inhibitor) of expression of an 48149 nucleic acid, such as an antisense, a ribozyme, or a triple helix molecule; or any combination thereof.

Preferably, the agent is a 48149-specific inhibitor. Examples of 48149-specific inhibitors include, but are not limited to, a small molecule 48149-specific inhibitor; or an anti-48149 antibody (e.g., a humanized, chimeric, human, or other recombinant (e.g., phage display) anti-48149 antibody).

In other embodiments, the agent is a non-specific aminopeptidase inhibitor (i.e., it inhibits two or more aminopeptidases).

In a preferred embodiment, the 48149-expressing cell or tissue is an atheroma-associated cell or tissue, e.g., a human atheroma-associated cell or tissue. Preferably, the atheroma-associated cell or tissue is an endothelial cell or tissue, a smooth muscle cell or tissue, or a monocyte or macrophage. In *in vivo* embodiments, the cell or tissue is associated with (e.g., located in or nearby) an atherosclerotic lesion or plaque, e.g., an early, intermediate or advanced atherosclerotic lesion or plaque. In one preferred embodiment, the cell or tissue is associated with (e.g., located in or nearby) an advanced or rupture-prone atherosclerotic lesion.

Examples of cardiovascular disorders (e.g., inflammatory disorders) that can be treated or prevented with the methods of the invention include, but are not limited to, atherosclerosis, myocardial infarction, stroke, thrombosis, aneurism, heart failure, ischemic heart disease, angina pectoris, sudden cardiac death, hypertensive heart disease; non-coronary vessel disease, such as arteriolosclerosis, small vessel disease, nephropathy, hypertriglyceridemia, hypercholesterolemia, hyperlipidemia, xanthomatosis, asthma, hypertension, emphysema and chronic pulmonary disease; or a cardiovascular condition associated with interventional procedures ("procedural vascular trauma"), such as restenosis following angioplasty, placement of a shunt, stent, synthetic or natural excision grafts, indwelling catheter, valve or other implantable devices. Preferred cardiovascular disorders include atherosclerosis, myocardial infarction, aneurism, and stroke.

In a preferred embodiment, the cardiovascular disorder is caused by aberrant lipid (e.g., fatty acid) metabolism. Examples of disorders involving aberrant lipid metabolism include, but are not limited to, atherosclerosis, arteriolosclerosis, hypertriglyceridemia, obesity, diabetes, hypercholesterolemia, xanthomatosis, and hyperlipidemia. Most preferable, the disorder is atherosclerosis.

In other preferred embodiments, the 48149-expressing cell is a macrophage (e.g., a monocyte-derived macrophage), an endothelial cells, a smooth muscle cell, or a liver cell. Since macrophages are involved in non-neutrophil mediated inflammatory conditions (e.g., chronic inflammatory conditions), the methods of the invention also encompass non-neutrophil mediated-inflammatory disorders, including but not limited to, an autoimmune disease (e.g., rheumatoid arthritis, allergy, multiple sclerosis, autoimmune diabetes, autoimmune uveitis and nephrotic syndrome), an infectious disease, a malignancy,

transplant rejection or graft-versus-host disease, a pulmonary disorder, a bone disorder, or an intestinal disorder as described herein.

In other embodiments, the 48149 expressing cell is an endothelial cell. Therefore, the methods of the invention can be used to treat, prevent and/or diagnose an endothelial cell mediated disorder, e.g., a disorder involving aberrant proliferation, migration, angiogenesis, or vascularization; or aberrant expression of cell surface adhesion molecules or genes associated with angiogenesis, e.g., TIE-2, FLT and FLK. Endothelial cell disorders include tumorigenesis, tumor metastasis, psoriasis, diabetic retinopathy, endometriosis, Grave's disease, ischemic disease (e.g., atherosclerosis), and chronic inflammatory diseases (e.g., rheumatoid arthritis).

In a preferred embodiment, the subject is a human suffering from, or at risk of, an 48149-mediated disorder or disease, e.g., a cardiovascular disorder, a non-neutrophil-mediated disorders, or an endothelial cell disorder, as described herein. For example, the subject is a patient undergoing a therapeutic or prophylactic protocol.

In a preferred embodiment, the subject is a human suffering from, or at risk of, atherosclerosis. For example, a human with early, intermediate or advanced atherosclerosis. Preferably, the subject is a human suffering from, or at risk of, rupture of an atherosclerostic plaque.

In other embodiments, the subject is a non-human animal, e.g., an experimental animal, e.g., monkey.

The agent(s) described herein can be administered by themselves, or in combination with at least one more agent (referred to herein as a "second agent(s)"), or procedures. In one embodiment, an 48149 specific agent is administered in combination with a non-specific aminopeptidase inhibitor, or a drug used to treat atherosclerosis.

In yet other embodiments, the agents of the invention can be administered alone or in combination with a cholesterol lowering agent. Examples of cholesterol lowering agents include bile acid sequestering resins (e.g. colestipol hydrochloride or cholestyramine), fibric acid derivatives (e.g. clofibrate, fenofibrate, or gemfibrozil), thiazolidenediones (e.g., troglitazone, pioglitazone, ciglitazone, englitazone, rosiglitazone), or hydroxymethylglutaryl coenzyme A reductase (HMG-CoA reductase) inhibitors (e.g. statins, such as fluvastatin sodium, lovastatin, pravastatin sodium, simvastatin, atorvastatin calcium, cerivastatin), an ApoAII-lowering agent, a VLDL lowering agent, an ApoAI-stimulating agent, as well as inhibitors of, nicotinic acid, niacin, or probucol. Preferred cholesterol lowering agents include inhibitors of HMG-CoA reductase (e.g., statins),

nicotinic acid, and niacin. Preferably, the cholesterol lowering agent results in a favorable plasma lipid profile (e.g., increased HDL and/or reduced LDL).

In other embodiments, the agent(s) of the invention is administered in combination with an interventional procedure ("procedural vascular trauma"). Examples of interventional procedures, include but are not limited to, angioplasty, placement of a shunt, stent, synthetic or natural excision grafts, indwelling catheter, valve and other implantable devices.

The second agent or procedure can be administered or effected prior to, at the same time, or after administration of the agent(s) of the invention, in single or multiple administration schedules. For example, the second agent and the agents of the invention can be administered continually over a preselected period of time, or administered in a series of spaced doses, i.e., intermittently, for a period of time.

In a preferred embodiment, the agent of the invention, alone or in combination with the second agent or procedure, inhibit (block, reduce or prevent) one or more of: inhibit atherosclerotic lesion formation, development or rupture; inhibit lipid accumulation, increase plaque stability or promote lesion regression; inhibit collagenolysis, e.g., degradation of type I, II, or III, preferably type I collagen, or the breakdown of intact, triple helical collagen; or inhibit rupture of atherosclerotic plaques.

In a preferred embodiment, the method further includes removing from the subject 48149 or 48149-expressing cells (e.g., macrophages, endothelial cells, smooth muscle cells, or liver cells), e.g., by drawing blood or obtaining a tissue biopsy.

In yet another aspect, the invention features a method of treating or preventing a cardiovascular disorder, e.g., a cardiovascular disorder as described herein (e.g., atherosclerosis), in a subject. The method includes administering to the subject an agent that inhibits the activity or expression of 48149, e.g., an agent as described herein, in an amount effective to treat or prevent the cardiovascular disorder.

The invention also features a method of diagnosing, or staging, a 48149-mediated disorder, e.g., a cardiovascular disorder (e.g., atherosclerosis), an endothelial cell diosrder, a liver disorder, or a non-neutrophil-mediated inflammatory disorder, in a subject. The method includes evaluating the expression, activity or processing, of a 48149 nucleic acid or polypeptide, thereby diagnosis or staging the disorder. In a preferred embodiment, the expression or activity is compared with a reference value, e.g., a difference in the expression or activity level of the 48149 nucleic or polypeptide relative to reference, e.g., a

normal subject or a cohort of normal subjects, is indicative of the disorder, or a stage in the disorder.

In a preferred embodiment, the subject is a human. For example, the subject is a human suffering from, or at risk of, a cardiovascular disorder as described herein. Preferably, subject is a human suffering from, or at risk of, atherosclerosis; a human with early, intermediate or advanced atherosclerosis; or a human suffering from, or at risk of, rupture of an atherosclerostic plaque. In other embodiments, the subject is a human suffering from, or at risk of, an endothelial cell disorder, a liver disorder, or a non-neutrophil-mediated inflammatory disorder as described herein.

In a preferred embodiment, the evaluating step occurs in vitro or ex vivo. For example, a sample, e.g., blood, plasma, a tissue sample, or a biopsy, is obtained from the subject. Preferably, the sample contains an 48149-expressing cell, e.g., an atheroma-associated cells (e.g., macrophages, endothelial cells, or smooth muscle cells). In one embodiment, the level of 48149 substrate breakdown products present in, e.g., a subject's plasma, can be evaluated.

In a preferred embodiment, the evaluating step occurs in vivo. For example, by administering to the subject a detectably labeled agent that interacts with the 48149-associated nucleic acid or polypeptide, such that a signal is generated relative to the level of activity or expression of the 48149 nucleic acid or polypeptide.

In preferred embodiments, the method is performed: on a sample from a subject, a sample from a human subject; e.g., a sample of a patient suffering from, or at risk of, a cardiovascular, endothelial, liver, or non-neutrophil-mediated inflammatory disorder as described herein; e.g., a sample of a patient suffering from, or at risk of, atherosclerosis (e.g., a human with early, intermediate or advanced atherosclerosis); or a sample of a human suffering from, or at risk of, rupture of an atherosclerostic plaque; to determine if the individual from which the target nucleic acid or protein is taken should receive a drug or other treatment; to diagnose an individual for a disorder or for predisposition to resistance to treatment, to stage a disease or disorder.

In a preferred embodiment, the level of expression of at least one, two, three or four atherosclerosis-associated nucleic acids or polypeptides is evaluated. Preferably, the atherosclerosis-associated nucleic acid or polypeptide is 48149, preferably human 48149.

In a preferred embodiment, the expression of the atherosclerosis-associated nucleic acid (e.g., 48149) is evaluated by evaluating the expression of a signal entity, e.g., a green

fluorescent protein or other marker protein, which is under the control or an atherosclerosis-associated gene control element, e.g., a 48149 promoter.

In some embodiments, the expression of one or more atherosclerosis-associated nucleic acid or polypeptide is evaluated by contacting said sample with, a nucleic acid probe that selectively hybridizes to one or more atherosclerosis-associated nucleic acids or polypeptides. An increase in the level of said one or more atherosclerosis-associated nucleic acids or polypeptides, relative to a control, indicates a disorder, or a stage in the disorder.

In some embodiments, nucleic acid (or protein) from the cell or sample is analyzed on positional arrays, e.g., DNA-chip arrays. Accordingly, in preferred embodiments the method further includes:

analyzing the sample by providing an array of a plurality of capture probes, wherein each of the capture probes is positionally distinguishable from other capture probes of the plurality on the array, and wherein each positional distinguishable capture probe includes a unique reagent, e.g., an antibody or a nucleic acid probe which can identify an atherosclerosis- (48149)-associated nucleic acid or polypeptide; and

hybridizing the sample with the array of capture probes, thereby analyzing the sample sequence.

In a preferred embodiment, the 48149-mediated disorder is a cardiovascular disorder, e.g., a cardiovascular disorder as described herein. Preferably, the disorder is atherosclerosis (e.g., early, intermediate or advanced atherosclerosis). Most preferably, the disorder is advanced stage atherosclerosis, e.g., an atherosclerotic stage characterized by rupture-prone atherosclerotic plaques or lesions.

In a preferred embodiment, the 48149-mediated disorder is an endothelial disorder as described herein.

In other embodiments, the 48149-mediated disorder is liver disorder or a non-neutrophil-mediated inflammatory disorder as described herein.

In a further aspect, the invention provides assays for determining the presence or absence of a genetic alteration in a 48149 nucleic acid or polypeptide, including for disease diagnosis, a response to cardiovascular therapy.

In a related aspect, the invention provides a method of evaluating a subject, e.g., to identify a predisposition to a 48149 mediated disorder (e.g., a cardiovascular, endothelial cell or non-neutrophil mediated inflammatory disorder), diagnose, or treat the subject. The method includes providing a nucleic acid of the subject; and either a) determining the

allelic identity of an atherosclerosis-associated nucleic acid (e.g., 48149, preferably, human 48149) or b) determining the sequence of at least a nucleotide of the nucleic acid. In a preferred embodiment, the method further includes comparing the allelic identity or sequence to a reference allele or reference sequence of the nucleic acid. The reference allele or reference sequence is associated with an immune disorder or a functional (e.g., normal) immune system. Allelic variants can be detected using, e.g., arrays, mismatch cleavage, electrophoretic assays, HPLC assays, and nucleic acid sequencing. Preferably, the assays detect nucleotide substitutions, and preferably, also insertions, deletions, translocations, and rearrangements of an atherosclerosis-associated nucleic acid (e.g., 48149, preferably, human 48149).

In a preferred embodiment, the method further includes diagnosing a subject, and/or choosing a therapeutic modality, e.g., a particular treatment, or a dosage thereof, based on the level of atherosclerosis-associated nucleic acid (e.g., 48149) expression or allelic identity.

In another aspect, the invention features, a method for evaluating the efficacy of a treatment of a disorder, e.g., an 48149-mediated disorder, e.g., a cardiovascular disorder (e.g., atherosclerosis), an endothelial cell disorder, a liver disorder, or a non-neutrophil-mediated inflammatory disorder, in a subject. The method includes evaluating the expression of one or more atherosclerosis-associated nucleic acids or polypeptides, thereby evaluating the efficacy of the treatment. In a preferred embodiment, the expression or activity is compared with a reference value. A change, e.g., decrease, in the level of said one or more atherosclerosis-associated nucleic acids or polypeptides in a sample obtained after treatment, relative to the level of expression before treatment, is indicative of the efficacy of the treatment of said disorder.

In a preferred embodiment, the subject is a human. For example, the subject is a human suffering from, or at risk of, a cardiovascular disorder as described herein. Preferably, subject is a human suffering from, or at risk of, atherosclerosis; a human with early, intermediate or advanced atherosclerosis; or a human suffering from, or at risk of, rupture of an atherosclerostic plaque. In other embodiments, the subject is a human suffering from, or at risk of, a liver disorder, a non-neutrophil-mediated inflammatory disorder, or an endothelial disorder, as described herein.

In a preferred embodiment, the evaluating step occurs in vitro or ex vivo. For example, a sample, e.g., blood, plasma, tissue sample, a biopsy, is obtained from the

subject. Preferably, the sample contains atheroma-associated cells, e.g., macrophages, endothelial cells, or smooth muscle cells, or liver cells.

For in vitro embodiments, the method includes providing a sample, e.g., a tissue, a bodily fluid (e.g., blood), a biopsy, from said subject; and

evaluating the expression of one or more atherosclerosis-associated nucleic acids or polypeptides, e.g., by contacting said sample with, a nucleic acid probe that selectively hybridizes to one or more atherosclerosis-associated nucleic acids, or an antibody that specifically binds to one or more atherosclerosis-associated polypeptides;

wherein a change, e.g., decrease, in the level of said one or more atherosclerosis-associated nucleic acids or polypeptides in a sample obtained after treatment, relative to the level of expression before treatment, is indicative of the efficacy of the treatment of said disorder.

In preferred embodiments, the method is performed: on a sample from a subject, a sample from a human subject; e.g., a sample of a patient suffering from, or at risk of, a cardiovascular, liver, or non-neutrophil-mediated inflammatory disorder as described herein; e.g., a sample of a patient suffering from, or at risk of, atherosclerosis (e.g., a human with early, intermediate or advanced atherosclerosis); or a sample of a human suffering from, or at risk of, rupture of an atherosclerostic plaque.

In a preferred embodiment, the atherosclerosis-associated nucleic acid or polypeptide is 48149, preferably human 48149.

In a preferred embodiment, the sample contains atheroma-associated cells, e.g., macrophages, endothelial cells, or smooth muscle cells.

In a preferred embodiment, the method further includes diagnosis and/or choosing a therapeutic modality, e.g., a particular treatment, or a dosage thereof, based on the level of atherosclerosis-associated nucleic acid expression (e.g., 48149 expression).

In a preferred embodiment, the expression of the atherosclerosis-associated nucleic acid (e.g., 48149) is evaluated by evaluating the expression of a signal entity, e.g., a green fluorescent protein or other marker protein, which is under the control or an atherosclerosis-associated gene control element e.g., a 48149 promoter.

In some embodiments, nucleic acid (or protein) from the cell or sample is analyzed on positional arrays, e.g., DNA-chip arrays. Accordingly, in preferred embodiments the method further includes:

analyzing the sample by providing an array of a plurality of capture probes, wherein each of the capture probes is positionally distinguishable from other capture

probes of the plurality on the array, and wherein each positional distinguishable capture probe includes a unique reagent, e.g., an antibody or a nucleic acid probe which can identify an atherosclerosis-associated nucleic acid or polypeptide, e.g., 48149; and

hybridizing the sample with the array of capture probes, thereby analyzing the sample sequence.

In a preferred embodiment, the evaluating step occurs in vivo. For example, by administering to the subject a detectably labeled agent that interacts with the 48149-associated nucleic acid or polypeptide, such that a signal is generated relative to the level of activity or expression of the 48149 nucleic acid or polypeptide.

In yet another aspect, the invention features a method of selecting a cell (e.g., a macrophage, endothelial cell, smooth muscle cell, or liver cell) having a selected level of 48149 expression or activity.

In a preferred embodiment, the method compares the expression of 48149 to a preselected standard, e.g., a control cell.

In a preferred embodiment, the method includes contacting said cell with an agent, e.g., an antibody, that selectively binds to 48149, under conditions that allow binding to occur. In one embodiment, the agent is coupled to, e.g., conjugated with, a moiety that allows separation (e.g., physical separation) of the bound agent-48149 complex.

In a preferred embodiment, the method includes determining resting from activated cells.

In yet another aspect, the invention features a method of evaluating, or identifying, an agent, e.g., an agent as described herein (e.g., a polypeptide, peptide, a peptide fragment, a peptidomimetic, a small molecule), for the ability to modulate, e.g. inhibit, the activity or expression of a 48149. Such agents are useful for treating or preventing cardiovascular disorders (e.g., atherosclerosis), metabolic disorders (e.g., a metabolic liver disorder), or non-neutrophil-mediated inflammatory disorders as described herein. The method includes:

providing a test agent, a 48149, or a cell expressing a 48149 (e.g., an atheroma-associated cell); and a 48149 substrate, e.g., an enkephalin or a chemokine;

contacting said test agent, said 48149 or said cell expressing said 48149, and said 48149 substrate, under conditions that allow an interaction (e.g., activity or expression) between said 48149 and said 48149 substrate to occur; and

determining whether said test agent modulates, e.g., inhibits, the expression or activity between said 48149 and said 48149 substrate,

wherein a change, e.g., a decrease, in the level of activity or expression between said 48149 and said 48149 substrate in the presence of the test agent relative to the activity or expression in the absence of the test agent, is indicative of modulation, e.g., inhibition, of the interaction between 48149 and the 48149 substrate.

In a preferred embodiment, the method further comprises the step of evaluating the test agent in an atheroma-associated cell, e.g., a macrophage, smooth muscle cell or endothelial cell, in vitro, or in vivo (e.g., in a subject, e.g., a patient having atherosclerosis), to thereby determine the effect of the test agent in the activity or expression of the 48149.

In a preferred embodiment, the contacting step occurs in vitro or ex vivo. For example, a sample, e.g., a blood sample, is obtained from the subject. Preferably, the sample contains an atheroma-associated cell, e.g., a macrophage, an endothelial cell or a smooth muscle cell.

In a preferred embodiment, the 48149 substrate is a fluorogenic substrate, e.g., an FITC-conjugated small peptide. Preferably, the fluorogenic substrate releases fluorescence upon cleavage.

In some embodiments, the 48149 substrate may interact with, e.g., bind to, other aminopeptidases.

In a preferred embodiment, the contacting step occurs in vivo. For example, by administering to the subject a detectably labeled agent that interacts with the 48149 nucleic acid or polypeptide, such that a signal is generated relative to the level of activity or expression of the 48149 nucleic acid or polypeptide.

In a preferred embodiment, the test agent is an inhibitor (partial or complete inhibitor) of the 48149 polypeptide activity or expression.

In preferred embodiments, the test agent is a peptide, a small molecule, e.g., a member of a combinatorial library (e.g., a peptide or organic combinatorial library, or a natural product library), or an antibody, or any combination thereof.

In additional preferred embodiments, the test agent is an antisense, a ribozyme, a triple helix molecule, or an atherosclerotic-associated nucleic acid, or any combination thereof.

In a preferred embodiment, a plurality of test agents, e.g., library members, is tested. In a preferred embodiment, the plurality of test agents, e.g., library members, includes at least 10, 102, 103, 104, 105, 106, 107, or 108 compounds. In a preferred

embodiment, the plurality of test agents, e.g., library members, share a structural or functional characteristic.

In a preferred embodiment, test agent is a peptide or a small organic molecule.

In a preferred embodiment, the method is performed in cell-free conditions (e.g., a reconstituted system).

In a preferred embodiment, the method further includes: contacting said agent with a test cell, or a test animal, to evaluate the effect of the test agent on the activity or expression of 48149.

In a preferred embodiment, the ability of the agent to modulate the activity or expression of 48149 is evaluated in a second system, e.g., a cell-free, cell-based, or an animal system.

In a preferred embodiment, the ability of the agent to modulate the activity or expression of 48149 is evaluated in a cell-based system, e.g., a two-hybrid assay.

In another aspect, the invention features a method of evaluating, or identifying, an agent, e.g., an agent as described herein (e.g., a polypeptide, peptide, a peptide fragment, a peptidomimetic, a small molecule), for the ability to modulate, e.g. enhance or decrease, transcription of an atherosclerotic- associated nucleic acid or polypeptide. The method includes:

contacting a cell, e.g., an atheroma-associated cell (e.g., a macrophage or a monocyte, an endothelial cell, or a smooth muscle cell), with a test agent; and

determining whether said test agent modulates, e.g., activates or inhibits, transcription of at least one atherosclerotic-associated nucleic acid, wherein a change, e.g., an increase or decrease, in the level of expression of said atherosclerotic-associated nucleic acid or polypeptide is indicative of a modulation, e.g., activation or inhibition, of the expression of atherosclerotic-associated nucleic acids.

In a preferred embodiment, the level of expression of at least one, two, three or four atherosclerotic-associated nucleic acid or polypeptide is evaluated. Preferably, the atherosclerosis-associated nucleic acid or polypeptide is 48149, preferably human 48149.

In a preferred embodiment, the level of expression of the at least one atherosclerotic-associated nucleic acid (e.g., a nucleic acid as described herein) is evaluated after stimulation of the cell, e.g., the atheroma-associated cell (e.g., a macrophage or a monocyte), with proinflammatory agent, e.g., a proinflammatory cytokine (e.g., IL-1b, CD40L, TNFa, or LPS).

In preferred embodiments, the test agent is a peptide, a small molecule, e.g., a member of a combinatorial library (e.g., a peptide or organic combinatorial library, or a natural product library), or an antibody, or any combination thereof.

In additional preferred embodiments, the test agent is an antisense, a ribozyme, a triple helix molecule, or an atherosclerotic-associated nucleic acid, or any combination thereof.

In a preferred embodiment, a plurality of test compounds, e.g., library members, is tested. In a preferred embodiment, the plurality of test compounds, e.g., library members, includes at least 10, 102, 103, 104, 105, 106, 107, or 108 compounds. In a preferred embodiment, the plurality of test compounds, e.g., library members, share a structural or functional characteristic.

In a preferred embodiment, test compound is a peptide or a small organic molecule.

In a preferred embodiment, the method is performed in cell-free conditions (e.g., a reconstituted system).

In a preferred embodiment, the method is performed in a cell, e.g., an atheroma-associated cell (e.g., a macrophage or a monocyte, an endothelial cell or a smooth muscle cell).

In a preferred embodiment, the method further includes: contacting said agent with a test cell, or a test animal, to evaluate the effect of the test agent on the transcription of the atherosclerotic-associated nucleic acid.

In a preferred embodiment, the ability of the agent to modulate transcription of the atherosclerotic-associated nucleic acid is evaluated in a second system, e.g., a cell-free, cell-based, or an animal system.

In a preferred embodiment, the ability of the agent to modulate transcription of the atherosclerotic-associated nucleic acid is evaluated in a cell-based system, e.g., a two-hybrid assay.

Also within the scope of the invention are agents identified using the methods described herein.

In another aspect, the invention features a pharmaceutical composition comprising an agent as described herein, and a pharmaceutically acceptable carrier. In one embodiment, the compositions of the invention, e.g., the pharmaceutical compositions, are administered in combination therapy, i.e., combined with other agents, e.g., therapeutic agents, that are useful for treating cardiovascular disorders, such as atherosclerosis. The

agent can be in the form of a prodrug, or a pharmaceutically acceptable salt or solvate thereof.

In another aspect, the invention features a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence. At least one address of the plurality has a capture probe that recognizes a 48149 molecule. In one embodiment, the capture probe is a nucleic acid, e.g., a probe complementary to a 48149 nucleic acid sequence. In another embodiment, the capture probe is a polypeptide, e.g., an antibody specific for 48149 polypeptides. Also featured is a method of analyzing a sample by contacting the sample to the aforementioned array and detecting binding of the sample to the array.

Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

Brief Description of the Drawings

Figure 1 depicts a hydropathy plot of human 48149. Relative hydrophobic residues are shown above the dashed horizontal line, and relative hydrophilic residues are below the dashed horizontal line. Numbers corresponding to positions in the amino acid sequence of human 48149 are indicated. Polypeptides of the invention include fragments which include: all or part of a hydrophobic sequence, i.e., a sequence above the dashed line, e.g., the sequence from about amino acid 10 to 35, from about 273 to 291, and from about 232 to 259 of SEQ ID NO:2; all or part of a hydrophilic sequence, i.e., a sequence below the dashed line, e.g., the sequence of from about amino acid 171 to 190, from about 250 to 272, and from about 321 to 339 of SEQ ID NO:2.

Figure 2 depicts an alignment of the membrane peptidase domain of human 48149 with a consensus amino acid sequence derived from a hidden Markov model (HMM) from PFAM. The upper sequence is the consensus amino acid sequence (SEQ ID NO:4), while the lower amino acid sequence corresponds to amino acids 76 to 480 of SEQ ID NO:2.

Figure 3 depicts increased 48149 mRNA expression in macrophage cells that have been treated with oxidized low density lipoprotein (OxLDL) for either 6 hours, 48 hours, or 96 hours, as well as macrophage cells grown in the absence of OxLDL for either 6 hours or 48 hours.

Figure 4 depicts the expression of 48149 mRNA in macrophages that have been incubated with modified low density lipoproteins (LDLs) in an 8 hour lipid-loading assay. The uptake of lipids and cholesterol by macrophages during a lipid-loading assay phenotypically mimics the formation of foam cells during atherogenesis (i.e., the formation of a atherosclerotic plaque). Thus, genes that are up-regulated during the course of the lipid-loading assay are thought to be involved in atherogenesis. The data demonstrates very high expression of 48149 mRNA when macrophages are exposed to mildly oxidized LDL (miOxLDL) for 8 hours, and high expression of 48149 mRNA when macrophages are exposed to moderately oxidized LDL (moOxLDL) or acetylated LDL (acLDL) for 8 hours. In contrast, 48149 mRNA expression is much lower in macrophage cells (monocytes) that have not been exposed to any LDL particles.

Figures 5A and 5B depicts the expression of mouse 48149 mRNA in ApoE knockout mice, which are genetically predisposed to developing atherosclerosis. In Figure 5A, the expression of 48149 in the arteries of ApoE knockout mice that were 8 to 30 weeks (wk) old is shown. As the mice age, the expression of 48149 is elevated in arterial regions where atherosclerotic plaques tend to form, i.e., the arch portion of the aorta, as compared to arterial regions where atherosclerotic plaques do not tend to form, i.e., the abdominal (descending) portion of the aorta (abd). In Figure 5B, the expression of 48149 mRNA in artery samples obtained from 38 week (wk) old ApoE knockout mice is displayed. In all mice analyzed, 48149 expression is elevated in aortic arch tissue as compared to aortic abd tissue, indicating a correlation between the formation of atherosclerotic lesions and the expression of 48149.

Figure 6 depicts the expression of 48149 mRNA in monkeys when they are eating a a diet (chow) that contains normal levels of dietary fatty acids, including saturated fatty acids, as compared to when they are eating a diet that contains only poly-unsaturated fatty acids. Six of the eight monkeys tested display elevated levels of 48149 expression when given the chow diet. This data is also presented in tabular form in Table 5 of Example 2.

The human 48149 sequence (see SEQ ID NO:1, as recited in Example 1), which is approximately 3494 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 2904 nucleotides, including the termination codon. The coding sequence encodes a 967 amino acid protein (see SEQ ID NO:2, as recited in Example 1).

Human 48149 contains the following regions or other structural features:

a aminopeptidase domain (PFAM Accession Number PF01433) located at about amino acid residues 76 to 480 of SEQ ID NO:2;

a neutral zinc protease motif (PS00142) located at about amino acid residues 385 to 394 of SEQ ID NO:2;

one transmembrane domain located at about amino acid residues 11 to 33 of SEQ ID NO:2; and

ten predicted N-glycosylation sites (PS00001) located at about amino acids 128 to 131, 234 to 237, 265 to 268, 319 to 322, 527 to 530, 573 to 576, 625 to 628, 681 to 684, 735 to 738, and 818 to 821 of SEQ ID NO:2.

For general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer *et al.* (1997) *Protein* 28:405-420 and http://www.psc.edu/general/ software/packages/pfam/pfam.html.

The 48149 protein contains a significant number of structural characteristics in common with members of the aminopeptidase family. In fact, human 48149 is identical to the sequence of human Myeloid Plasma Membrane Glycoprotein DC13, which is also known as Aminopeptidase N (Look et al. (1989), *J. Clin. Invest.* 83(4):1299-307). The term "family" when referring to the protein and nucleic acid molecules of the invention means two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin, or alternatively, can contain homologues of non-human origin, e.g., rat or mouse proteins. Members of a family can also have common functional characteristics.

An aminopeptidase family of proteins is characterized by a common fold. The fold is typically located extracellularly, and can be associated with the plasma membrane via a transmembrane domain. Aminopeptidase family members catalyze the removal of NH_2-terminal amino acids from peptides, typically with a preference for neutral, basic, or acidic residues, although aminopetidases can have broad specificity despite showing such a preference. The natural substrates appear to be peptides rather than large proteins, but aminopeptidases are more effective in the removal of residues from oligopeptides rather than dipeptides. Aminopeptidases are believed to play an important role in the final stages of the digestion of small peptides, e.g., in the intestine, and have been postulated to function in the hydrolytic inactivation of regulatory peptides including those involved in

signal transduction at the cell membrane. The aminopeptidase domain typically contains a neutral zinc protease motif that serves as both a metal-binding site and the catalytic site involved in proteolysis.

A 48149 polypeptide can include an "aminopeptidase domain" or regions homologous with an "aminopeptidase domain".

As used herein, the term "aminopeptidase domain" includes an amino acid sequence of about 200 to 600 amino acid residues in length and having a bit score for the alignment of the sequence to the membrane peptidase domain profile (Pfam HMM) of at least 300. Preferably, a membrane peptidase domain includes at least about 200 to 500 amino acids, more preferably about 300 to 450 amino acid residues, or about 400 amino acids and has a bit score for the alignment of the sequence to the membrane peptidase domain (HMM) of at least 400, preferably 550, or more preferably 700 or greater. The membrane peptidase domain (HMM) has been assigned the PFAM Accession Number PF01433 (http;//genome.wustl.edu/Pfam/.html). An alignment of the membrane peptidase domain (amino acids 76 to 480 of SEQ ID NO:2) of human 48149 with a consensus amino acid sequence (SEQ ID NO:4) derived from a hidden Markov model is depicted in Figure 2.

In a preferred embodiment 48149 polypeptide or protein has a "membrane peptidase domain" or a region which includes at least about 200 to 600, more preferably about 200 to 500, or 300 to 450 amino acid residues, and has at least about 50%, 60%, 70% 80% 90% 95%, 99%, or 100% homology with an "aminopeptidase domain," e.g., the aminopeptidase domain of human 48149 (e.g., residues 76 to 480 of SEQ ID NO:2).

To identify the presence of an "aminopeptidase" domain in a 48149 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against the PFAM database of HMMs (e.g., the PFAM database, release 2.1) using the default parameters (http://www.sanger.ac.uk/Software/Pfam/HMM_search). For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (e.g., to 8 bits). A description of the PFAM database can be found in Sonhammer *et al.* (1997) *Proteins* 28(3):405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.*(1990) *Meth. Enzymol.* 183:146-159; Gribskov *et al.*(1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.*(1994) *J. Mol. Biol.*

235:1501-1531; and Stultz *et al.*(1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference. A search was performed against the PFAM HMM database resulting in the identification of an "aminopeptidase" domain in the amino acid sequence of human 48149 at about residues 76 to 480 of SEQ ID NO:2 (see Figure 2).

In preferred embodiments, a 48149 protein includes at least one neutral zinc protease motif. As used herein, a "neutral zinc protease motif" includes a sequence of at least five amino acid residues defined by the sequence: [GSTALIVN]-X-X-H-E-[LIVMFYW]-[DEHRKP]-H-x-[LIVMFYWGSPQ] (SEQ ID NO:5). A neutral zinc protease motif, as defined, can be involved in the coordination of a zinc ion and the enzymatic hydrolysis of of a polypeptide substrate. More preferably, a neutral zinc protease motif includes 7, or even more preferably 10 amino acid residues. Neutral zinc protease motifs have been described in, e.g., Bode et al. (1996), Zoology 99:237-46, the contents of which are incorporated herein by reference. Human 48149 contains a neutral zinc protease motif located at about amino acid residues 385 to 394 of SEQ ID NO:2.

In a preferred embodiment, a 48149 polypeptide or protein has at least one neutral zinc protease motif, or a region which includes at least 5, 7, or preferably 10 amino acid residues and has at least 70%, 80%, 90%, or 100% homology with a "neutral zinc protease motif" of human 48149, e.g., about amino acid residues 385 to 394 of SEQ ID NO:2.

In some embodiments, a 48149 protein includes at least one transmembrane domain. As used herein, the term "transmembrane domain" includes an amino acid sequence of about 15 amino acid residues in length that spans a phospholipid membrane. More preferably, a transmembrane domain includes about at least 20, 22 or 25 amino acid residues and spans a phospholipid membrane. Transmembrane domains are rich in hydrophobic residues, and typically have an a-helical structure. In a preferred embodiment, at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, e.g., leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains are described in, for example, Zagotta W.N. et al, (1996) Annual Rev. Neurosci. 19: 235-63, the contents of which are incorporated herein by reference.

In preferred embodiments, a 48149 polypeptide or protein contains at least one transmembrane domain or a region which includes at least 20, 22 or 25 amino acid residues and has at least about 70%, 80%, 90%, 95%, 99%, or 100% homology with a "transmembrane domain", e.g., at least one transmembrane domain of human 48149 (e.g., from about amino acid residues 11 to 33 of SEQ ID NO:2).

A 48149 family member can include at least one aminopeptidase domain. Furthermore, a 48149 family member can include at least one neutral zinc protease motif (PS00142); at least one transmembrane domain; and at least one, two, three, four, five, six, seven, eight, nine, and preferably ten N-glycosylation sites (PS00001).

As the 48149 polypeptides of the invention may modulate 48149-mediated activities, they may be useful as of for developing novel diagnostic and therapeutic agents for 48149-mediated or related disorders, as described below.

As used herein, a "48149 activity", "biological activity of 48149" or "functional activity of 48149", refers to an activity exerted by a 48149 protein, polypeptide or nucleic acid molecule. For example, a 48149 activity can be an activity exerted by 48149 in a physiological milieu on, e.g., a 48149-responsive cell or on a 48149 substrate, e.g., a protein substrate. A 48149 activity can be determined *in vivo* or *in vitro*. In one embodiment, a 48149 activity is a direct activity, such as an association with a 48149 target molecule. A "target molecule" or "binding partner" is a molecule with which a 48149 protein binds or interacts in nature. In an exemplary embodiment, a 48149 protein is an enzyme that hydrolyzes polypeptide substrates, e.g., small polypeptides that are involved in signaling, e.g, enkephalin or the chemokine MCP-1.

A 48149 activity can also be an indirect activity, e.g., a cellular signaling activity mediated by interaction of the 48149 protein with a 48149-binding protein or a signaling molecule that is a 48149 substrate. The features of the 48149 molecules of the present invention can provide similar biological activities as aminopeptidase family members. For example, the 48149 proteins of the present invention can have one or more of the following activities: (1) hydrolysis of polypeptides; (2) catalysis of NH_2-terminal amino acids from polypeptides; and (3) digestion of polypeptides involved in cellular signaling. Furthermore, the 48149 proteins of the invention can have a function as described above in the tissues in which they are expressed. Thus, the 48149 proteins can have one or more of the following activities: (4) modulation of macrophage activity; (5) modulation of macrophage activation; (6) modulation of leukocyte acitivation; (7) modulation of fatty acid and cholesterol uptake by macrophages; (8) promotion of atherosclerotic plaque formation; or (9) modulation of fatty acid and cholesterol metabolism.

Thus, the 48149 molecules can act as novel diagnostic targets and therapeutic agents for controlling cardiovascular disorders, metabolic disorders, e.g., relating to lipid or cholesterol metabolism, e.g., in the liver, or non-neutrophil-mediated inflammatory disorders. The term "cardiovascular disorders" or "disease" includes heart disorders, as

well as disorders of the blood vessels of the circulation system caused by, e.g., abnormally high concentrations of lipids in the blood vessels.

Examples of cardiovascular disorders or diseases include e.g., atherosclerosis, aneurism, thrombosis, heart failure, ischemic heart disease, angina pectoris, myocardial infarction, sudden cardiac death, hypertensive heart disease; non-coronary vessel disease, such as arteriolosclerosis, small vessel disease, nephropathy, hypertriglyceridemia, hypercholesterolemia, hyperlipidemia, hypertension; or a cardiovascular condition associated with interventional procedures ("procedural vascular trauma"), such as restenosis following angioplasty, placement of a shunt, stent, synthetic or natural excision grafts, indwelling catheter, valve or other implantable devices.

Disorders involving the heart, include but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, aneurism, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and

atresia, and aortic stenosis and atresia, asthma, emphysema and chronic pulmonary disease and disorders involving cardiac transplantation.

Disorders involving blood vessels include, but are not limited to, responses of vascular cell walls to injury, such as endothelial dysfunction and endothelial activation and intimal thickening; vascular diseases including, but not limited to, congenital anomalies, such as arteriovenous fistula, atherosclerosis, and hypertensive vascular disease, such as hypertension; inflammatory disease--the vasculitides, such as giant cell (temporal) arteritis, Takayasu arteritis, polyarteritis nodosa (classic), Kawasaki syndrome (mucocutaneous lymph node syndrome), microscopic polyanglitis (microscopic polyarteritis, hypersensitivity or leukocytoclastic anglitis), Wegener granulomatosis, thromboanglitis obliterans (Buerger disease), vasculitis associated with other disorders, and infectious arteritis; Raynaud disease; aneurisms and dissection, such as abdominal aortic aneurisms, syphilitic (luetic) aneurisms, and aortic dissection (dissecting hematoma); disorders of veins and lymphatics, such as varicose veins, thrombophlebitis and phlebothrombosis, obstruction of superior vena cava (superior vena cava syndrome), obstruction of inferior vena cava (inferior vena cava syndrome), and lymphangitis and lymphedema; tumors, including benign tumors and tumor-like conditions, such as hemangioma, lymphangioma, glomus tumor (glomangioma), vascular ectasias, and bacillary angiomatosis, and intermediate-grade (borderline low-grade malignant) tumors, such as Kaposi sarcoma and hemangioendothelioma, and malignant tumors, such as angiosarcoma and hemangiopericytoma; and pathology of therapeutic interventions in vascular disease, such as balloon angioplasty and related techniques and vascular replacement, such as coronary artery bypass graft surgery.

Small vessel disease includes, but is not limited to, vascular insufficiency in the limbs, peripheral neuropathy and retinopathy, e.g., diabetic retinopathy.

Endothelial cell disorders include disorders characterized by aberrant, unregulated, or unwanted endothelial cell activity, e.g., proliferation, migration, angiogenesis, or vascularization; or aberrant expression of cell surface adhesion molecules or genes associated with angiogenesis, e.g., TIE-2, FLT and FLK. Endothelial cell disorders include tumorigenesis, tumor metastasis, psoriasis, diabetic retinopathy, endometriosis, Grave's disease, ischemic disease (e.g., atherosclerosis), and chronic inflammatory diseases (e.g., rheumatoid arthritis).

Liver disorders include, but are not limited to, liver injury associated with a hepatic manifestation of a vascular disorder such as obstruction of either the intrahepatic or

extrahepatic bile flow or an alteration in hepatic circulation resulting, e.g., from chronic heart failure, veno-occlusive disease, portal vein thrombosis or Budd-Chiari syndrome. Other liver disorders include, but are not limited to disorders associated with an accumulation in the liver of fibrous tissue, such as that resulting from an imbalance between production and degradation of the extracellular matrix accompanied by the collapse and condensation of preexisting fibers; and hepatocellular necrosis or injury induced by a wide variety of agents including processes which disturb homeostasis, such as an inflammatory process, tissue damage resulting from toxic injury or altered hepatic blood flow, and infections (e.g., bacterial, viral and parasitic). For example, such disorders include hepatic injury, such as portal hypertension or hepatic fibrosis. Additional liver disorder include liver fibrosis attributed to inborn errors of metabolism, for example, fibrosis resulting from a storage disorder such as Gaucher's disease (lipid abnormalities) or a glycogen storage disease, A1-antitrypsin deficiency; a disorder mediating the accumulation (e.g., storage) of an exogenous substance, for example, hemochromatosis (iron-overload syndrome) and copper storage diseases (Wilson's disease), disorders resulting in the accumulation of a toxic metabolite (e.g., tyrosinemia, fructosemia and galactosemia) and peroxisomal disorders (e.g., Zellweger syndrome).

As used herein, the term "atherosclerosis" is intended to have its clinical meaning. This term refers to a cardiovascular condition occurring as a result of lesion formation in the arterial walls. The narrowing is due to the formation of plaques or streaks in the inner lining of the arteries. These plaques consist of foam cells filled with modified low-density lipoproteins, oxidized-LDL, decaying smooth muscle cells, fibrous tissue, clumps of blood platelets, cholesterol, and sometimes calcium. They tend to form in regions of disturbed blood flow and are found most often in people with high concentrations of cholesterol in the bloodstream. The number and thickness of plaques increase with age, causing loss of the smooth lining of the blood vessels and encouraging the formation of thrombi (blood clots). Sometimes fragments of thrombi break off and form emboli, which travel through the bloodstream and block smaller vessels. The thrombi or emboli can restrict the blood supply to the heart, brain, kidney and other organs eventually leading to end organ damage or death. The major causes of atherosclerosis are hypercholesterolemia, hypoalphoproteinemia, and hyperlipidemia marked by high circulating triglycerides in the blood. These lipids are deposited in the arterial walls, obstructing the blood flow and forming atherosclerotic plaques leading to death.

As used herein the term "hypercholesterolemia" is a condition with elevated levels of circulating total cholesterol, LDL-cholesterol and VLDL-cholesterol as per the guidelines of the Expert Panel Report of the National Cholesterol Educational Program (NCEP) of Detection, Evaluation of Treatment of high cholesterol in adults (see, Arch. Int. Med. (1988) 148, 36-39).

As used herein the term "hyperlipidemia" or "hyperlipemia" is a condition where the blood lipid parameters are elevated in the blood. This condition manifests an abnormally high concentration of fats. The lipid fractions in the circulating blood are, total cholesterol, low density lipoproteins, very low density lipoproteins and triglycerides.

As used herein the term "lipoprotein" such as VLDL, LDL and HDL, refers to a group of proteins found in the serum, plasma and lymph and are important for lipid transport. The chemical composition of each lipoprotein differs in that the HDL has a higher proportion of protein versus lipid, whereas the VLDL has a lower proportion of protein versus lipid.

As used herein, the term "triglyceride" means a lipid or neutral fat consisting of glycerol combined with three fatty acid molecules.

As used herein the term "xanthomatosis" is a disease evidenced by a yellowish swelling or plaques in the skin resulting from deposits of fat. The presence of xanthomas are usually accompanied by raised blood cholesterol levels.

As used herein the terms "apolipoprotein B", "apoprotein B", or "Apo B" and "apolipoprotein E", "apoprotein E", or "Apo E" refer to protein components of LDL cholesterol transport lipoprotein particles. Cholesterol synthesized de novo is transported from the liver and intestine to peripheral tissues in the form of lipoprotein particles. Most of the apolipoprotein B is secreted into the circulatory system as VLDL.

As used herein the term "apolipoprotein A" or "apoprotein A" or "Apo A" refers to the protein component of the HDL cholesterol transport proteins.

"Procedural vascular trauma" includes the effects of surgical/medical-mechanical interventions into mammalian vasculature, but does not include vascular trauma due to the organic vascular pathologies listed hereinabove, or to unintended traumas, such as due to an accident. Thus, procedural vascular traumas within the scope of the present treatment method include (1) organ grafting or transplantation, such as transplantation and grafting of heart, kidney, liver and the like, e.g., involving vessel anastomosis; (2) vascular surgery, such as coronary bypass surgery, biopsy, heart valve replacement, atheroectomy, thrombectomy, and the like; (3) transcatheter vascular therapies (TVT) including

angioplasty, e.g., laser angioplasty and PTCA procedures discussed hereinbelow, employing balloon catheters, or indwelling catheters; (4) vascular grafting using natural or synthetic materials, such as in saphenous vein coronary bypass grafts, dacron and venous grafts used for peripheral arterial reconstruction, etc.; (5) placement of a mechanical shunt, such as a PTFE hemodialysis shunt used for arteriovenous communications; and (6) placement of an intravascular stent, which may be metallic, plastic or a biodegradable polymer. See U.S. patent application Ser. No. 08/389,712, filed Feb. 15, 1995, which is incorporated by reference herein. For a general discussion of implantable devices and biomaterials from which they can be formed, see H. Kambic et al., "Biomaterials in Artificial Organs", Chem. Eng. News, 30 (Apr. 14, 1986), the disclosure of which is incorporated by reference herein.

The term "agent" is used herein to denote a chemical compound, a mixture of chemical compounds, a biological macromolecule, a peptide, polypeptide (e.g., an antibody), small molecule, member of a combinatorial library, a peptide fragment, a peptidomimetic, or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian) cells or tissues. Agents can be evaluated for 48149 inhibitory activity by inclusion in screening assays described, for example, hereinbelow.

As used herein, a "therapeutically effective amount" of an agent refers to an amount of an 48149 inhibitor which is effective, upon single or multiple dose administration to the subject, e.g., a patient, at inhibiting 48149 expression or activity, or in prolonging the survival of the subject with a non-neutrophil-mediated inflammatory disorder, cardiovascular or endothelial disorder, or disorder beyond that expected in the absence of such treatment.

As used herein, "inhibiting the expression or activity" of 48149 refers to a reduction, blockade of the expression or activity, e.g., collagenolysis (e.g., degradation of collagen I) and does not necessarily indicate a total elimination of the 48149 expression or activity.

As used herein, "a prophylactically effective amount" of an agent refers to an amount of an 48149 inhibitor which is effective, upon single- or multiple-dose administration to the patient, in preventing or delaying the occurrence of the onset or recurrence of a disorder as described herein.

The terms "induce", "inhibit", "potentiate", "elevate", "increase", "decrease" or the like, e.g., which denote quantitative differences between two states, refer to at least

statistically significant differences between the two states. For example, "an amount effective to inhibit the activity or expression of 48149 means that the level of activity or expression of 48149 in a treated sample will differ statistically significantly from the level of 48149 activity or expression in untreated cells. Such terms are applied herein to, for example, levels of expression, and levels activity.

In some embodiments, the therapeutic and prophylactic uses of the agents of the invention, further include the administration of a second agent, e.g., a non-specific MMP inhibitor, a cholesterol lowering agent, or an interventional as a combination therapies. The term "in combination" in this context means that the agents, or agent and procedures are given substantially contemporaneously, either simultaneously or sequentially. If given sequentially, at the onset of administration of the second agent or procedure, the first agent is preferably still detectable at effective concentrations at the site of treatment. For example, the combination therapy can include an agent of the present invention coformulated with, and/or coadministered with, one or more additional therapeutic agents, e.g., one or more MMP inhibitors, cytotoxic or cytostatic agents and/or immunosuppressants. For example, the agents of the invention or antibody binding fragments thereof may be coformulated with, and/or coadministered with, one or more additional MMP inhibitors.

The agents of the invention may be administered in combination with lipid lowering agents. Current combination therapy therapies using combinations of niacin and statins are being used with positive results to treat hyperlipidemia (Guyton, JR. (1999) Curr Cardiol Rep. 1(3):244-250; Otto, C. et al. (1999) Internist (Berl) 40(12):1338-45). Other useful drug combinations include those derived by addition of fish oil, bile acid binding resins, or stanol esters, as well as nonstatin combinations such as niacin-resin or fibrate-niacin (Guyton, JR. (1999) supra). For examples of dosages and administration schedules of the cholesterol lowering agents, the teachings of Guyton, JR. (1999), supra, Otto, C. et al. (1999), supra, Guyton, JR et al. (1998), *Am J Cardiol* 82(12A):82U-86U; Guyton, JR et al. (1998), *Am J Cardiol.* 82(6):737-43; Vega, GL et al. (1998), *Am J. Cardiol.* 81(4A):36B-42B; Schectman, G. (1996), *Ann Intern Med.* 125(12):990-1000; Nakamura, H. et al. (1993) Nippon Rinsho 51(8):2101-7; Goldberg, A. et al. (2000), *Am J Cardiol* 85(9):1100-5; Morgan, JM et al. (1996), *J Cardiovasc. Pharmac. Ther.* 1(3):195-202; Stein, EA et al. (1996), *J Cardiovasc Pharmacol Ther* 1(2):107-116; and Goldberg, AC (1998), *Am J Cardiol* 82(12A):35U-41U, are expressly incorporated by reference.

As used herein, "cholesterol lowering agents" include agents which are useful for lowering serum cholesterol such as for example bile acid sequestering resins (e.g. colestipol hydrochloride or cholestyramine), fish oil, stanol esters, an ApoAII-lowering agent, a VLDL lowering agent, an ApoAI-stimulating agent, fibric acid derivatives (e.g. clofibrate, fenofibrate, or gemfibrozil), thiazolidenediones (e.g. troglitazone, pioglitazone, ciglitazone, englitazone, rosiglitazone), or HMG-CoA reductase inhibitors (e.g. statins, such as fluvastatin sodium, lovastatin, pravastatin sodium, simvastatin, atorvastatin calcium, cerivastatin), as well as nicotinic acid, niacin, or probucol.

As used herein, the term "VLDL-lowering agent" includes an agent which decreases the hepatic synthesis of triglyceride-rich lipoproteins or increases the catabolism of triglyceride-rich lipoproteins, e.g., fibrates such as gemfibrozil, or the statins, increases the expression of the apoE-mediated clearance pathway, or improves insulin sensitivity in diabetics, e.g., the thiazolidene diones.

The 48149 protein, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO:2 thereof are collectively referred to as "polypeptides or proteins of the invention" or "48149 polypeptides or proteins". Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as "nucleic acids of the invention" or "48149 nucleic acids." 48149 molecules refer to 48149 nucleic acids, polypeptides, and antibodies.

As used herein, the term "nucleic acid molecule" includes DNA molecules (e.g., a cDNA or genomic DNA), RNA molecules (e.g., an mRNA) and analogs of the DNA or RNA. A DNA or RNA analog can be synthesized from nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated nucleic acid molecule" or "purified nucleic acid molecule" includes nucleic acid molecules that are separated from other nucleic acid molecules present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which naturally flank the nucleic acid

molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

As used herein, the term "hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions" describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6, which is incorporated by reference. Aqueous and nonaqueous methods are described in that reference and either can be used. Specific hybridization conditions referred to herein are as follows: 1) low stringency hybridization conditions in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by two washes in 0.2X SSC, 0.1% SDS at least at 50 °C (the temperature of the washes can be increased to 55°C for low stringency conditions); 2) medium stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C; 3) high stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C; and preferably 4) very high stringency hybridization conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Very high stringency conditions (4) are the preferred conditions and the ones that should be used unless otherwise specified.

Preferably, an isolated nucleic acid molecule of the invention that hybridizes under a stringency condition described herein to the sequence of SEQ ID NO:1 or SEQ ID NO:3, corresponds to a naturally-occurring nucleic acid molecule.

As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature. For example a naturally occurring nucleic acid molecule can encode a natural protein.

As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include at least an open reading frame encoding a 48149 protein. The gene can optionally further include non-coding sequences, e.g., regulatory sequences and introns. Preferably, a gene encodes a mammalian 48149 protein or derivative thereof.

An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when

chemically synthesized. "Substantially free" means that a preparation of 48149 protein is at least 10% pure. In a preferred embodiment, the preparation of 48149 protein has less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-48149 protein (also referred to herein as a "contaminating protein"), or of chemical precursors or non-48149 chemicals. When the 48149 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The invention includes isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 48149 without abolishing or substantially altering a 48149 activity. Preferably the alteration does not substantially alter the 48149 activity, e.g., the activity is at least 20%, 40%, 60%, 70% or 80% of wild-type. An "essential" amino acid residue is a residue that, when altered from the wild-type sequence of 48149, results in abolishing a 48149 activity such that less than 20% of the wild-type activity is present. For example, conserved amino acid residues in 48149 are predicted to be particularly unamenable to alteration.

A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 48149 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 48149 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 48149 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1 or SEQ ID NO:3, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

As used herein, a "biologically active portion" of a 48149 protein includes a fragment of a 48149 protein which participates in an interaction, e.g., an intramolecular or an inter-molecular interaction. An inter-molecular interaction can be a specific binding interaction or an enzymatic interaction (e.g., the interaction can be transient and a covalent bond is formed or broken). An inter-molecular interaction can be between a 48149 molecule and a non-48149 molecule or between a first 48149 molecule and a second 48149 molecule (e.g., a dimerization interaction). Biologically active portions of a 48149 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the 48149 protein, e.g., the amino acid sequence shown in SEQ ID NO:2, which include less amino acids than the full length 48149 proteins, and exhibit at least one activity of a 48149 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 48149 protein, e.g., hydolysis of polypeptide substrates, e.g., small peptides, or hydrolytic inactivation of regulatory peptides, e.g., peptides involved in signaling events at the cell membrane. A biologically active portion of a 48149 protein can be a polypeptide which is, for example, 10, 25, 50, 100, 200 or more amino acids in length. Biologically active portions of a 48149 protein can be used as targets for developing agents which modulate a 48149 mediated activity, e.g., hydolysis of polypeptide substrates, e.g., small peptides, or hydrolytic inactivation of regulatory peptides, e.g., peptides involved in signaling events at the cell membrane.

Calculations of homology or sequence identity between sequences (the terms are used interchangeably herein) are performed as follows.

To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology").

The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch ((1970) *J. Mol. Biol.* 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used unless otherwise specified) are a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller ((1989) CABIOS, 4:11-17) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 48149 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 48149 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25:3389-3402. When utilizing BLAST and Gapped BLAST

programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

Particularly preferred 48149 polypeptides of the present invention have an amino acid sequence substantially identical to the amino acid sequence of SEQ ID NO:2. In the context of an amino acid sequence, the term "substantially identical" is used herein to refer to a first amino acid that contains a sufficient or minimum number of amino acid residues that are i) identical to, or ii) conservative substitutions of aligned amino acid residues in a second amino acid sequence such that the first and second amino acid sequences can have a common structural domain and/or common functional activity. For example, amino acid sequences that contain a common structural domain having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%. 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:2 are termed substantially identical.

In the context of nucleotide sequence, the term "substantially identical" is used herein to refer to a first nucleic acid sequence that contains a sufficient or minimum number of nucleotides that are identical to aligned nucleotides in a second nucleic acid sequence such that the first and second nucleotide sequences encode a polypeptide having common functional activity, or encode a common structural polypeptide domain or a common functional polypeptide activity. For example, nucleotide sequences having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%. 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:1 or 3 are termed substantially identical.

"Misexpression or aberrant expression", as used herein, refers to a non-wildtype pattern of gene expression at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over- or under-expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of altered, e.g., increased or decreased, expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, translated amino acid sequence, post-transitional modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

"Subject," as used herein, refers to human and non-human animals. The term "non-human animals" of the invention includes all vertebrates, e.g., mammals, such as non-human primates (particularly higher primates), sheep, dog, rodent (e.g., mouse or rat), guinea pig, goat, pig, cat, rabbits, cow, and non-mammals, such as chickens, amphibians, reptiles, etc. In a preferred embodiment, the subject is a human. In another embodiment, the subject is an experimental animal or animal suitable as a disease model.

A "purified preparation of cells", as used herein, refers to an in vitro preparation of cells. In the case cells from multicellular organisms (e.g., plants and animals), a purified preparation of cells is a subset of cells obtained from the organism, not the entire intact organism. In the case of unicellular microorganisms (e.g., cultured cells and microbial cells), it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

Various aspects of the invention are described in further detail below.

Isolated Nucleic Acid Molecules

In one aspect, the invention provides, an isolated or purified, nucleic acid molecule that encodes a 48149 polypeptide described herein, e.g., a full-length 48149 protein or a fragment thereof, e.g., a biologically active portion of 48149 protein. Also included is a nucleic acid fragment suitable for use as a hybridization probe, which can be used, e.g., to identify a nucleic acid molecule encoding a polypeptide of the invention, 48149 mRNA, and fragments suitable for use as primers, e.g., PCR primers for the amplification or mutation of nucleic acid molecules.

In one embodiment, an isolated nucleic acid molecule of the invention includes the nucleotide sequence shown in SEQ ID NO:1, or a portion of any of these nucleotide sequences. In one embodiment, the nucleic acid molecule includes sequences encoding the human 48149 protein (i.e., "the coding region" of SEQ ID NO:1, as shown in SEQ ID NO:3), as well as 5' untranslated sequences. Alternatively, the nucleic acid molecule can include only the coding region of SEQ ID NO:1 (e.g., SEQ ID NO:3) and, e.g., no flanking sequences which normally accompany the subject sequence. In another embodiment, the nucleic acid molecule encodes a sequence corresponding to a fragment of the protein from about amino acid 76 to 480.

In another embodiment, an isolated nucleic acid molecule of the invention includes a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ

ID NO:1 or SEQ ID NO:3, or a portion of any of these nucleotide sequences. In other embodiments, the nucleic acid molecule of the invention is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, such that it can hybridize (e.g., under a stringency condition described herein) to the nucleotide sequence shown in SEQ ID NO:1 or 3, thereby forming a stable duplex.

In one embodiment, an isolated nucleic acid molecule of the present invention includes a nucleotide sequence which is at least about: 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the entire length of the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, or a portion, preferably of the same length, of any of these nucleotide sequences.

48149 Nucleic Acid Fragments

A nucleic acid molecule of the invention can include only a portion of the nucleic acid sequence of SEQ ID NO:1 or 3. For example, such a nucleic acid molecule can include a fragment which can be used as a probe or primer or a fragment encoding a portion of a 48149 protein, e.g., an immunogenic or biologically active portion of a 48149 protein. A fragment can comprise those nucleotides of SEQ ID NO:1, which encode a membrane peptidase domain of human 48149. The nucleotide sequence determined from the cloning of the 48149 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other 48149 family members, or fragments thereof, as well as 48149 homologues, or fragments thereof, from other species.

In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5' or 3' noncoding region. Other embodiments include a fragment which includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly fragments thereof which are at least 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or more amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

A nucleic acid fragment can include a sequence corresponding to a domain, region, or functional site described herein. A nucleic acid fragment can also include one or more domain, region, or functional site described herein. Thus, for example, a 48149 nucleic acid

fragment can include a sequence corresponding to an aminopeptidase domain, a hydrophobic signal sequence that is retained and functions as the transmembrane domain, or an extracellular carboxyterminal domain.

48149 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under a stringency condition described herein to at least about 7, 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense or antisense sequence of SEQ ID NO:1 or SEQ ID NO:3, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1 or SEQ ID NO:3. Preferably, an oligonucleotide is less than about 200, 150, 120, or 100 nucleotides in length.

In one embodiment, the probe or primer is attached to a solid support, e.g., a solid support described herein.

One exemplary kit of primers includes a forward primer that anneals to the coding strand and a reverse primer that anneals to the non-coding strand. The forward primer can anneal to the start codon, e.g., the nucleic acid sequence encoding amino acid residue 1 of SEQ ID NO:2. The reverse primer can anneal to the ultimate codon, e.g., the codon immediately before the stop codon, e.g., the codon encoding amino acid residue 968 of SEQ ID NO:2. In a preferred embodiment, the annealing temperatures of the forward and reverse primers differ by no more than 5, 4, 3, or 2°C.

In a preferred embodiment the nucleic acid is a probe which is at least 10, 12, 15, 18, 20 and less than 200, more preferably less than 100, or less than 50, nucleotides in length. It should be identical, or differ by 1, or 2, or less than 5 or 10 nucleotides, from a sequence disclosed herein. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

A probe or primer can be derived from the sense or anti-sense strand of a nucleic acid which encodes: an aminopeptidase domain, e.g., located at about amino acid residues 76 to 480 or SEQ ID NO:2; a sequence that includes a hydrophobic signal sequence that is retained and functions as the transmembrane domain, e.g., located at about amino acid residues 1 to 75 of SEQ ID NO:2; or an extracellular carboxyterminal domain, e.g., located at about amino acid residues 481 to 967 of SEQ ID NO:2.

In another embodiment a set of primers is provided, e.g., primers suitable for use in a PCR, which can be used to amplify a selected region of a 48149 sequence, e.g., a

domain, region, site or other sequence described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differs by one base from a sequence disclosed herein or from a naturally occurring variant. For example, primers suitable for amplifying all or a portion of any of the following regions are provided: an aminopeptidase domain, e.g., located at about amino acid residues 76 to 480 or SEQ ID NO:2; a sequence that includes a hydrophobic signal sequence that is retained and functions as the transmembrane domain, e.g., located at about amino acid residues 1 to 75 of SEQ ID NO:2; or an extracellular carboxyterminal domain, e.g., located at about amino acid residues 481 to 967 of SEQ ID NO:2.

A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

A nucleic acid fragment encoding a "biologically active portion of a 48149 polypeptide" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1 or 3, which encodes a polypeptide having a 48149 biological activity (e.g., the biological activities of the 48149 proteins are described herein), expressing the encoded portion of the 48149 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 48149 protein. For example, a nucleic acid fragment encoding a biologically active portion of 48149 includes a membrane peptidase domain, e.g., amino acid residues about 76 to 480 of SEQ ID NO:2. A nucleic acid fragment encoding a biologically active portion of a 48149 polypeptide, may comprise a nucleotide sequence which is greater than 300 or more nucleotides in length.

In preferred embodiments, a nucleic acid includes a nucleotide sequence which is about 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3200, 3400, or more nucleotides in length and hybridizes under a stringency condition described herein to a nucleic acid molecule of SEQ ID NO:1, or SEQ ID NO:3.

In a preferred embodiment, a nucleic acid fragment differs by at least 1, 2, 3, 10, 20, or more nucleotides from the sequence of Genbank accession number M22324. Differ can include differing in length or sequence identity. For example, a nucleic acid fragment can: include one or more nucleotides from SEQ ID NO: 1 or SEQ ID NO:3 outside the region of nucleotides 1 to 3494 of the sequence of Genbank accession number M22324; not include all of the nucleotides of the sequence of Genbank accession number M22324, e.g., can be one or more nucleotides shorter (at one or both ends) than the sequence of

Genbank accession number M22324; or can differ by one or more nucleotides in the region of overlap.

48149 Nucleic Acid Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3. Such differences can be due to degeneracy of the genetic code (and result in a nucleic acid which encodes the same 48149 proteins as those encoded by the nucleotide sequence disclosed herein. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence which differs, by at least 1, but less than 5, 10, 20, 50, or 100 amino acid residues that shown in SEQ ID NO:2. If alignment is needed for this comparison the sequences should be aligned for maximum homology. The encoded protein can differ by no more than 5, 4, 3, 2, or 1 amino acid. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

Nucleic acids of the inventor can be chosen for having codons, which are preferred, or non-preferred, for a particular expression system. E.g., the nucleic acid can be one in which at least one codon, at preferably at least 10%, or 20% of the codons has been altered such that the sequence is optimized for expression in *E. coli*, yeast, human, insect, or CHO cells.

Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism) or can be non naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

In a preferred embodiment, the nucleic acid differs from that of SEQ ID NO: 1 or 3, e.g., as follows: by at least one but less than 10, 20, 30, or 40 nucleotides; at least one but less than 1%, 5%, 10% or 20% of the nucleotides in the subject nucleic acid. The nucleic acid can differ by no more than 5, 4, 3, 2, or 1 nucleotide. If necessary for this analysis the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%, at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more identical to the nucleotide sequence shown in SEQ ID NO:2 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under a stringency condition described herein, to the nucleotide sequence shown in SEQ ID NO 2 or a fragment of the sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of the 48149 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the 48149 gene.

Preferred variants include those that are correlated with aminopeptidase activity, e.g., hydrolysis of polypeptide substrates, e.g., short polypeptide substrates or signaling molecules that function at the cell surface, e.g., enkalphins, cytokines, or molecules of the TGFθ family.

Allelic variants of 48149, e.g., human 48149, include both functional and non-functional proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the 48149 protein within a population that maintain their aminopeptidase activity, e.g., their ability to hydrolyze polypeptide substrates, e.g., short polypeptide substrates or signaling molecules that function at the cell surface, e.g., enkalphins, cytokines, or molecules of the TGFθ family. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein. Non-functional allelic variants are naturally-occurring amino acid sequence variants of the 48149, e.g., human 48149, protein within a population that do not have aminopeptidase activity, e.g., the ability to hydrolyze polypeptide substrates, e.g., short polypeptide substrates or signaling molecules that function at the cell surface, e.g., enkalphins, cytokines, or molecules of the TGFθ family. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion, or premature truncation of the amino acid sequence of SEQ ID NO:2, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

Moreover, nucleic acid molecules encoding other 48149 family members and, thus, which have a nucleotide sequence which differs from the 48149 sequences of SEQ ID NO:1 or SEQ ID NO:3 are intended to be within the scope of the invention.

Antisense Nucleic Acid Molecules, Ribozymes and Modified 48149 Nucleic Acid Molecules

In another aspect, the invention features, an isolated nucleic acid molecule which is antisense to 48149. An "antisense" nucleic acid can include a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire 48149 coding strand, or to only a portion thereof (e.g., the coding region of human 48149 corresponding to SEQ ID NO:3). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding 48149 (e.g., the 5' and 3' untranslated regions).

An antisense nucleic acid can be designed such that it is complementary to the entire coding region of 48149 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of 48149 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 48149 mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length.

An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. The antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject (e.g., by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 48149 protein

to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

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

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. A ribozyme having specificity for a 48149-encoding nucleic acid can include one or more sequences complementary to the nucleotide sequence of a 48149 cDNA disclosed herein (i.e., SEQ ID NO:1 or SEQ ID NO:3), and a sequence having known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach (1988) *Nature* 334:585-591). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 48149-encoding mRNA. See, e.g., Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, 48149 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

48149 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 48149 (e.g., the 48149 promoter and/or enhancers) to form triple helical structures that prevent transcription of the 48149 gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6:569-84; Helene, C. i

(1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14:807-15. The potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent, radioactive, or colorimetric.

A 48149 nucleic acid molecule can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For non-limiting examples of synthetic oligonucleotides with modifications see Toulmé (2001) *Nature Biotech.* 19:17 and Faria *et al.* (2001) *Nature Biotech.* 19:40-44. Such phosphoramidite oligonucleotides can be effective antisense agents.

For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4: 5-23). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic, e.g., a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra* and Perry-O'Keefe *et al. Proc. Natl. Acad. Sci.* 93: 14670-675.

PNAs of 48149 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 48149 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup B. *et al.* (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci.*

USA 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (see, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a 48149 nucleic acid of the invention, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantitating the presence of the 48149 nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi *et al.*, U.S. Patent No. 5,854,033; Nazarenko *et al.*, U.S. Patent No. 5,866,336, and Livak *et al.*, U.S. Patent 5,876,930.

Isolated 48149 Polypeptides

In another aspect, the invention features, an isolated 48149 protein, or fragment, e.g., a biologically active portion, for use as immunogens or antigens to raise or test (or more generally to bind) anti-48149 antibodies. 48149 protein can be isolated from cells or tissue sources using standard protein purification techniques. 48149 protein or fragments thereof can be produced by recombinant DNA techniques or synthesized chemically.

Polypeptides of the invention include those which arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and post-translational events. The polypeptide can be expressed in systems, e.g., cultured cells, which result in substantially the same post-translational modifications present when expressed the polypeptide is expressed in a native cell, or in systems which result in the alteration or omission of post-translational modifications, e.g., glycosylation or cleavage, present when expressed in a native cell.

In a preferred embodiment, a 48149 polypeptide has one or more of the following characteristics:

(i) it has aminopeptidase activity, e.g., the ability to hydrolyze polypeptide substrates, e.g., short polypeptide substrates or signaling molecules that function at the cell surface, e.g., enkalphins, cytokines or molecules of the TGFβ family;

(ii) it has a molecular weight, e.g., a deduced molecular weight, preferably ignoring any contribution of post translational modifications, amino acid composition or other physical characteristic of a 48149 polypeptide, e.g., a polypeptide of SEQ ID NO:2;

(iii) it has an overall sequence similarity of at least 60%, more preferably at least 70%, 80%, 90%, 95%, 98%, or 99% with a polypeptide a of SEQ ID NO:2;

(iv) it can be found on macrophages, endothelial cells, smooth muscle cells, liver cells, epithelial cells, e.g., of the kidney and small instestine;

(v) it has a aminopeptidase domain which shares preferably about 70%, 80%, 90%, 95%, 98%, or 99% homology with amino acid residues about 76 to 480 of SEQ ID NO:2;

(vi) it has a neutral zinc protease motif (PS00142); and

(vii) it can functionally collaborate with the protein CD10/neutral endopeptidase.

In a preferred embodiment the 48149 protein, or fragment thereof, differs from the corresponding sequence in SEQ ID:2. In one embodiment it differs by at least one but by less than 15, 10 or 5 amino acid residues. In another it differs from the corresponding sequence in SEQ ID NO:2 by at least one residue but less than 20%, 15%, 10% or 5% of the residues in it differ from the corresponding sequence in SEQ ID NO:2. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) The differences are, preferably, differences or changes at a non-essential residue or a conservative substitution. In a preferred embodiment the differences are not in the aminopeptidase domain e.g., about amino acid residues 76 to 480 of SEQ ID NO:2. In another preferred embodiment one or more differences are in the aminopeptidase domain e.g., about amino acid residues 76 to 480 of SEQ ID NO:2.

Other embodiments include a protein that contain one or more changes in amino acid sequence, e.g., a change in an amino acid residue which is not essential for activity. Such 48149 proteins differ in amino acid sequence from SEQ ID NO:2, yet retain biological activity.

In one embodiment, the protein includes an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, or more homologous to SEQ ID NO:2.

A 48149 protein or fragment is provided which varies from the sequence of SEQ ID NO:2 in regions defined by amino acids about 34 to 75 and 481 to 967 by at least one but by less than 15, 10 or 5 amino acid residues in the protein or fragment but which does

not differ from SEQ ID NO:2 in regions defined by amino acids about 1 to 33 and 76 to 480. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) In some embodiments the difference is at a non-essential residue or is a conservative substitution, while in others the difference is at an essential residue or is a non-conservative substitution.

In one embodiment, a biologically active portion of a 48149 protein includes a membrane peptidase domain. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native 48149 protein.

In a preferred embodiment, the 48149 protein has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, the 48149 protein is substantially identical to SEQ ID NO:2. In yet another embodiment, the 48149 protein is substantially identical to SEQ ID NO:2 and retains the functional activity of the protein of SEQ ID NO:2, as described in detail in the subsections above.

In a preferred embodiment, a fragment differs by at least 1, 2, 3, 10, 20, or more amino acid residues from a protein sequence encoded by the sequence of Genbank accession number M22324. Differ can include differing in length or sequence identity. For example, a fragment can: include one or more amino acid residues from SEQ ID NO: 2 outside the region encoded by nucleotides 120 to 3021 of SEQ ID NO:1; not include all of the amino acid residues of a sequence in M22324, e.g., can be one or more amino acid residues shorter (at one or both ends) than a sequence in M22324; or can differ by one or more amino acid residues in the region of overlap.

48149 Chimeric or Fusion Proteins

In another aspect, the invention provides 48149 chimeric or fusion proteins. As used herein, a 48149 "chimeric protein" or "fusion protein" includes a 48149 polypeptide linked to a non-48149 polypeptide. A "non-48149 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the 48149 protein, e.g., a protein which is different from the 48149 protein and which is derived from the same or a different organism. The 48149 polypeptide of the fusion protein can correspond to all or a portion e.g., a fragment described herein of a 48149 amino acid sequence. In a preferred embodiment, a 48149 fusion protein includes

at least one (or two) biologically active portion of a 48149 protein. The non-48149 polypeptide can be fused to the N-terminus or C-terminus of the 48149 polypeptide.

The fusion protein can include a moiety which has a high affinity for a ligand. For example, the fusion protein can be a GST-48149 fusion protein in which the 48149 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 48149. Alternatively, the fusion protein can be a 48149 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of 48149 can be increased through use of a heterologous signal sequence.

Fusion proteins can include all or a part of a serum protein, e.g., an IgG constant region, or human serum albumin.

The 48149 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 48149 fusion proteins can be used to affect the bioavailability of a 48149 substrate. 48149 fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a 48149 protein; (ii) mis-regulation of the 48149 gene; and (iii) aberrant post-translational modification of a 48149 protein.

Moreover, the 48149-fusion proteins of the invention can be used as immunogens to produce anti-48149 antibodies in a subject, to purify 48149 ligands and in screening assays to identify molecules which inhibit the interaction of 48149 with a 48149 substrate.

Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A 48149-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 48149 protein.

Variants of 48149 Proteins

In another aspect, the invention also features a variant of a 48149 polypeptide, e.g., which functions as an agonist (mimetics) or as an antagonist. Variants of the 48149 proteins can be generated by mutagenesis, e.g., discrete point mutation, the insertion or deletion of sequences or the truncation of a 48149 protein. An agonist of the 48149 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 48149 protein. An antagonist of a 48149 protein can inhibit one or more of the activities of the naturally occurring form of the 48149 protein by, for example, competitively modulating a 48149-mediated activity of a 48149 protein. Thus,

specific biological effects can be elicited by treatment with a variant of limited function. Preferably, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 48149 protein.

Variants of a 48149 protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a 48149 protein for agonist or antagonist activity.

Libraries of fragments e.g., N terminal, C terminal, or internal fragments, of a 48149 protein coding sequence can be used to generate a variegated population of fragments for screening and subsequent selection of variants of a 48149 protein. Variants in which a cysteine residues is added or deleted or in which a residue which is glycosylated is added or deleted are particularly preferred.

Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property are known in the art. Such methods are adaptable for rapid screening of the gene libraries generated by combinatorial mutagenesis of 48149 proteins. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 48149 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6:327-331).

Cell based assays can be exploited to analyze a variegated 48149 library. For example, a library of expression vectors can be transfected into a cell line, e.g., a cell line, which ordinarily responds to 48149 in a substrate-dependent manner. The transfected cells are then contacted with 48149 and the effect of the expression of the mutant on signaling by the 48149 substrate can be detected, e.g., by measuring monocyte/macrophage lipid loading processes. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the 48149 substrate, and the individual clones further characterized.

In another aspect, the invention features a method of making a 48149 polypeptide, e.g., a peptide having a non-wild type activity, e.g., an antagonist, agonist, or super agonist of a naturally occurring 48149 polypeptide, e.g., a naturally occurring 48149 polypeptide. The method includes: altering the sequence of a 48149 polypeptide, e.g., altering the sequence , e.g., by substitution or deletion of one or more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

In another aspect, the invention features a method of making a fragment or analog of a 48149 polypeptide a biological activity of a naturally occurring 48149 polypeptide. The method includes: altering the sequence, e.g., by substitution or deletion of one or more residues, of a 48149 polypeptide, e.g., altering the sequence of a non-conserved region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

Anti-48149 Antibodies

In another aspect, the invention provides an anti-48149 antibody, or a fragment thereof (e.g., an antigen-binding fragment thereof). The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. As used herein, the term "antibody" refers to a protein comprising at least one, and preferably two, heavy (H) chain variable regions (abbreviated herein as VH), and at least one and preferably two light (L) chain variable regions (abbreviated herein as VL). The VH and VL regions can be further subdivided into regions of hypervariability, termed "complementarity determining regions" ("CDR"), interspersed with regions that are more conserved, termed "framework regions" (FR). The extent of the framework region and CDR's has been precisely defined (see, Kabat, E.A., *et al.* (1991) *Sequences of Proteins of Immunological Interest, Fifth Edition*, U.S. Department of Health and Human Services, NIH Publication No. 91-3242, and Chothia, C. et al. (1987) J. Mol. Biol. 196:901-917, which are incorporated herein by reference). Each VH and VL is composed of three CDR's and four FRs, arranged from amino-terminus to carboxy-terminus in the following order: FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4. The anti-48149 antibody can further include a heavy and light chain constant region, to thereby form a heavy and light immunoglobulin chain, respectively. In one embodiment, the antibody is a tetramer of two heavy immunoglobulin chains and two light immunoglobulin chains, wherein the heavy and light immunoglobulin chains are inter-connected by, e.g., disulfide bonds. The heavy chain constant region is comprised of three domains, CH1, CH2 and CH3. The light chain constant region is comprised of one domain, CL. The variable region of the heavy and light chains contains a binding domain that interacts with an antigen. The constant regions of the antibodies typically mediate the binding of the antibody to host tissues or factors, including various cells of the immune

system (e.g., effector cells) and the first component (Clq) of the classical complement system.

As used herein, the term "immunoglobulin" refers to a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes. The recognized human immunoglobulin genes include the kappa, lambda, alpha (IgA1 and IgA2), gamma (IgG1, IgG2, IgG3, IgG4), delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Full-length immunoglobulin "light chains" (about 25 KDa or 214 amino acids) are encoded by a variable region gene at the NH2-terminus (about 110 amino acids) and a kappa or lambda constant region gene at the COOH--terminus. Full-length immunoglobulin "heavy chains" (about 50 KDa or 446 amino acids), are similarly encoded by a variable region gene (about 116 amino acids) and one of the other aforementioned constant region genes, e.g., gamma (encoding about 330 amino acids).

The term "antigen-binding fragment" of an antibody (or simply "antibody portion," or "fragment"), as used herein, refers to one or more fragments of a full-length antibody that retain the ability to specifically bind to the antigen, e.g., 48149 polypeptide or fragment thereof. Examples of antigen-binding fragments of the anti-48149 antibody include, but are not limited to: (i) a Fab fragment, a monovalent fragment consisting of the VL, VH, CL and CH1 domains; (ii) a $F(ab')_2$ fragment, a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; (iii) a Fd fragment consisting of the VH and CH1 domains; (iv) a Fv fragment consisting of the VL and VH domains of a single arm of an antibody, (v) a dAb fragment (Ward *et al.*, (1989) *Nature* 341:544-546), which consists of a VH domain; and (vi) an isolated complementarity determining region (CDR). Furthermore, although the two domains of the Fv fragment, VL and VH, are coded for by separate genes, they can be joined, using recombinant methods, by a synthetic linker that enables them to be made as a single protein chain in which the VL and VH regions pair to form monovalent molecules (known as single chain Fv (scFv); see *e.g.*, Bird *et al.* (1988) *Science* 242:423-426; and Huston *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:5879-5883). Such single chain antibodies are also encompassed within the term "antigen-binding fragment" of an antibody. These antibody fragments are obtained using conventional techniques known to those with skill in the art, and the fragments are screened for utility in the same manner as are intact antibodies.

The anti-48149 antibody can be a polyclonal or a monoclonal antibody. In other embodiments, the antibody can be recombinantly produced, e.g., produced by phage display or by combinatorial methods.

Phage display and combinatorial methods for generating anti-48149 antibodies are known in the art (as described in, e.g., Ladner et al. U.S. Patent No. 5,223,409; Kang et al. International Publication No. WO 92/18619; Dower et al. International Publication No. WO 91/17271; Winter et al. International Publication WO 92/20791; Markland et al. International Publication No. WO 92/15679; Breitling et al. International Publication WO 93/01288; McCafferty et al. International Publication No. WO 92/01047; Garrard et al. International Publication No. WO 92/09690; Ladner et al. International Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum Antibod Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffths et al. (1993) *EMBO J* 12:725-734; Hawkins et al. (1992) *J Mol Biol* 226:889-896; Clackson et al. (1991) *Nature* 352:624-628; Gram et al. (1992) *PNAS* 89:3576-3580; Garrad et al. (1991) *Bio/Technology* 9:1373-1377; Hoogenboom et al. (1991) *Nuc Acid Res* 19:4133-4137; and Barbas et al. (1991) *PNAS* 88:7978-7982, the contents of all of which are incorporated by reference herein).

In one embodiment, the anti-48149 antibody is a fully human antibody (e.g., an antibody made in a mouse which has been genetically engineered to produce an antibody from a human immunoglobulin sequence), or a non-human antibody, e.g., a rodent (mouse or rat), goat, primate (e.g., monkey), camel antibody. Preferably, the non-human antibody is a rodent (mouse or rat antibody). Method of producing rodent antibodies are known in the art.

Human monoclonal antibodies can be generated using transgenic mice carrying the human immunoglobulin genes rather than the mouse system. Splenocytes from these transgenic mice immunized with the antigen of interest are used to produce hybridomas that secrete human mAbs with specific affinities for epitopes from a human protein (see, e.g., Wood et al. International Application WO 91/00906, Kucherlapati et al. PCT publication WO 91/10741; Lonberg et al. International Application WO 92/03918; Kay et al. International Application 92/03917; Lonberg, N. et al. 1994 *Nature* 368:856-859; Green, L.L. et al. 1994 *Nature Genet.* 7:13-21; Morrison, S.L. et al. 1994 *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Bruggeman et al. 1993 *Year Immunol* 7:33-40; Tuaillon et al. 1993 *PNAS* 90:3720-3724; Bruggeman et al. 1991 *Eur J Immunol* 21:1323-1326).

An anti-48149 antibody can be one in which the variable region, or a portion thereof, e.g., the CDR's, are generated in a non-human organism, e.g., a rat or mouse. Chimeric, CDR-grafted, and humanized antibodies are within the invention. Antibodies generated in a non-human organism, e.g., a rat or mouse, and then modified, e.g., in the variable framework or constant region, to decrease antigenicity in a human are within the invention.

Chimeric antibodies can be produced by recombinant DNA techniques known in the art. For example, a gene encoding the Fc constant region of a murine (or other species) monoclonal antibody molecule is digested with restriction enzymes to remove the region encoding the murine Fc, and the equivalent portion of a gene encoding a human Fc constant region is substituted (see Robinson et al., International Patent Publication PCT/US86/02269; Akira, et al., European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al., European Patent Application 173,494; Neuberger et al., International Application WO 86/01533; Cabilly et al. U.S. Patent No. 4,816,567; Cabilly et al., European Patent Application 125,023; Better et al. (1988 *Science* 240:1041-1043); Liu et al. (1987) *PNAS* 84:3439-3443; Liu et al., 1987, *J. Immunol.* 139:3521-3526; Sun et al. (1987) *PNAS* 84:214-218; Nishimura et al., 1987, *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al., 1988, *J. Natl Cancer Inst.* 80:1553-1559).

A humanized or CDR-grafted antibody will have at least one or two but generally all three recipient CDR's (of heavy and or light immuoglobulin chains) replaced with a donor CDR. The antibody may be replaced with at least a portion of a non-human CDR or only some of the CDR's may be replaced with non-human CDR's. It is only necessary to replace the number of CDR's required for binding of the humanized antibody to a 48149 or a fragment thereof. Preferably, the donor will be a rodent antibody, e.g., a rat or mouse antibody, and the recipient will be a human framework or a human consensus framework. Typically, the immunoglobulin providing the CDR's is called the "donor" and the immunoglobulin providing the framework is called the "acceptor." In one embodiment, the donor immunoglobulin is a non-human (e.g., rodent). The acceptor framework is a naturally-occurring (e.g., a human) framework or a consensus framework, or a sequence about 85% or higher, preferably 90%, 95%, 99% or higher identical thereto.

As used herein, the term "consensus sequence" refers to the sequence formed from the most frequently occurring amino acids (or nucleotides) in a family of related sequences (See e.g., Winnaker, From Genes to Clones (Verlagsgesellschaft, Weinheim, Germany 1987). In a

family of proteins, each position in the consensus sequence is occupied by the amino acid occurring most frequently at that position in the family. If two amino acids occur equally frequently, either can be included in the consensus sequence. A "consensus framework" refers to the framework region in the consensus immunoglobulin sequence.

An antibody can be humanized by methods known in the art. Humanized antibodies can be generated by replacing sequences of the Fv variable region which are not directly involved in antigen binding with equivalent sequences from human Fv variable regions. General methods for generating humanized antibodies are provided by Morrison, S. L., 1985, *Science* 229:1202-1207, by Oi et al., 1986, *BioTechniques* 4:214, and by Queen et al. US 5,585,089, US 5,693,761 and US 5,693,762, the contents of all of which are hereby incorporated by reference. Those methods include isolating, manipulating, and expressing the nucleic acid sequences that encode all or part of immunoglobulin Fv variable regions from at least one of a heavy or light chain. Sources of such nucleic acid are well known to those skilled in the art and, for example, may be obtained from a hybridoma producing an antibody against a 48149 polypeptide or fragment thereof. The recombinant DNA encoding the humanized antibody, or fragment thereof, can then be cloned into an appropriate expression vector.

Humanized or CDR-grafted antibodies can be produced by CDR-grafting or CDR substitution, wherein one, two, or all CDR's of an immunoglobulin chain can be replaced. See e.g., U.S. Patent 5,225,539; Jones et al. 1986 *Nature* 321:552-525; Verhoeyan et al. 1988 *Science* 239:1534; Beidler et al. 1988 *J. Immunol.* 141:4053-4060; Winter US 5,225,539, the contents of all of which are hereby expressly incorporated by reference. Winter describes a CDR-grafting method which may be used to prepare the humanized antibodies of the present invention (UK Patent Application GB 2188638A, filed on March 26, 1987; Winter US 5,225,539), the contents of which is expressly incorporated by reference.

Also within the scope of the invention are humanized antibodies in which specific amino acids have been substituted, deleted or added. Preferred humanized antibodies have amino acid substitutions in the framework region, such as to improve binding to the antigen. For example, a humanized antibody will have framework residues identical to the donor framework residue or to another amino acid other than the recipient framework residue. To generate such antibodies, a selected, small number of acceptor framework residues of the humanized immunoglobulin chain can be replaced by the corresponding donor amino acids. Preferred locations of the substitutions include amino acid residues

adjacent to the CDR, or which are capable of interacting with a CDR (see e.g., US 5,585,089). Criteria for selecting amino acids from the donor are described in US 5,585,089, e.g., columns 12-16 of US 5,585,089, the e.g., columns 12-16 of US 5,585,089, the contents of which are hereby incorporated by reference. Other techniques for humanizing antibodies are described in Padlan et al. EP 519596 A1, published on December 23, 1992.

In preferred embodiments an antibody can be made by immunizing with purified 48149 antigen, or a fragment thereof, e.g., a membrane peptidase fragment described herein, membrane associated antigen, tissue, e.g., crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions, e.g., membrane fractions.

A full-length 48149 protein or, antigenic peptide fragment of 48149 can be used as an immunogen or can be used to identify anti-48149 antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. The antigenic peptide of 48149 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of 48149. Preferably, the antigenic peptide includes at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Fragments of 48149 can be used, e.g., as immunogens or to characterize the specificity of an antibody. For example, fragments of 48149 which include residues about 171 to 190, about 250 to 272, or about 321 to 339 can be used to make antibodies against hydrophilic regions of the 48149 protein. Similarly, fragments of 48149 which include residues about 10 to 37, about 273 to 291, or about 232 to 259 can be used to make an antibody against a hydrophobic region of the 48149 protein; fragments of 48149 which include residues about 25 to 968, about 30 to 968, or about 35 to 968 can be used to make an antibody against an extracellular region of the 48149 protein; a fragment of 48149 which include residues about 40 to 700, about 55 to 600, or about 70 to 500 can be used to make an antibody against the aminopeptidase region of the 48149 protein.

Antibodies reactive with, or specific for, any of these regions, or other regions or domains described herein are provided.

Antibodies which bind only native 48149 protein, only denatured or otherwise non-native 48149 protein, or which bind both, are with in the invention. Antibodies with linear or conformational epitopes are within the invention. Conformational epitopes can

sometimes be identified by identifying antibodies which bind to native but not denatured 48149 protein.

Preferred epitopes encompassed by the antigenic peptide are regions of 48149 are located on the surface of the protein, e.g., hydrophilic regions, as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human 48149 protein sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the 48149 protein and are thus likely to constitute surface residues useful for targeting antibody production.

In a preferred embodiment the antibody can bind to the extracellular portion of the 48149 protein, e.g., it can bind to a whole cell which expresses the 48149 protein. In preferred embodiments antibodies can bind one or more of purified antigen, membrane associated antigen, tissue, e.g., tissue sections, whole cells, preferably living cells, lysed cells, cell fractions, e.g., membrane fractions.

The anti-48149 antibody can be a single chain antibody. A single-chain antibody (scFV) may be engineered (see, for example, Colcher, D. *et al.* (1999) *Ann N Y Acad Sci* 880:263-80; and Reiter, Y. (1996) *Clin Cancer Res* 2:245-52). The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target 48149 protein.

In a preferred embodiment the antibody has effector function and/or can fix complement. In other embodiments the antibody does not recruit effector cells; or fix complement.

In a preferred embodiment, the antibody has reduced or no ability to bind an Fc receptor. For example, it is a isotype or subtype, fragment or other mutant, which does not support binding to an Fc receptor, e.g., it has a mutagenized or deleted Fc receptor binding region.

In a preferred embodiment, an anti-48149 antibody alters (e.g., increases or decreases) the membrane peptidase activity, e.g., monocyte/macrophage lipid loading processes of a 48149 polypeptide.

The antibody can be coupled to a toxin, e.g., a polypeptide toxin, e,g, ricin or diphtheria toxin or active fragment hereof, or a radioactive nucleus, or imaging agent, e.g. a radioactive, enzymatic, or other, e.g., imaging agent, e.g., a NMR contrast agent. Labels which produce detectable radioactive emissions or fluorescence are preferred.

An anti-48149 antibody (e.g., monoclonal antibody) can be used to isolate 48149 by standard techniques, such as affinity chromatography or immunoprecipitation.

Moreover, an anti-48149 antibody can be used to detect 48149 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-48149 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labelling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

The invention also includes a nucleic acid which encodes an anti-48149 antibody, e.g., an anti-48149 antibody described herein. Also included are vectors which include the nucleic acid and cells transformed with the nucleic acid, particularly cells which are useful for producing an antibody, e.g., mammalian cells, e.g. CHO or lymphatic cells.

The invention also includes cell lines, e.g., hybridomas, which make an anti-48149 antibody, e.g., and antibody described herein, and method of using said cells to make a 48149 antibody.

Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells

In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses.

A vector can include a 48149 nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the recombinant expression vector includes one or

more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as described herein (e.g., 48149 proteins, mutant forms of 48149 proteins, fusion proteins, and the like).

The recombinant expression vectors of the invention can be designed for expression of 48149 proteins in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be used in 48149 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for 48149 proteins. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six weeks).

To maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

The 48149 expression vector can be a yeast expression vector, a vector for expression in insect cells, e.g., a baculovirus expression vector or a vector suitable for expression in mammalian cells.

When used in mammalian cells, the expression vector's control functions can be provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

In another embodiment, the promoter is an inducible promoter, e.g., a promoter regulated by a steroid hormone, by a polypeptide hormone (e.g., by means of a signal transduction pathway), or by a heterologous polypeptide (e.g., the tetracycline-inducible systems, "Tet-On" and "Tet-Off"; see, e.g., Clontech Inc., CA, Gossen and Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547, and Paillard (1989) *Human Gene Therapy* 9:983).

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989)

Proc. Natl. Acad. Sci. USA 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example, the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus.

Another aspect the invention provides a host cell which includes a nucleic acid molecule described herein, e.g., a 48149 nucleic acid molecule within a recombinant expression vector or a 48149 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a 48149 protein can be expressed in bacterial cells (such as *E. coli*), insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells (African green monkey kidney cells CV-1 origin SV40 cells; Gluzman (1981) *Cell*23:175-182)). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

A host cell of the invention can be used to produce (i.e., express) a 48149 protein. Accordingly, the invention further provides methods for producing a 48149 protein using

the host cells of the invention. In one embodiment, the method includes culturing the host cell of the invention (into which a recombinant expression vector encoding a 48149 protein has been introduced) in a suitable medium such that a 48149 protein is produced. In another embodiment, the method further includes isolating a 48149 protein from the medium or the host cell.

In another aspect, the invention features, a cell or purified preparation of cells which include a 48149 transgene, or which otherwise misexpress 48149. The cell preparation can consist of human or non-human cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a 48149 transgene, e.g., a heterologous form of a 48149, e.g., a gene derived from humans (in the case of a non-human cell). The 48149 transgene can be misexpressed, e.g., overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene that mis-expresses an endogenous 48149, e.g., a gene the expression of which is disrupted, e.g., a knockout. Such cells can serve as a model for studying disorders that are related to mutated or mis-expressed 48149 alleles or for use in drug screening.

In another aspect, the invention features, a human cell, e.g., a macrophage, an endothelial cell, a liver cell, a hematopoietic stem cell, or a hepatic stem cell, transformed with nucleic acid which encodes a subject 48149 polypeptide.

Also provided are cells, preferably human cells, e.g., hematopoetic (e.g., macrophage), endothelial, liver, or fibroblast cells, in which an endogenous 48149 is under the control of a regulatory sequence that does not normally control the expression of the endogenous 48149 gene. The expression characteristics of an endogenous gene within a cell, e.g., a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous 48149 gene. For example, an endogenous 48149 gene which is "transcriptionally silent," e.g., not normally expressed, or expressed only at very low levels, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, e.g., Chappel, US 5,272,071; WO 91/06667, published in May 16, 1991.

In a preferred embodiment, recombinant cells described herein can be used for replacement therapy in a subject. For example, a nucleic acid encoding a 48149 polypeptide operably linked to an inducible promoter (e.g., a steroid hormone receptor-regulated promoter) is introduced into a human or nonhuman, e.g., mammalian, e.g.,

porcine recombinant cell. The cell is cultivated and encapsulated in a biocompatible material, such as poly-lysine alginate, and subsequently implanted into the subject. See, e.g., Lanza (1996) *Nat. Biotechnol.* 14:1107; Joki *et al.* (2001) *Nat. Biotechnol.* 19:35; and U.S. Patent No. 5,876,742. Production of 48149 polypeptide can be regulated in the subject by administering an agent (e.g., a steroid hormone) to the subject. In another preferred embodiment, the implanted recombinant cells express and secrete an antibody specific for a 48149 polypeptide. The antibody can be any antibody or any antibody derivative described herein.

Transgenic Animals

The invention provides non-human transgenic animals. Such animals are useful for studying the function and/or activity of a 48149 protein and for identifying and/or evaluating modulators of 48149 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, e.g., a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, other transgenes, e.g., a knockout, reduce expression. Thus, a transgenic animal can be one in which an endogenous 48149 gene has been altered by, e.g., by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a transgene of the invention to direct expression of a 48149 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a 48149 transgene in its genome and/or expression of 48149 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals

carrying a transgene encoding a 48149 protein can further be bred to other transgenic animals carrying other transgenes.

48149 proteins or polypeptides can be expressed in transgenic animals or plants, e.g., a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

The invention also includes a population of cells from a transgenic animal, as discussed, e.g., below.

Uses

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic).

The isolated nucleic acid molecules of the invention can be used, for example, to express a 48149 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect a 48149 mRNA (e.g., in a biological sample) or a genetic alteration in a 48149 gene, and to modulate 48149 activity, as described further below. The 48149 proteins can be used to treat disorders characterized by insufficient or excessive production of a 48149 substrate or production of 48149 inhibitors. In addition, the 48149 proteins can be used to screen for naturally occurring 48149 substrates, to screen for drugs or compounds which modulate 48149 activity, as well as to treat disorders characterized by insufficient or excessive production of 48149 protein or production of 48149 protein forms which have decreased, aberrant or unwanted activity compared to 48149 wild type protein (e.g., aberrant monocyte/macrophage lipid loading processes, atheroma, or aberrant lipid metabolism in the liver). Moreover, the anti-48149 antibodies of the invention can be used to detect and isolate 48149 proteins, regulate the bioavailability of 48149 proteins, and modulate 48149 activity.

A method of evaluating a compound for the ability to interact with, e.g., bind, a subject 48149 polypeptide is provided. The method includes: contacting the compound with the subject 48149 polypeptide; and evaluating ability of the compound to interact with, e.g., to bind or form a complex with the subject 48149 polypeptide. This method can

be performed in vitro, e.g., in a cell free system, or in vivo, e.g., in a two-hybrid interaction trap assay. This method can be used to identify naturally occurring molecules that interact with subject 48149 polypeptide. It can also be used to find natural or synthetic inhibitors of subject 48149 polypeptide. Screening methods are discussed in more detail below.

Screening Assays

The invention provides methods (also referred to herein as "screening assays") for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to 48149 proteins, have a stimulatory or inhibitory effect on, for example, 48149 expression or 48149 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 48149 substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., 48149 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions.

In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a 48149 protein or polypeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds that bind to or modulate an activity of a 48149 protein or polypeptide or a biologically active portion thereof.

In one embodiment, an activity of a 48149 protein can be assayed as follows. First, a recombinant fragment of 48149 protein that lacks the transmembrane domain but contains a purification tag, e.g., a His-6 tag, can be produced and purified, e.g., by metal ion chelation chromatography. Such a fragment, following purification, can be incubated in vitro with a substrate molecule, e.g., an enkaphalin, cytokine, of TGFθ family molecule, and proteolysis of the substrate molecule can observed, e.g., by immunoprecipitating the substrate molecule and detemining whether any of its' physical properties have been altered. For example, the mobility of the proteolyzed substrate molecule may be altered on a polyacrylamide gel. Alternatively, the assay can be a cell based assay where 48149 protein is expressed in the cell, e.g., a macrophage cell, and the activity of the 48149 protein is assayed by determining how the cell responds to a signaling molecule, e.g., an enkaphalin, cytokine, or TGFθ molecule, that is a 48149 substrate.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, e.g., Zuckermann, R.N. *et al.* (1994) *J. Med. Chem.* 37:2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner U.S. Patent No. 5,223,409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382; Felici (1991) *J. Mol. Biol.* 222:301-310; Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a 48149 protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate 48149 activity is determined. Determining the ability of the test compound to modulate 48149 activity can be accomplished by monitoring, for example, monocyte/macrophage lipid loading processing. The cell, for example, can be of mammalian origin, e.g., human.

The ability of the test compound to modulate 48149 binding to a compound, e.g., a 48149 substrate, or to bind to 48149 can also be evaluated. This can be accomplished, for example, by coupling the compound, e.g., the substrate, with a radioisotope or enzymatic label such that binding of the compound, e.g., the substrate, to 48149 can be determined by

detecting the labeled compound, e.g., substrate, in a complex. Alternatively, 48149 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate 48149 binding to a 48149 substrate in a complex. For example, compounds (e.g., 48149 substrates) can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

The ability of a compound (e.g., a 48149 substrate) to interact with 48149 with or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with 48149 without the labeling of either the compound or the 48149. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and 48149.

In yet another embodiment, a cell-free assay is provided in which a 48149 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the 48149 protein or biologically active portion thereof is evaluated. Preferred biologically active portions of the 48149 proteins to be used in assays of the present invention include fragments which participate in interactions with non-48149 molecules, e.g., fragments with high surface probability scores.

Soluble and/or membrane-bound forms of isolated proteins (e.g., 48149 proteins or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET) (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos, *et al.*, U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

In another embodiment, determining the ability of the 48149 protein to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

It may be desirable to immobilize either 48149, an anti-48149 antibody or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test

compound to a 48149 protein, or interaction of a 48149 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/48149 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 48149 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 48149 binding or activity determined using standard techniques.

Other techniques for immobilizing either a 48149 protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated 48149 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

In one embodiment, this assay is performed utilizing antibodies reactive with 48149 protein or target molecules but which do not interfere with binding of the 48149

protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or 48149 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 48149 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 48149 protein or target molecule.

Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas, G., and Minton, A.P., (1993) *Trends Biochem Sci* 18:284-7); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel, F. *et al.*, eds. Current Protocols in Molecular Biology 1999, J. Wiley: New York.); and immunoprecipitation (see, for example, Ausubel, F. *et al.*, eds. (1999) *Current Protocols in Molecular Biology*, J. Wiley: New York). Such resins and chromatographic techniques are known to one skilled in the art (see, e.g., Heegaard, N.H., (1998) *J Mol Recognit* 11:141-8; Hage, D.S., and Tweed, S.A. (1997) *J Chromatogr B Biomed Sci Appl.* 699:499-525). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

In a preferred embodiment, the assay includes contacting the 48149 protein or biologically active portion thereof with a known compound which binds 48149 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 48149 protein, wherein determining the ability of the test compound to interact with a 48149 protein includes determining the ability of the test compound to preferentially bind to 48149 or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

The target gene products of the invention can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the 48149 genes herein identified. In an

alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of a 48149 protein through modulation of the activity of a downstream effector of a 48149 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

In yet another aspect, the 48149 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with 48149 ("48149-binding proteins" or "48149-bp") and are involved in 48149 activity. Such 48149-bps can be activators or inhibitors of signals by the 48149 proteins or 48149 targets as, for example, downstream elements of a 48149-mediated signaling pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 48149 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively the: 48149 protein can be the fused to the activator domain.) If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a 48149-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., lacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the 48149 protein.

In another embodiment, modulators of 48149 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of 48149 mRNA or protein evaluated relative to the level of expression of 48149 mRNA or protein in the absence of the candidate compound. When expression of 48149 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 48149 mRNA or protein expression. Alternatively, when expression of 48149 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 48149 mRNA or protein expression. The level of 48149 mRNA or protein expression can be determined by methods described herein for detecting 48149 mRNA or protein.

In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a 48149 protein can be confirmed *in vivo*, e.g., in an animal such as an animal model for atherosclerosis.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., a 48149 modulating agent, an antisense 48149 nucleic acid molecule, a 48149-specific antibody, or a 48149-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be used for treatments as described herein.

Detection Assays

Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome e.g., to locate gene regions associated with genetic disease or to associate 48149 with a disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

Chromosome Mapping

The 48149 nucleotide sequences or portions thereof can be used to map the location of the 48149 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the 48149 sequences with genes associated with disease.

Briefly, 48149 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 48149 nucleotide sequences. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 48149 sequences will yield an amplified fragment.

A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse

chromosomes, can allow easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et al.* (1983) *Science* 220:919-924).

Other mapping strategies e.g., in situ hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries can be used to map 48149 to a chromosomal location.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, Human Chromosomes: A Manual of Basic Techniques ((1988) Pergamon Press, New York).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

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

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 48149 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence.

Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

48149 sequences can be used to identify individuals from biological samples using, e.g., restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, e.g., in a Southern blot, and probed to yield bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 48149 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 can provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from 48149 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

Use of Partial 48149 Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 (e.g., fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 bases, preferably at least 30 bases) are particularly appropriate for this use.

The 48149 nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 48149 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., 48149 primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

Generally, the invention provides, a method of determining if a subject is at risk for a disorder related to a lesion in or the misexpression of a gene which encodes 48149.

Such disorders include, e.g., a disorder associated with the misexpression of 48149 gene, a cardiovascular disorder (e.g., atherosclerosis), a metabolic disorder (e.g., a metabolic disorder of the liver), or a non-neutrophil-mediated inflammatory disorder.

The method includes one or more of the following:

detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the 48149 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, e.g., a mutation in the 5' control region;

detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the 48149 gene;

detecting, in a tissue of the subject, the misexpression of the 48149 gene, at the mRNA level, e.g., detecting a non-wild type level of a mRNA ;

detecting, in a tissue of the subject, the misexpression of the gene, at the protein level, e.g., detecting a non-wild type level of a 48149 polypeptide.

In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the 48149 gene; an insertion of one or more nucleotides into the gene, a point mutation, e.g., a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, e.g., a translocation, inversion, or deletion.

For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence from SEQ ID NO:1, or naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the 48149 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting, by hybridization, e.g., *in situ* hybridization, of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion.

In preferred embodiments detecting the misexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the 48149 gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of 48149.

Methods of the invention can be used prenatally or to determine if a subject's offspring will be at risk for a disorder.

In preferred embodiments the method includes determining the structure of a 48149 gene, an abnormal structure being indicative of risk for the disorder.

In preferred embodiments the method includes contacting a sample from the subject with an antibody to the 48149 protein or a nucleic acid, which hybridizes specifically with the gene. These and other embodiments are discussed below.

Diagnostic and Prognostic Assays

Diagnostic and prognostic assays of the invention include method for assessing the expression level of 48149 molecules and for identifying variations and mutations in the sequence of 48149 molecules.

Expression Monitoring and Profiling. The presence, level, or absence of 48149 protein or nucleic acid in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 48149 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes 48149 protein such that the presence of 48149 protein or nucleic acid is detected in the biological sample. The term "biological sample" includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression of the 48149 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the 48149 genes; measuring the amount of protein encoded by the 48149 genes; or measuring the activity of the protein encoded by the 48149 genes.

The level of mRNA corresponding to the 48149 gene in a cell can be determined both by *in situ* and by *in vitro* formats.

The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length 48149 nucleic acid, such as the nucleic acid of SEQ ID NO:1, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 48149 mRNA or genomic DNA. The probe can be disposed on an address of an array, e.g., an array described below. Other suitable probes for use in the diagnostic assays are described herein.

In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the probes, for example, in a two-dimensional gene chip array described below. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the 48149 genes.

The level of mRNA in a sample that is encoded by one of 48149 can be evaluated with nucleic acid amplification, e.g., by rtPCR (Mullis (1987) U.S. Patent No. 4,683,202), ligase chain reaction (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193), self sustained sequence replication (Guatelli *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.*, (1989), *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, (1988) *Bio/Technology* 6:1197), rolling circle replication (Lizardi *et al.*, U.S. Patent No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

For *in situ* methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the 48149 gene being analyzed.

In another embodiment, the methods further contacting a control sample with a compound or agent capable of detecting 48149 mRNA, or genomic DNA, and comparing the presence of 48149 mRNA or genomic DNA in the control sample with the presence of 48149 mRNA or genomic DNA in the test sample. In still another embodiment, serial analysis of gene expression, as described in U.S. Patent No. 5,695,937, is used to detect 48149 transcript levels.

A variety of methods can be used to determine the level of protein encoded by 48149. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample.

In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.

The detection methods can be used to detect 48149 protein in a biological sample *in vitro* as well as *in vivo*. *In vitro* techniques for detection of 48149 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western blot analysis. *In vivo* techniques for detection of 48149 protein include introducing into a subject a labeled anti-48149 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. In another embodiment, the sample is labeled, e.g., biotinylated and then contacted to the antibody, e.g., an anti-48149 antibody positioned on an antibody array (as described below). The sample can be detected, e.g., with avidin coupled to a fluorescent label.

In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting 48149 protein, and comparing the presence of 48149 protein in the control sample with the presence of 48149 protein in the test sample.

The invention also includes kits for detecting the presence of 48149 in a biological sample. For example, the kit can include a compound or agent capable of detecting 48149 protein or mRNA in a biological sample; and a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 48149 protein or nucleic acid.

For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a

polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also includes a buffering agent, a preservative, or a protein stabilizing agent. The kit can also includes components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed or aberrant or unwanted 48149 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as atherosclerosis, aberrant monocyte/macrophage lipid loading processing or deregulated cell proliferation.

In one embodiment, a disease or disorder associated with aberrant or unwanted 48149 expression or activity is identified. A test sample is obtained from a subject and 48149 protein or nucleic acid (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of 48149 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 48149 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest, including a biological fluid (e.g., serum), cell sample, or tissue.

The prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 48149 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder associated with the misexpression of 48149 gene, a cardiovascular disorder (e.g., atherosclerosis), a metabolic disorder (e.g., a metabolic disorder of the liver), or a non-neutrophil-mediated inflammatory disorder.

In another aspect, the invention features a computer medium having a plurality of digitally encoded data records. Each data record includes a value representing the level of expression of 48149 in a sample, and a descriptor of the sample. The descriptor of the sample can be an identifier of the sample, a subject from which the sample was derived

(e.g., a patient), a diagnosis, or a treatment (e.g., a preferred treatment). In a preferred embodiment, the data record further includes values representing the level of expression of genes other than 48149 (e.g., other genes associated with a 48149-disorder, or other genes on an array). The data record can be structured as a table, e.g., a table that is part of a database such as a relational database (e.g., a SQL database of the Oracle or Sybase database environments).

Also featured is a method of evaluating a sample. The method includes providing a sample, e.g., from the subject, and determining a gene expression profile of the sample, wherein the profile includes a value representing the level of 48149 expression. The method can further include comparing the value or the profile (i.e., multiple values) to a reference value or reference profile. The gene expression profile of the sample can be obtained by any of the methods described herein (e.g., by providing a nucleic acid from the sample and contacting the nucleic acid to an array). The method can be used to diagnose atherosclerosis in a subject wherein an increase in 48149 expression is an indication that the subject has or is disposed to having atherosclerosis. The method can be used to monitor a treatment for atherosclerosis in a subject. For example, the gene expression profile can be determined for a sample from a subject undergoing treatment. The profile can be compared to a reference profile or to a profile obtained from the subject prior to treatment or prior to onset of the disorder (see, e.g., Golub *et al.* (1999) *Science* 286:531).

In yet another aspect, the invention features a method of evaluating a test compound (see also, "Screening Assays", above). The method includes providing a cell and a test compound; contacting the test compound to the cell; obtaining a subject expression profile for the contacted cell; and comparing the subject expression profile to one or more reference profiles. The profiles include a value representing the level of 48149 expression. In a preferred embodiment, the subject expression profile is compared to a target profile, e.g., a profile for a normal cell or for desired condition of a cell. The test compound is evaluated favorably if the subject expression profile is more similar to the target profile than an expression profile obtained from an uncontacted cell.

In another aspect, the invention features, a method of evaluating a subject. The method includes: a) obtaining a sample from a subject, e.g., from a caregiver, e.g., a caregiver who obtains the sample from the subject; b) determining a subject expression profile for the sample. Optionally, the method further includes either or both of steps: c) comparing the subject expression profile to one or more reference expression profiles; and d) selecting the reference profile most similar to the subject reference profile. The subject

expression profile and the reference profiles include a value representing the level of 48149 expression. A variety of routine statistical measures can be used to compare two reference profiles. One possible metric is the length of the distance vector that is the difference between the two profiles. Each of the subject and reference profile is represented as a multi-dimensional vector, wherein each dimension is a value in the profile.

The method can further include transmitting a result to a caregiver. The result can be the subject expression profile, a result of a comparison of the subject expression profile with another profile, a most similar reference profile, or a descriptor of any of the aforementioned. The result can be transmitted across a computer network, e.g., the result can be in the form of a computer transmission, e.g., a computer data signal embedded in a carrier wave.

Also featured is a computer medium having executable code for effecting the following steps: receive a subject expression profile; access a database of reference expression profiles; and either i) select a matching reference profile most similar to the subject expression profile or ii) determine at least one comparison score for the similarity of the subject expression profile to at least one reference profile. The subject expression profile, and the reference expression profiles each include a value representing the level of 48149 expression.

Arrays and Uses Thereof

In another aspect, the invention features an array that includes a substrate having a plurality of addresses. At least one address of the plurality includes a capture probe that binds specifically to a 48149 molecule (e.g., a 48149 nucleic acid or a 48149 polypeptide). The array can have a density of at least than 10, 50, 100, 200, 500, 1,000, 2,000, or 10,000 or more addresses/cm^2, and ranges between. In a preferred embodiment, the plurality of addresses includes at least 10, 100, 500, 1,000, 5,000, 10,000, 50,000 addresses. In a preferred embodiment, the plurality of addresses includes equal to or less than 10, 100, 500, 1,000, 5,000, 10,000, or 50,000 addresses. The substrate can be a two-dimensional substrate such as a glass slide, a wafer (e.g., silica or plastic), a mass spectroscopy plate, or a three-dimensional substrate such as a gel pad. Addresses in addition to address of the plurality can be disposed on the array.

In a preferred embodiment, at least one address of the plurality includes a nucleic acid capture probe that hybridizes specifically to a 48149 nucleic acid, e.g., the sense or anti-sense strand. In one preferred embodiment, a subset of addresses of the plurality of addresses has a nucleic acid capture probe for 48149. Each address of the subset can include a capture probe that hybridizes to a different region of a 48149 nucleic acid. In another preferred embodiment, addresses of the subset include a capture probe for a 48149 nucleic acid. Each address of the subset is unique, overlapping, and complementary to a different variant of 48149 (e.g., an allelic variant, or all possible hypothetical variants). The array can be used to sequence 48149 by hybridization (see, e.g., U.S. Patent No. 5,695,940).

An array can be generated by various methods, e.g., by photolithographic methods (see, e.g., U.S. Patent Nos. 5,143,854; 5,510,270; and 5,527,681), mechanical methods (e.g., directed-flow methods as described in U.S. Patent No. 5,384,261), pin-based methods (e.g., as described in U.S. Pat. No. 5,288,514), and bead-based techniques (e.g., as described in PCT US/93/04145).

In another preferred embodiment, at least one address of the plurality includes a polypeptide capture probe that binds specifically to a 48149 polypeptide or fragment thereof. The polypeptide can be a naturally-occurring interaction partner of 48149 polypeptide. Preferably, the polypeptide is an antibody, e.g., an antibody described herein (see "Anti-48149 Antibodies," above), such as a monoclonal antibody or a single-chain antibody.

In another aspect, the invention features a method of analyzing the expression of 48149. The method includes providing an array as described above; contacting the array with a sample and detecting binding of a 48149-molecule (e.g., nucleic acid or polypeptide) to the array. In a preferred embodiment, the array is a nucleic acid array. Optionally the method further includes amplifying nucleic acid from the sample prior or during contact with the array.

In another embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array, particularly the expression of 48149. If a sufficient number of diverse samples is analyzed, clustering (e.g., hierarchical clustering, k-means clustering, Bayesian clustering and the like) can be used to identify other genes which are co-regulated with 48149. For example, the array can be used for the quantitation of the expression of multiple genes. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertained. Quantitative data

can be used to group (e.g., cluster) genes on the basis of their tissue expression *per se* and level of expression in that tissue.

For example, array analysis of gene expression can be used to assess the effect of cell-cell interactions on 48149 expression. A first tissue can be perturbed and nucleic acid from a second tissue that interacts with the first tissue can be analyzed. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined, e.g., to monitor the effect of cell-cell interaction at the level of gene expression.

In another embodiment, cells are contacted with a therapeutic agent. The expression profile of the cells is determined using the array, and the expression profile is compared to the profile of like cells not contacted with the agent. For example, the assay can be used to determine or analyze the molecular basis of an undesirable effect of the therapeutic agent. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

In another embodiment, the array can be used to monitor expression of one or more genes in the array with respect to time. For example, samples obtained from different time points can be probed with the array. Such analysis can identify and/or characterize the development of a 48149-associated disease or disorder; and processes, such as a cellular transformation associated with a 48149-associated disease or disorder. The method can also evaluate the treatment and/or progression of a 48149-associated disease or disorder

The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (*e.g.*, including 48149) that could serve as a molecular target for diagnosis or therapeutic intervention.

In another aspect, the invention features an array having a plurality of addresses. Each address of the plurality includes a unique polypeptide. At least one address of the plurality has disposed thereon a 48149 polypeptide or fragment thereof. Methods of producing polypeptide arrays are described in the art, e.g., in De Wildt *et al.* (2000). *Nature Biotech. 18*, 989-994; Lueking *et al.* (1999). *Anal. Biochem. 270*, 103-111; Ge, H. (2000). *Nucleic Acids Res. 28*, e3, I-VII; MacBeath, G., and Schreiber, S.L. (2000).

Science 289, 1760-1763; and WO 99/51773A1. In a preferred embodiment, each addresses of the plurality has disposed thereon a polypeptide at least 60, 70, 80,85, 90, 95 or 99 % identical to a 48149 polypeptide or fragment thereof. For example, multiple variants of a 48149 polypeptide (e.g., encoded by allelic variants, site-directed mutants, random mutants, or combinatorial mutants) can be disposed at individual addresses of the plurality. Addresses in addition to the address of the plurality can be disposed on the array.

The polypeptide array can be used to detect a 48149 binding compound, e.g., an antibody in a sample from a subject with specificity for a 48149 polypeptide or the presence of a 48149-binding protein or ligand.

The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (*e.g.*, ascertaining the effect of 48149 expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

In another aspect, the invention features a method of analyzing a plurality of probes. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which express 48149 or from a cell or subject in which a 48149 mediated response has been elicited, e.g., by contact of the cell with 48149 nucleic acid or protein, or administration to the cell or subject 48149 nucleic acid or protein; providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which does not express 48149 (or does not express as highly as in the case of the 48149 positive plurality of capture probes) or from a cell or subject which in which a 48149 mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); contacting the array with one or more inquiry probes (which is preferably other than a 48149 nucleic acid, polypeptide, or antibody), and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody.

In another aspect, the invention features a method of analyzing a plurality of probes or a sample. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality having a unique capture probe, contacting the array with a first sample from a cell or subject which express or mis-express 48149 or from a cell or subject in which a 48149-mediated response has been elicited, e.g., by contact of the cell with 48149 nucleic acid or protein, or administration to the cell or subject 48149 nucleic acid or protein; providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, and contacting the array with a second sample from a cell or subject which does not express 48149 (or does not express as highly as in the case of the 48149 positive plurality of capture probes) or from a cell or subject which in which a 48149 mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); and comparing the binding of the first sample with the binding of the second sample. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody. The same array can be used for both samples or different arrays can be used. If different arrays are used the plurality of addresses with capture probes should be present on both arrays.

In another aspect, the invention features a method of analyzing 48149, e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 48149 nucleic acid or amino acid sequence; comparing the 48149 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 48149.

Detection of Variations or Mutations

The methods of the invention can also be used to detect genetic alterations in a 48149 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 48149 protein activity or nucleic acid expression, such as a cardiovascular disorder (e.g., atherosclerosis), a metabolic disorder (e.g., a metabolic disorder of the liver), or a non-neutrophil-mediated inflammatory disorder. In preferred embodiments, the methods include detecting, in a sample from the subject, the presence or

absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a 48149-protein, or the mis-expression of the 48149 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 48149 gene; 2) an addition of one or more nucleotides to a 48149 gene; 3) a substitution of one or more nucleotides of a 48149 gene, 4) a chromosomal rearrangement of a 48149 gene; 5) an alteration in the level of a messenger RNA transcript of a 48149 gene, 6) aberrant modification of a 48149 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a 48149 gene, 8) a non-wild type level of a 48149-protein, 9) allelic loss of a 48149 gene, and 10) inappropriate post-translational modification of a 48149-protein.

An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the 48149-gene. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both) from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 48149 gene under conditions such that hybridization and amplification of the 48149-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein. Alternatively, other amplification methods described herein or known in the art can be used.

In another embodiment, mutations in a 48149 gene from a sample cell can be identified by detecting alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined, e.g., by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in 48149 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, two-dimensional arrays, e.g., chip

based arrays. Such arrays include a plurality of addresses, each of which is positionally distinguishable from the other. A different probe is located at each address of the plurality. A probe can be complementary to a region of a 48149 nucleic acid or a putative variant (e.g., allelic variant) thereof. A probe can have one or more mismatches to a region of a 48149 nucleic acid (e.g., a destabilizing mismatch). The arrays can have a high density of addresses, e.g., can contain hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in 48149 can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 48149 gene and detect mutations by comparing the sequence of the sample 48149 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry.

Other methods for detecting mutations in the 48149 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242; Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295).

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 48149 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662; U.S. Patent No. 5,459,039).

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 48149 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control 48149 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). A further method of detecting point mutations is the chemical ligation of oligonucleotides as described in Xu *et al.* ((2001) *Nature Biotechnol.* 19:148). Adjacent oligonucleotides, one of which selectively anneals to the query site, are ligated together if the nucleotide at the query site of the sample nucleic acid is complementary to the query oligonucleotide; ligation can be monitored, e.g., by fluorescent dyes coupled to the oligonucleotides.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of

interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

In another aspect, the invention features a set of oligonucleotides. The set includes a plurality of oligonucleotides, each of which is at least partially complementary (e.g., at least 50%, 60%, 70%, 80%, 90%, 92%, 95%, 97%, 98%, or 99% complementary) to a 48149 nucleic acid.

In a preferred embodiment the set includes a first and a second oligonucleotide. The first and second oligonucleotide can hybridize to the same or to different locations of SEQ ID NO:1 or the complement of SEQ ID NO:1. Different locations can be different but overlapping, or non-overlapping on the same strand. The first and second oligonucleotide can hybridize to sites on the same or on different strands.

The set can be useful, e.g., for identifying SNP's, or identifying specific alleles of 48149. In a preferred embodiment, each oligonucleotide of the set has a different nucleotide at an interrogation position. In one embodiment, the set includes two oligonucleotides, each complementary to a different allele at a locus, e.g., a biallelic or polymorphic locus.

In another embodiment, the set includes four oligonucleotides, each having a different nucleotide (e.g., adenine, guanine, cytosine, or thymidine) at the interrogation position. The interrogation position can be a SNP or the site of a mutation. In another preferred embodiment, the oligonucleotides of the plurality are identical in sequence to one another (except for differences in length). The oligonucleotides can be provided with differential labels, such that an oligonucleotide that hybridizes to one allele provides a signal that is distinguishable from an oligonucleotide that hybridizes to a second allele. In still another embodiment, at least one of the oligonucleotides of the set has a nucleotide change at a position in addition to a query position, e.g., a destabilizing mutation to

decrease the T_m of the oligonucleotide. In another embodiment, at least one oligonucleotide of the set has a non-natural nucleotide, e.g., inosine. In a preferred embodiment, the oligonucleotides are attached to a solid support, e.g., to different addresses of an array or to different beads or nanoparticles.

In a preferred embodiment the set of oligo nucleotides can be used to specifically amplify, e.g., by PCR, or detect, a 48149 nucleic acid.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.,* in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a 48149 gene.

Use of 48149 Molecules as Surrogate Markers

The 48149 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 48149 molecules of the invention may be detected, and may be correlated with one or more biological states in vivo. For example, the 48149 molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (e.g., with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (e.g., early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (e.g., an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art

include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

The 48149 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug in vivo. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker (e.g., a 48149 marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-48149 antibodies may be employed in an immune-based detection system for a 48149 protein marker, or 48149-specific radiolabeled probes may be used to detect a 48149 mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am, J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

The 48149 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g., McLeod *et al.* (1999) *Eur. J. Cancer* 35:1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the

presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA, or protein (e.g., 48149 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 48149 DNA may correlate 48149 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

Pharmaceutical Compositions

The nucleic acid and polypeptides, fragments thereof, as well as anti-48149 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions. Such compositions typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

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

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

Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of

the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

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

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

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

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

It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated;

each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit high therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective

amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

For antibodies, the preferred dosage is 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank *et al.* ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e.,. including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific

compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

An antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, puromycin, maytansinoids, *e.g.*, maytansinol (see US Patent No. 5,208,020), CC-1065 (see US Patent Nos. 5,475,092, 5,585,499, 5,846,545) and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, CC-1065, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine, vinblastine, taxol and maytansinoids). Radioactive ions include, but are not limited to iodine, yttrium and praseodymium.

The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, α-interferon, β-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophase colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

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

Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 48149 expression or activity. As used herein, the term "treatment" is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 48149 molecules of the present invention or 48149 modulators according to that individual's drug response genotype. Pharmacogenomics

allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted 48149 expression or activity, by administering to the subject a 48149 or an agent which modulates 48149 expression or at least one 48149 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 48149 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 48149 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 48149 aberrance, for example, a 48149, 48149 agonist or 48149 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

It is possible that some 48149 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

The 48149 molecules can act as novel diagnostic targets and therapeutic agents for controlling cardiovascular disorders or disorders associated with accumulation in the liver of fibrous tissue. Additionally, 48149 molecules may play an important role in the etiology of certain viral diseases, including but not limited to Hepatitis B, Hepatitis C and Herpes Simplex Virus (HSV). Modulators of 48149 activity could be used to control viral diseases. The modulators can be used in the treatment and/or diagnosis of viral infected tissue or virus-associated tissue fibrosis, especially liver and liver fibrosis. Also, 48149 modulators can be used in the treatment and/or diagnosis of virus-associated carcinoma, especially hepatocellular cancer.

As described in the appended examples, expression of 48149 mRNA was detected in a number of tumor samples, including breast, ovary, lung, colon, and cervical tumors, as well as in the normal tissues from which the tumors were derived. Expression of 48149 tended to be reduced in breast and colon tumors, as compared to normal breast and colon tissue, respectively. In contrast, expression of 48149 mRNA was elevated in all ovary tumors tested, as compared to normal ovary tissue. Accordingly, 48149 molecules can act as novel diagnostic targets and therapeutic agents for controlling disorders involving

aberrant activity of any of the aforesaid in which the 48149 gene is expressed, e.g., cancerous or non-malignant disorders of the breast, ovary, lung, colon, and cervix.

As discussed, successful treatment of 48149 disorders can be brought about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds, e.g., an agent identified using an assays described above, that proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of 48149 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, $F(ab')_2$ and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof).

Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular protein, it can be preferable to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

Another method by which nucleic acid molecules may be utilized in treating or preventing a disease characterized by 48149 expression is through the use of aptamer molecules specific for 48149 protein. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to protein ligands (see, e.g., Osborne, *et al.* (1997) *Curr. Opin. Chem Biol.* 1: 5-9; and Patel, D.J. (1997) *Curr Opin Chem Biol* 1:32-46). Since nucleic acid molecules may in many cases be more conveniently introduced into target cells than therapeutic protein molecules may be,

aptamers offer a method by which 48149 protein activity may be specifically decreased without the introduction of drugs or other molecules which may have pluripotent effects.

Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies may, therefore, by administered in instances whereby negative modulatory techniques are appropriate for the treatment of 48149 disorders. For a description of antibodies, see the Antibody section above.

In circumstances wherein injection of an animal or a human subject with a 48149 protein or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against 48149 through the use of anti-idiotypic antibodies (see, for example, Herlyn, D. (1999) *Ann Med* 31:66-78; and Bhattacharya-Chatterjee, M., and Foon, K.A. (1998) *Cancer Treat Res.* 94:51-68). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the 48149 protein. Vaccines directed to a disease characterized by 48149 expression may also be generated in this fashion.

In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see e.g., Marasco *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:7889-7893).

The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or ameliorate 48149 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures as described above.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or

no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography. Another example of determination of effective dose for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays may utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. The compound which is able to modulate 48149 activity is used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell, R. J. *et al* (1996) *Current Opinion in Biotechnology* **7**:89-94 and in Shea, K.J. (1994) *Trends in Polymer Science* 2:166-173. Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, G. *et al* (1993) *Nature* 361:645-647. Through the use of isotope-labeling, the "free" concentration of compound which modulates the expression or activity of 48149 can be readily monitored and used in calculations of IC_{50}.

Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50}. An rudimentary example of such a "biosensor" is discussed in Kriz, D. *et al* (1995) *Analytical Chemistry* 67:2142-2144.

Another aspect of the invention pertains to methods of modulating 48149 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a

48149 or agent that modulates one or more of the activities of 48149 protein activity associated with the cell. An agent that modulates 48149 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 48149 protein (e.g., a 48149 substrate or receptor), a 48149 antibody, a 48149 agonist or antagonist, a peptidomimetic of a 48149 agonist or antagonist, or other small molecule.

In one embodiment, the agent stimulates one or 48149 activities. Examples of such stimulatory agents include active 48149 protein and a nucleic acid molecule encoding 48149. In another embodiment, the agent inhibits one or more 48149 activities. Examples of such inhibitory agents include antisense 48149 nucleic acid molecules, anti-48149 antibodies, and 48149 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 48149 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up regulates or down regulates) 48149 expression or activity. In another embodiment, the method involves administering a 48149 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 48149 expression or activity.

Stimulation of 48149 activity is desirable in situations in which 48149 is abnormally downregulated and/or in which increased 48149 activity is likely to have a beneficial effect. For example, stimulation of 48149 activity is desirable in situations in which a 48149 is downregulated and/or in which increased 48149 activity is likely to have a beneficial effect. Likewise, inhibition of 48149 activity is desirable in situations in which 48149 is abnormally upregulated and/or in which decreased 48149 activity is likely to have a beneficial effect.

Pharmacogenomics

The 48149 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 48149 activity (e.g., 48149 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 48149 associated disorders (e.g., cardiovascular disorders (e.g., atherosclerosis), metabolic disorder (e.g., a metabolic

disorder of the liver), or non-neutrophil-mediated inflammatory disorders) associated with aberrant or unwanted 48149 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 48149 molecule or 48149 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 48149 molecule or 48149 modulator.

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

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease

process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach," can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (e.g., a 48149 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

Alternatively, a method termed the "gene expression profiling," can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 48149 molecule or 48149 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 48149 molecule or 48149 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the 48149 genes of the present invention, wherein these products may be associated with resistance of the cells to a therapeutic agent. Specifically, the activity of the proteins encoded by the 48149 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, e.g., human cells, will become sensitive to treatment with an agent that the unmodified target cells were resistant to.

Monitoring the influence of agents (e.g., drugs) on the expression or activity of a 48149 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 48149 gene expression,

protein levels, or upregulate 48149 activity, can be monitored in clinical trials of subjects exhibiting decreased 48149 gene expression, protein levels, or downregulated 48149 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 48149 gene expression, protein levels, or downregulate 48149 activity, can be monitored in clinical trials of subjects exhibiting increased 48149 gene expression, protein levels, or upregulated 48149 activity. In such clinical trials, the expression or activity of a 48149 gene, and preferably, other genes that have been implicated in, for example, a 48149-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

48149 Informatics

The sequence of a 48149 molecule is provided in a variety of media to facilitate use thereof. A sequence can be provided as a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a 48149. Such a manufacture can provide a nucleotide or amino acid sequence, e.g., an open reading frame, in a form which allows examination of the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form. The sequence information can include, but is not limited to, 48149 full-length nucleotide and/or amino acid sequences, partial nucleotide and/or amino acid sequences, polymorphic sequences including single nucleotide polymorphisms (SNPs), epitope sequence, and the like. In a preferred embodiment, the manufacture is a machine-readable medium, e.g., a magnetic, optical, chemical or mechanical information storage device.

As used herein, "machine-readable media" refers to any medium that can be read and accessed directly by a machine, e.g., a digital computer or analogue computer. Non-limiting examples of a computer include a desktop PC, laptop, mainframe, server (e.g., a web server, network server, or server farm), handheld digital assistant, pager, mobile telephone, and the like. The computer can be stand-alone or connected to a communications network, e.g., a local area network (such as a VPN or intranet), a wide area network (e.g., an Extranet or the Internet), or a telephone network (e.g., a wireless, DSL, or ISDN network). Machine-readable media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM,

ROM, EPROM, EEPROM, flash memory, and the like; and hybrids of these categories such as magnetic/optical storage media.

A variety of data storage structures are available to a skilled artisan for creating a machine-readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of data processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

In a preferred embodiment, the sequence information is stored in a relational database (such as Sybase or Oracle). The database can have a first table for storing sequence (nucleic acid and/or amino acid sequence) information. The sequence information can be stored in one field (e.g., a first column) of a table row and an identifier for the sequence can be store in another field (e.g., a second column) of the table row. The database can have a second table, e.g., storing annotations. The second table can have a field for the sequence identifier, a field for a descriptor or annotation text (e.g., the descriptor can refer to a functionality of the sequence, a field for the initial position in the sequence to which the annotation refers, and a field for the ultimate position in the sequence to which the annotation refers. Non-limiting examples for annotation to nucleic acid sequences include polymorphisms (e.g., SNP's) translational regulatory sites and splice junctions. Non-limiting examples for annotations to amino acid sequence include polypeptide domains, e.g., a domain described herein; active sites and other functional amino acids; and modification sites.

By providing the nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. A search is used to identify fragments or regions of the sequences of the invention which

match a particular target sequence or target motif. The search can be a BLAST search or other routine sequence comparison, e.g., a search described herein.

Thus, in one aspect, the invention features a method of analyzing 48149, e.g., analyzing structure, function, or relatedness to one or more other nucleic acid or amino acid sequences. The method includes: providing a 48149 nucleic acid or amino acid sequence; comparing the 48149 sequence with a second sequence, e.g., one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database to thereby analyze 48149. The method can be performed in a machine, e.g., a computer, or manually by a skilled artisan.

The method can include evaluating the sequence identity between a 48149 sequence and a database sequence. The method can be performed by accessing the database at a second site, e.g., over the Internet.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. Typical sequence lengths of a target sequence are from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI).

Thus, the invention features a method of making a computer readable record of a sequence of a 48149 sequence which includes recording the sequence on a computer readable matrix. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

In another aspect, the invention features, a method of analyzing a sequence. The method includes: providing a 48149 sequence, or record, in machine-readable form;

comparing a second sequence to the 48149 sequence; thereby analyzing a sequence. Comparison can include comparing to sequences for sequence identity or determining if one sequence is included within the other, e.g., determining if the 48149 sequence includes a sequence being compared. In a preferred embodiment the 48149 or second sequence is stored on a first computer, e.g., at a first site and the comparison is performed, read, or recorded on a second computer, e.g., at a second site. E.g., the 48149 or second sequence can be stored in a public or proprietary database in one computer, and the results of the comparison performed, read, or recorded on a second computer. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

In another aspect, the invention provides a machine-readable medium for holding instructions for performing a method for determining whether a subject has a 48149-associated disease or disorder or a pre-disposition to a 48149-associated disease or disorder, wherein the method comprises the steps of determining 48149 sequence information associated with the subject and based on the 48149 sequence information, determining whether the subject has a 48149-associated disease or disorder or a pre-disposition to a 48149-associated disease or disorder and/or recommending a particular treatment for the disease, disorder or pre-disease condition.

The invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a 48149-associated disease or disorder or a pre-disposition to a disease associated with a 48149 wherein the method comprises the steps of determining 48149 sequence information associated with the subject, and based on the 48149 sequence information, determining whether the subject has a 48149-associated disease or disorder or a pre-disposition to a 48149-associated disease or disorder, and/or recommending a particular treatment for the disease, disorder or pre-disease condition. In a preferred embodiment, the method further includes the step of receiving information, e.g., phenotypic or genotypic information, associated with the subject and/or acquiring from a network phenotypic information associated with the subject. The information can be stored in a database, e.g., a relational database. In another embodiment, the method further includes accessing the database, e.g., for records relating to other subjects, comparing the 48149 sequence of the subject to the 48149 sequences in the database to

thereby determine whether the subject as a 48149-associated disease or disorder, or a pre-disposition for such.

The present invention also provides in a network, a method for determining whether a subject has a 48149 associated disease or disorder or a pre-disposition to a 48149-associated disease or disorder associated with 48149, said method comprising the steps of receiving 48149 sequence information from the subject and/or information related thereto, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to 48149 and/or corresponding to a 48149-associated disease or disorder (e.g., a cardiovascular disorder (e.g., atherosclerosis), a metabolic disorder (e.g., a metabolic disorder of the liver), or a non-neutrophil-mediated inflammatory disorder), and based on one or more of the phenotypic information, the 48149 information (*e.g.*, sequence information and/or information related thereto), and the acquired information, determining whether the subject has a 48149-associated disease or disorder or a pre-disposition to a 48149-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

The present invention also provides a method for determining whether a subject has a 48149 -associated disease or disorder or a pre-disposition to a 48149-associated disease or disorder, said method comprising the steps of receiving information related to 48149 (*e.g.*, sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to 48149 and/or related to a 48149-associated disease or disorder, and based on one or more of the phenotypic information, the 48149 information, and the acquired information, determining whether the subject has a 48149-associated disease or disorder or a pre-disposition to a 48149-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

This invention is further illustrated by the following examples that should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

EXAMPLES

Example 1: Identification and Characterization of Human 48149 cDNA

The human 48149 nucleic acid sequence is recited as follows:

TAATTTTTGCCCAGTCTGCCTGTTGTGGGGCTCCTCCCCTTTGGGGATATAAGCCCGGCC
TGGGGCTGCTCCGTTCTCTGCCTGGCCTGAGGCTCCCTGAGCCGCCTCCCCACCATCACC
ATGGCCAAGGGCTTCTATATTTCCAAGTCCCTGGGCATCCTGGGGATCCTCCTGGGCGTG
GCAGCCGTGTGCACAATCATCGCACTGTCAGTGGTGTACTCCCAGGAGAAGAACAAGAAC
GCCAACAGCTCCCCCGTGGCCTCCACCACCCCGTCCGCCTCAGCCACCACCAACCCCGCC
TCGGCCACCACCTTGGACCAAAGTAAAGCGTGGAATCGTTACCGCCTCCCCAACACGCTG
AAACCCGATTCCTACCAGGTGACGCTGAGACCGTACCTCACCCCCAATGACAGGGGCCTG
TACGTTTTTAAGGGCTCCAGCACCGTCCGTTTCACCTGCAAGGAGGCCACTGACGTCATC
ATCATCCACAGCAAGAAGCTCAACTACACCCTCAGCCAGGGGCACAGGGTGGTCCTGCGT
GGTGTGGGAGGCTCCCAGCCCCCCGACATTGACAAGACTGAGCTGGTGGAGCCCACCGAG
TACCTGGTGGTGCACCTCAAGGGCTCCCTGGTGAAGGACAGCCAGTATGAGATGGACAGC
GAGTTCGAGGGGGAGTTGGCAGATGACCTGGCGGGCTTCTACCGCAGCGAGTACATGGAG
GGCAATGTCAGAAAGGTGGTGGCCACTACACAGATGCAGGCTGCAGATGCCCGGAAGTCC
TTCCCATGCTTCGATGAGCCGGCCATGAAGGCCGAGTTCAACATCACGCTTATCCACCCC
AAGGACCTGACAGCCCTGTCCAACATGCTTCCCAAAGGTCCCAGCACCCCACTTCCAGAA
GACCCCAACTGGAATGTCACTGAGTTCCACACCACGCCCAAGATGTCCACGTACTTGCTG
GCCTTCATTGTCAGTGAGTTCGACTACGTGGAGAAGCAGGCATCCAATGGTGTCTTGATC
CGGATCTGGGCCCGGCCCAGTGCCATTGCGGCGGGCCACGGCGATTATGCCCTGAACGTG
ACGGGCCCCATCCTTAACTTCTTTGCTGGTCATTATGACACACCCTACCCACTCCCAAAA
TCAGACCAGATTGGCCTGCCAGACTTCAACGCCGGCGCCATGGAGAACTGGGGACTGGTG
ACCTACCGGGAGAACTCCCTGCTGTTCGACCCCCTGTCCTCCTCCAGCAGCAACAAGGAG
CGGGTGGTCACTGTGATTGCTCATGAGCTGGCCCACCAGTGGTTCGGGAACCTGGTGACC
ATAGAGTGGTGGAATGACCTGTGGCTGAACGAGGGCTTCGCCTCCTACGTGGAGTACCTG
GGTGCTGACTATGCGGAGCCCACCTGGAACTTGAAAGACCTCATGGTGCTGAATGATGTG
TACCGCGTGATGGCAGTGGATGCACTGGCCTCCTCCCACCCGCTGTCCACACCCGCCTCG
GAGATCAACACGCCGGCCCAGATCAGTGAGCTGTTTGACGCCATCTCCTACAGCAAGGGC
GCCTCAGTCCTCAGGATGCTCTCCAGCTTCCTGTCCGAGGACGTATTCAAGCAGGGCCTG
GCGTCCTACCTCCACACCTTTGCCTACCAGAACACCATCTACCTGAACCTGTGGGACCAC
CTGCAGGAGGCTGTGAACAACCGGTCCATCCAACTCCCCACCACCGTGCGGGACATCATG
AACCGCTGGACCCTGCAGATGGGCTTCCCGGTCATCACGGTGGATACCAGCACGGGGACC
CTTTCCCAGGAGCACTTCCTCCTTGACCCCGATTCCAATGTTACCCGCCCCTCAGAATTC
AACTACGTGTGGATTGTGCCCATCACATCCATCAGAGATGGCAGACAGCAGCAGGACTAC
TGGCTGATAGATGTAAGAGCCCAGAACGATCTCTTCAGCACATCAGGCAATGAGTGGGTC

CTGCTGAACCTCAATGTGACGGGCTATTACCGGGTGAACTACGACGAAGAGAACTGGAGG
AAGATTCAGACTCAGCTGCAGAGAGACCACTCGGCCATCCCTGTCATCAATCGGGCACAG
ATCATTAATGACGCCTTCAACCTGGCCAGTGCCCATAAGGTCCCTGTCACTCTGGCGCTG
AACAACACCCTCTTCCTGATTGAAGAGAGACAGTACATGCCCTGGGAGGCCGCCCTGAGC
AGCCTGAGCTACTTCAAGCTCATGTTTGACCGCTCCGAGGTCTATGGCCCCATGAAGAAC
TACCTGAAGAAGCAGGTCACACCCCTCTTCATTCACTTCAGAAATAATACCAACAACTGG
AGGGAGATCCCAGAAAACCTGATGGACCAGTACAGCGAGGTTAATGCCATCAGCACCGCC
TGCTCCAACGGAGTTCCAGAGTGTGAGGAGATGGTCTCTGGCCTTTTCAAGCAGTGGATG
GAGAACCCCAATAATAACCCGATCCACCCCAACCTGCGGTCCACCGTCTACTGCAACGCT
ATCGCCCAGGGCGGGGAGGAGGAGTGGGACTTCGCCTGGGAGCAGTTCCGAAATGCCACA
CTGGTCAATGAGGCTGACAAGCTCCGGGCAGCCCTGGCCTGCAGCAAAGAGTTGTGGATC
CTGAACAGGTACCTGAGCTACACCCTGAACCCGGACTTAATCCGGAAGCAGGACGCCACC
TCTACCATCATCAGCATTACCAACAACGTCATTGGGCAAGGTCTGGTCTGGGACTTTGTC
CAGAGCAACTGGAAGAAGCTTTTTAACGATTATGGTGGTGGCTCGTTCTCCTTCTCCAAC
CTCATCCAGGCAGTGACACGACGATTCTCCACCGAGTATGAGCTGCAGCAGCTGGAGCAG
TTCAAGAAGGACAACGAGGAAACAGGCTTCGGCTCAGGCACCCGGGCCCTGGAGCAAGCC
CTGGAGAAGACGAAAGCCAACATCAAGTGGGTGAAGGAGAACAAGGAGGTGGTGCTCCAG
TGGTTCACAGAAAACAGCAAA**TAG**TCCCCAGCCCTTGAAGTCACCCGGCCCCGATGCAAG
GTGCCCACATGTGTCCATCCCAGCGGCTGGTGCAGGGCCTCCATTCCTGGAGCCCGAGGC
ACCAGTGTCCTCCCCTCAAGGACAAAGTCTCCAGCCCACGTTCTCTCTGCCTGTGAGCCA
GTCTAGTTCCTGATGACCCAGGCTGCCTGAGCACCTCCCAGCCCCTGCCCCTCATGCCAA
CCCCGCCCTAGGCCTGGCATGGCACCTGTCGCCCAGTGCCCTGGGGCTGATCTCAGGGAA
GCCCAGCTCCAGGGCCAGATGAGCAGAAGCTCTCGATGGACAATGAACGGCCTTGCTGGG
GGCCGCCCTGTACCCTCTTTCACCTTTCCCTAAAGACCCTAAATCTGAGGAATCAACAGG
GCAGCAGATCTGTATATTTTTTTCTAAGAGAAAATGTAAATAAAGGATTTCTAGATGAAA
AAAAAAAAAAAAAA (SEQ ID NO:1).

The human 48149 sequence (SEQ ID NO:1), which is approximately 3494 nucleotides long. The nucleic acid sequence includes an initiation codon (ATG) and a termination codon (TAG) which are underscored above. The region between and inclusive of the initiation codon and the termination codon is a methionine-initiated coding sequence of about 2904 nucleotides, including the termination codon (nucleotides indicated as "coding" of SEQ ID NO:1; SEQ ID NO:3). The coding sequence encodes a 967 amino acid protein (SEQ ID NO:2), which is recited as follows:

MAKGFYISKSLGILGILLGVAAVCTIIALSVVYSQEKNKNANSSPVASTTPSASATTNPA
SATTLDQSKAWNRYRLPNTLKPDSYQVTLRPYLTPNDRGLYVFKGSSTVRFTCKEATDVI
IIHSKKLNYTLSQGHRVVLRGVGGSQPPDIDKTELVEPTEYLVVHLKGSLVKDSQYEMDS
EFEGELADDLAGFYRSEYMEGNVRKVVATTQMQAADARKSFPCFDEPAMKAEFNITLIHP

KDLTALSNMLPKGPSTPLPEDPNWNVTEFHTTPKMSTYLLAFIVSEFDYVEKQASNGVLI
RIWARPSAIAAGHGDYALNVTGPILNFFAGHYDTPYPLPKSDQIGLPDFNAGAMENWGLV
TYRENSLLFDPLSSSSSNKERVVTVIAHELAHQWFGNLVTIEWWNDLWLNEGFASYVEYL
GADYAEPTWNLKDLMVLNDVYRVMAVDALASSHPLSTPASEINTPAQISELFDAISYSKG
ASVLRMLSSFLSEDVFKQGLASYLHTFAYQNTIYLNLWDHLQEAVNNRSIQLPTTVRDIM
NRWTLQMGFPVITVDTSTGTLSQEHFLLDPDSNVTRPSEFNYVWIVPITSIRDGRQQQDY
WLIDVRAQNDLFSTSGNEWVLLNLNVTGYYRVNYDEENWRKIQTQLQRDHSAIPVINRAQ
IINDAFNLASAHKVPVTLALNNTLFLIEERQYMPWEAALSSLSYFKLMFDRSEVYGPMKN
YLKKQVTPLFIHFRNNTNNWREIPENLMDQYSEVNAISTACSNGVPECEEMVSGLFKQWM
ENPNNNPIHPNLRSTVYCNAIAQGGEEEWDFAWEQFRNATLVNEADKLRAALACSKELWI
LNRYLSYTLNPDLIRKQDATSTIISITNNVIGQGLVWDFVQSNWKKLFNDYGGGSFSFSN
LIQAVTRRFSTEYELQQLEQFKKDNEETGFGSGTRALEQALEKTKANIKWVKENKEVVLQ
WFTENSK (SEQ ID NO:2).

Example 2: Tissue Distribution of 48149 mRNA by TaqMan Analysis

Endogenous human 48149 gene expression was determined using the Perkin-Elmer/ABI 7700 Sequence Detection System which employs TaqMan technology. Briefly, TaqMan technology relies on standard RT-PCR with the addition of a third gene-specific oligonucleotide (referred to as a probe) which has a fluorescent dye coupled to its 5' end (typically 6-FAM) and a quenching dye at the 3' end (typically TAMRA). When the fluorescently tagged oligonucleotide is intact, the fluorescent signal from the 5' dye is quenched. As PCR proceeds, the 5' to 3' nucleolytic activity of Taq polymerase digests the labeled primer, producing a free nucleotide labeled with 6-FAM, which is now detected as a fluorescent signal. The PCR cycle where fluorescence is first released and detected is directly proportional to the starting amount of the gene of interest in the test sample, thus providing a quantitative measure of the initial template concentration. Samples can be internally controlled by the addition of a second set of primers/probe specific for a housekeeping gene such as GAPDH which has been labeled with a different fluorophore on the 5' end (typically VIC).

To determine the level of 48149 in various human tissues a primer/probe set was designed. Total RNA was prepared from a series of human tissues using an RNeasy kit from Qiagen. First strand cDNA was prepared from 1 μg total RNA using an oligo-dT primer and Superscript II reverse transcriptase (Gibco/BRL). cDNA obtained from approximately 50 ng total RNA was used per TaqMan reaction. Tissues tested include the human tissues and several cell lines shown in Tables 1-5. 48149 mRNA was detected in in

a wide range of tissues, with notable expression in colon, small intestine, coronary smooth muscle cells, pancreas, liver, blood cells, kidney and heart (Tables 1-2), as well as in a variety of tumors and some lung tumor cell lines (Tables 3-4). Importantly, 48149 expression was found to be elevated in monkeys that were fed a diet containing saturated fatty acids, as compared to monkeys that were fed a diet that contained poly-unsaturated fats (Table 5).

Table 1: Normalized Expression of 48149 mRNA

Tissue	Relative Expression
Artery normal	0.1427
Aorta diseased	0.0641
Vein normal	0.2433
Coronary SMC	9.5519
HUVEC	1.791
Hemangioma	0.509
Heart normal	0.1929
Heart CHF	0.0369
Kidney	0.5908
Skeletal Muscle	9.8545
Adipose normal	0.7504
Pancreas	6.3899
primary osteoblasts	1.4905
Osteoclasts differentiated	0.5325
Skin normal	1.4548
Spinal cord normal	0.043
Brain Cortex normal	0.007
Brain Hypothalamus normal	0.0094
Nerve	0.0995
DRG (Dorsal Root Ganglion)	0.0675
Breast normal	1.543
Breast tumor	0.1534
Ovary normal	0.2133

Ovary Tumor	0.2441
Prostate Normal	2.9707
Prostate Tumor	1.603
Salivary glands	0.0786
Colon normal	138.2163
Colon Tumor	2.022
Lung normal	1.57
Lung tumor	0.1354
Lung COPD	1.4957
Colon IBD	3.9745
Liver normal	5.5627
Liver fibrosis	1.3526
Spleen normal	0.22
Tonsil normal	0.2662
Lymph node normal	0.0837
Small intestine normal	19.0377
Macrophages	5.8595
Synovium	0.7224
BM-MNC	6.0034
Activated PBMC	0.7504
Neutrophils	2.2986
Megakaryocytes	2.9604
Erythroid	0.0168
positive control	0.1208

Expression of human 48149 mRNA was detected in all tissues tested, with the exception of brain cortex tissue. The highest levels of 48149 expression were detected in normal colon tissue. Other tissues that displayed a relatively high level of 48149 expression include the small intestine, coronary smooth muscle cells, pancreas, liver, and blood cells (e.g., macrophages and bone marrow mononuclear cells). Most tissue obtained from tumors, e.g., breast, prostate, colon, and lung, displayed a reduction in 48149 expression as compared to the corresponding normal tissue. Abbreviations used in Table 1 include: SMC, smooth muscle cells; HUVEC, human umbilical vein endothelial cells;

CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; IBD, inflammatory bowel disease; BM-MNC, bone marrow mononuclear cells; and PBMC, peripheral blood mononuclear cells.

Table 2: Expression of 48149 mRNA in Highly Vascularized Tissues

Tissue	Relative Expression
ONC 101 Hemangioma	17.28
ONC 102 Hemangioma	3.61
ONC 103 Hemangioma	16.92
CHT 1273 Glioblastoma	3.08
CHT 216 Glioblastoma	2.15
CHT 501 Glioblastoma	21.27
NDR 203 Normal Kidney	81.05
PIT 213 Renal Cell Carcinoma	17.89
CHT 732 Wilms Tumor	2.78
CHT 765 Wilms Tumor	16.18
NDR 295 Skin	6.30
CHT 1424 Uterine Adenocarcinoma	3.13
CHT 1238 Neuroblastoma	2.79
BWH 78 Fetal Adrenal	1.71
BWH 74 Fetal Kidney	65.38
BWH 4 Fetal Heart	5.41
MPI 849 Normal Heart	30.93
CLN 746 Spinal cord	4.10
TCH002 Hemangioma	0.00
TCH003 Hemangioma	6.80
TCH004 Hemangioma	2.21

Expression of 48149 mRNA was detected in a number of highly vascularized tissues, including hemagiomas, glioblastomas, kidney (fetal and normal adult), Wilms Tumors, skin, a uterine carcinoma, a neuroblastoma, fetal adrenal gland, heart (fetal and normal adult), and spinal chord. Expression of 48149 was highest in the adult and fetal

kidney. High levels of 48149 expression were also detected in normal heart tissue and some abnormal growths, e.g., hemangiomas, a glioblastoma, a renal cell carcinoma, and a Wilms tumor. The level of 48149 mRNA expression detected in the renal cell carcinoma was much lower than the level detected in fetal or normal adult kidney tissue.

5

Table 3: Expression of 48149 mRNA in Oncology Samples

Tissue	Relative Expression
PIT 400 Breast N	67.45
PIT 372 Breast N	67.45
CHT 558 Breast N	3.51
CLN 168 Breast T: IDC	0.74
MDA 304 Breast T: MD-IDC	5.84
NDR 58 Breast T: IDC	8.12
NDR 05 Breast T: IDC	24.01
CHT 562 Breast T: IDC	1.82
NDR 138 Breast T ILC (LG)	15.15
CHT 1841 Lymph node (Breast met)	10.31
PIT 58 Lung (Breast met)	14.83
PIT 208 Ovary N	0.00
CHT 620 Ovary N	2.10
CLN 03 Ovary T	5.72
CLN 17 Ovary T	639.49
MDA 25 Ovary T	15.57
MDA 216 Ovary T	49.04
CLN 012 Ovary T	44.04
MDA 185 Lung N	9.59
CLN 930 Lung N	16.01
MDA 183 Lung N	24.69
MPI 215 Lung T--SmC	1.85
MDA 259 Lung T-PDNSCCL	21.79
CHT 832 Lung T-PDNSCCL	13.89
MDA 262 Lung T-SCC	14.28

CHT 793 Lung T-ACA	0.34
CHT 331 Lung T-ACA	5.96
CHT 405 Colon N	246.56
CHT 523 Colon N	380.24
CHT 371 Colon N	843.82
CHT 382 Colon T: MD	4.96
CHT 528 Colon T: MD	510.51
CLN 609 Colon T	3.68
CHT 372 Colon T: MD-PD	96.72
CHT 340 Colon-Liver Met	115.82
NDR 100 Colon-Liver Met	46.39
PIT 260 Liver N (female)	69.35
CHT 1653 Cervix Squamous CC	1.40
CHT 569 Cervix Squamous CC	0.10
A24 HMVEC-Arr	70.56
C48 HMVEC-Prol	173.74

Expression of 48149 mRNA was detected in a number of tumor samples, including breast, ovary, lung, colon, and cervical tumors, as well as in the normal tissues from which the tumors were derived. Expression of 48149 tended to be reduced in breast and colon tumors, as compared to normal breast and colon tissue, respectively. In contrast, expression of 48149 mRNA was elevated in all ovary tumors tested, as compared to normal ovary tissue. Finally, 48149 mRNA expression was detected in human vascular endothelial cells (HMVEC), where it was upregulated in proliferating HMVECs as compared to arrested HMVECs. Abbreviations used in Table 3 include: N, normal; T, tumor; IDC, Invasive Ductal Carcinoma; ILC, Invasive Lobular Carcinoma; MD, moderately differentiated; SmC, small cell carcinoma; PDNSCC, poorly differentiated non-small cell carcinoma; SCC, squamous cell carcinoma; andACA, adenocarcinoma.

Table 4: Expression of 48149 mRNA in Lung Cell Lines

Tissue	Relative Expression
NHBE	24.77

A549 (BA)	32.02
H460 (LCLC)	0.00
H23 (AC)	5.66
H522 (AC)	0.00
H125 (AC/SCC)	435.28
H520 (SCC)	0.00
H69 (SCLC)	0.00
H345 (SCLC)	0.00
H460 INCX 24hr	0.00
H460 p16 24hr	0.00
H460 INCX 48hr	0.00
H460 p16 48hr	0.00
H460 INCX Stable Plas	0.00
H460 p16 Stable Plas	0.00
H460 NA-Agar	0.00
H460 Incx stable Agar	0.00
H460 p16 stable Agar	0.00
H125 Incx 96hr	646.18
H125 p53 96hr	257.03
H345 Mock 144hr	0.00
H345 Gluc 144hr	0.00
H345 VIP 144hr	0.00

Elevated expression of 48149 mRNA was detected in several lung cell lines, including NHBE cells, the lung carcinoma cell line A549, and H125 adenosquamous carcinoma cells. Expression of 48149 mRNA was reduced in H125 cells that were transfected with a p53 expression construct, as compared to H125 cells transfected with an empty vector.

Table 5: Expression of 48149 mRNA in Wake Forest Samples

Animal/Diet	Relative Expression
#800 CHOW DIET	2.216

#813 CHOW DIET	4.749
#842 CHOW DIET	22.592
#909 CHOW DIET	6.354
#816 CHOW DIET	5.456
#824 CHOW DIET	0.268
#830 CHOW DIET	0.623
#905 CHOW DIET	1.492
#826 CHOW DIET	1.212
#844 CHOW DIET	0.489
#846 CHOW DIET	0.875
#908 CHOW DIET	0.095
#815 CHOW DIET	0.112
#798 CHOW DIET	0.194
#932 CHOW DIET	0.645
#800 POLY DIET W/OUT CHOL	0.998
#813 POLY DIET W/OUT CHOL	1.055
#842 POLY DIET W/OUT CHOL	0.046
#909 POLY DIET W/OUT CHOL	0.082
#816 POLY DIET W/OUT CHOL	0.160
#824 POLY DIET W/OUT CHOL	0.543
#830 POLY DIET W/OUT CHOL	0.985
#905 POLY DIET W/OUT CHOL	0.581
#826 MONO DIET W/OUT CHOL	0.032
#844 MONO DIET W/OUT CHOL	0.044
#846 MONO DIET W/OUT CHOL	0.147
#908 MONO DIET W/OU CHOL	0.299
#815 MONO DIET W/OUT CHOL	0.558
#798 MONO DIET W/OUT CHOL	0.269
#932 MONO DIET W/OUT CHOL	0.028
#926 MONO DIET W/OUT CHOL	0.113
#800 POLY DIET W/ CHOL	0.482
#813 POLY DIET W/ CHOL	12.7355
#842 POLY DIET W/ CHOL	1.714

#909 POLY DIET W/ CHOL	1.070
#816 POLY DIET W/CHOL	0.082
#824 POLY DIET W/CHOL	0.170
#830 POLY DIET W/CHOL	0.585
#905 POLY DIET W/CHOL	1.246
#826 MONO DIET W/ CHOL	1.556
#844 MONO DIET W/CHOL	1.345
#846 MONO DIET W/CHOL	0.054
#908 MONO DIET W/ CHOL	0.230
#815 MONO DIET W/ CHOL	0.938
#798 MONO DIET W/CHOL	2.262
#932 MONO DIET W/CHOL	4.401
#926 MONO DIET W/CHOL	7.987

Table 5 depicts the expression of 48149 mRNA in the liver of Wake Forest monkeys that were given a particular diet. The data demonstrate reduced 48149 expression in six out of eight monkeys (see #800, #813, #842, #909, #816, #824, #830, and #905) when they were fed a polyunsaturated diet, as compared to when they were fed a "chow" diet, i.e., a diet containing normal dietary levels of fatty acids, including saturated fatty acids. A diet that emphasizes polyunsaturated fats is expected to be cardioprotective and limit or prevent the formation of atherosclerotic plaques in arteries. Thus, genes that are down-regulated in response to such a diet can be involved in the formation of atherosclerotic plaques. In addition, the data demonstrate an increase in 48149 mRNA expression in three out of eight marmosets (see #826, #844, #846, #908, #815, #798, #932, and #926) when they were fed a diet containing monounsaturated fat supplemented with cholesteyramin, as compared to when they were fed either a "chow" diet or a diet containing monounsaturated fat that lacked cholesteyramin. Cholesteyramin treatment stimulate the formation of atherosclerotic plaques. Thus, genes that display an increase in expression in response to cholesteyramin treatment are thought to be involved in atherosclerosis.

To follow up the observation that 48149 expression is positively correlated with atherosclerosis, artery tissue samples that was either normal or contained atherosclerotic lesions were analyzed by in-situ hybridization using a 48149 nucleic acid probe. All of the atherosclerotic tissue samples contained elevated expression of 48149 mRNA as compared

to the normal artery samples analyzed. Similarly, artery tissue from ApoE knockout mice, which are genetically susceptible to atherosclerosis, was analyzed by in-situ hybridization. All of the ApoE knockout-derived artery samples that contained atherosclerotic plaques displayed higher expression of 48149 mRNA relative to the normal artery samples. Significantly, expression of 48149 mRNA was localized to macrophages present in the atherosclerotic plaques of the ApoE knockout-derived arteries, again suggesting a role for 48149 in atherosclerosis.

Example 3: Tissue Distribution of 48149 mRNA by Northern Analysis

Northern blot hybridizations with various RNA samples can be performed under standard conditions and washed under stringent conditions, i.e., 0.2xSSC at 65°C. A DNA probe corresponding to all or a portion of the 48149 cDNA (SEQ ID NO:1) can be used. The DNA was radioactively labeled with ^{32}P-dCTP using the Prime-It Kit (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters containing mRNA from mouse hematopoietic and endocrine tissues, and cancer cell lines (Clontech, Palo Alto, CA) can be probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

Example 4: Recombinant Expression of 48149 in Bacterial Cells

In this example, 48149 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, 48149 is fused to GST and this fusion polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-48149 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

Example 5: Expression of Recombinant 48149 Protein in COS Cells

To express the 48149 gene in COS cells (e.g., COS-7 cells, CV-1 origin SV40 cells; Gluzman (1981) *CellI*23:175-182), the pcDNA/Amp vector by Invitrogen

Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire 48149 protein and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

To construct the plasmid, the 48149 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the 48149 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the 48149 coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the 48149_gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5α, SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the 48149-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. The expression of the 48149 polypeptide is detected by radiolabelling (^{35}S-methionine or ^{35}S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) using an HA specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with ^{35}S-methionine (or ^{35}S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40,

0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the 48149 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the 48149 polypeptide is detected by radiolabelling and immunoprecipitation using a 48149 specific monoclonal antibody.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. A method for identifying a compound capable of treating a cardiovascular disorder, comprising assaying the ability of the compound to modulate 48149 nucleic acid expression or 48149 polypeptide activity, thereby identifying a compound capable of treating a cardiovascular disorder.

2. A method for identifying a compound capable of modulating lipid levels comprising:

a) contacting a cell which expresses 48149 with a test compound; and

b) assaying the ability of the test compound to modulate the expression of a 48149 nucleic acid or the activity of a 48149 polypeptide, thereby identifying a compound capable of modulating lipid levels.

3. A method for modulating lipid production in a cell comprising contacting a cell with a 48149 modulator, thereby modulating lipid production in the cell.

4. The method of claim 2, wherein the cell is a liver cell.

5. The method of claim 3, wherein the 48149 modulator is a small organic molecule, peptide, antibody or antisense nucleic acid molecule.

6. The method of claim 3, wherein the 48149 modulator is capable of modulating 48149 polypeptide activity.

7. The method of claim 6, wherein the 48149 modulator is a small organic molecule, peptide, antibody or antisense nucleic acid molecule.

8. The method of claim 6, wherein the 48149 modulator is capable of modulating 48149 nucleic acid expression.

9. A method for treating a subject having a cardiovascular disorder characterized by aberrant 48149 polypeptide activity or aberrant 48149 nucleic acid

expression comprising administering to the subject a 48149 modulator, thereby treating said subject having a cardiovascular disorder.

10. The method of claim 9, wherein said cardiovascular disorder is selected from the group consisting of atherosclerosis, myocardial infarction, stroke, thrombosis, aneurism, heart failure, ischemic heart disease, angina pectoris, sudden cardiac death, hypertensive heart disease; non-coronary vessel disease, such as arteriolosclerosis, small vessel disease, nephropathy, hypertriglyceridemia, hypercholesterolemia, hyperlipidemia, xanthomatosis, asthma, hypertension, emphysema and chronic pulmonary disease.

11. The method of claim 9, wherein said 48149 modulator is administered in a pharmaceutically acceptable formulation.

12. The method of claim 9, wherein the 48149 modulator is a small organic molecule, peptide, antibody or antisense nucleic acid molecule.

13. The method of claim 9, wherein the 48149 modulator is capable of modulating 48149 polypeptide activity.

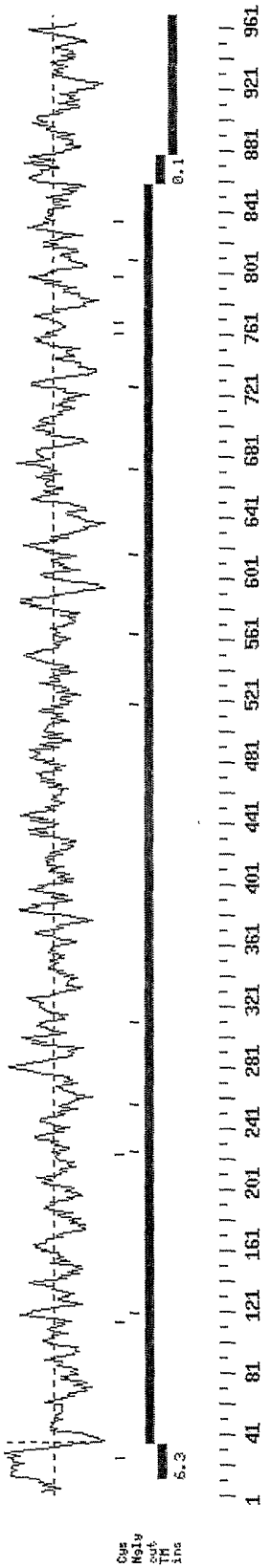

1
41
81
121
161
201
241
281
321
361
401
441
481
521
561
601
641
681
721
761
801
841
881
921
961
0.1
6.3
Gas
Ngly
out
TM
ins

FIGURE 1

```
          Peptidase_M1: domain 1 of 1, from 76 to 480: score 722.8, E = 2.7e-24
Begin SEQ ID NO:4              *->LPttvkPlhYdLtlkpkfgflpekpnyadeknftFsGsvtItltnq
                                  LP+t+kP +Y++tl p   +l++     d ++++F+Gs+t+++  ++
SEQ ID NO:2          48149    76    LPNTLKPDSYQVTLRP---YLTPN----DRGLYVFKGSSTVRF--TC 113

                               tkaatdeIvLhakdltisstgegvrvtlvlvngsqklpesvefslqdetd
                               + +atd+I++h+k+l+++  ++g+rv+l++v gsq+ p +  ++l+++t
                     48149   114 K-EATDVIIIHSKKLNYTL-SQGHRVVLRGVGGSQP-PDIDKTELVEPT- 159

                               flavddnkekltinlpealsagqggspytLeIeyegklndismlGfYrse
                                       e+l++ l+ +l++++   +y+ + e+eg+l+d +++GfYrse
                     48149   160 --------EYLVVHLKGSLVKDS---QYEMDSEFEGELAD-DLAGFYRSE 197

                               YtdgdgetkymatTQfeeptdARraFPCfDEPsfKATFtitiihpkgtta
                               Y++ ++ +k++atTQ++ ++dAR+ FPCfDEP++KA F+it+ihpk++ta
                     48149   198 YME-GNVRKVVATTQMQ-AADARKSFPCFDEPAMKAEFNITLIHPKDLTA 245

                               lSNmpeis.ttkdddgptrvittFetTpkMStYLlAfiVGeleyietetk
                               lSNm+++ ++t++ ++p +++t+F+tTpkMStYLlAfiV+e++y+e +
                     48149   246 LSNMLPKGpSTPLPEDPNWNVTEFHTTPKMSTYLLAFIVSEFDYVEKQAS 295

                               dgysarevpvrvyarpgaknagqgqyalevtkkllefyeeyfgipYplpK
                               +g     v++r++arp+a++ag+g+yal+vt+++l f++ ++++pYplpK
                     48149   296 NG-----VLIRIWARPSAIAAGHGDYALNVTGPILNFFAGHYDTPYPLPK 340

                               lDqvAvPdFsaGAMENWGLiTyrepaLLydprsstnsdkqrvAeViaHEL
                               +Dq+++PdF+aGAMENWGL+Tyre++LL+dp ss++s+k+rv +ViaHEL
                     48149   341 SDQIGLPDFNAGAMENWGLVTYRENSLLFDPLSSSSSNKERVVTVIAHEL 390

                               AHQWFGNLVTmkWWddLWLNEGFAtymEylgtdelggepewnieaqfllr
                               AHQWFGNLVT +WW+dLWLNEGFA+y+Eylg+d++  ep+wn +++++l
                     48149   391 AHQWFGNLVTIEWWNDLWLNEGFASYVEYLGADYA--EPTWNLKDLMVL- 437

                               ddvaqlalasDslgsshPitnklvevntpaeiseiFdsaitYaKG<-*  End SEQ ID NO:4
                               +dv ++++a D+l+sshP++++ +e+ntpa+ise+Fd ai+Y+KG
                     48149   438 NDV-YRVMAVDALASSHPLSTPASEINTPAQISELFD-AISYSKG    480   SEQ ID NO:2
```

FIGURE 2

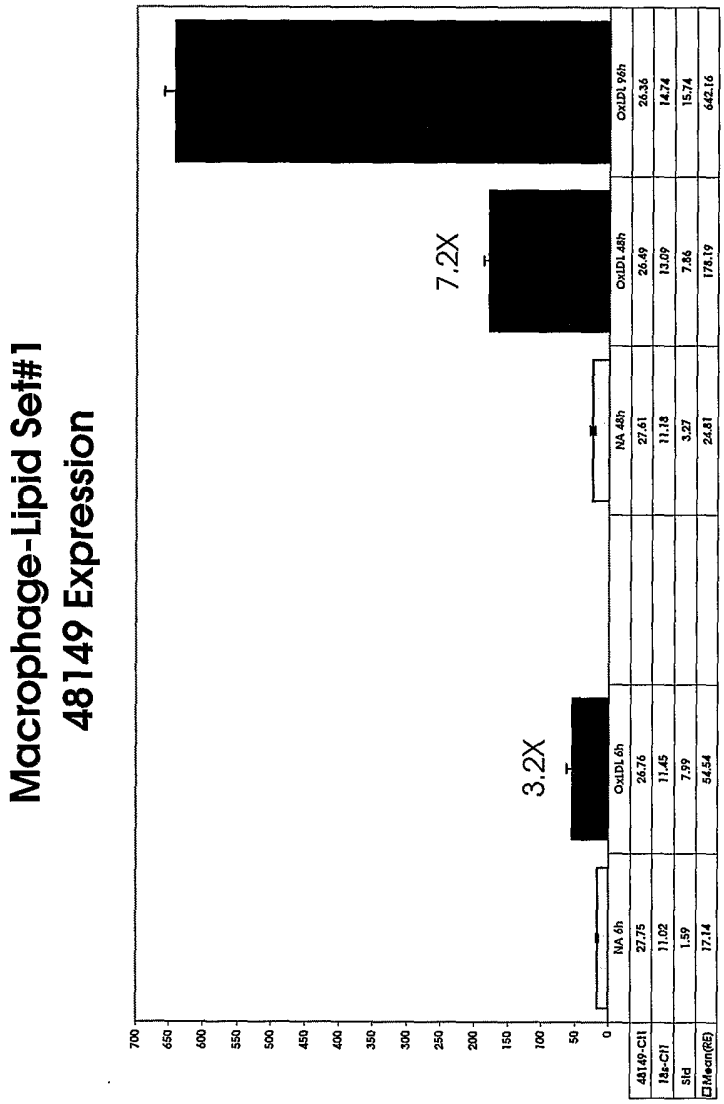

Macrophage-Lipid Set#1
48149 Expression
3.2X
7.2X
700
650
600
550
500
450
400
350
300
250
200
150
100
50
0
NA 6h
OxLDL 6h
NA 48h
OxLDL 48h
OxLDL 96h
48149-Ctl
18s-Ctl
Std
Mean(RE)
27.75
26.76
27.61
26.49
26.36
11.02
11.45
11.18
13.09
14.74
1.59
7.59
3.27
7.86
15.74
17.14
54.54
24.81
178.19
642.16

FIGURE 3

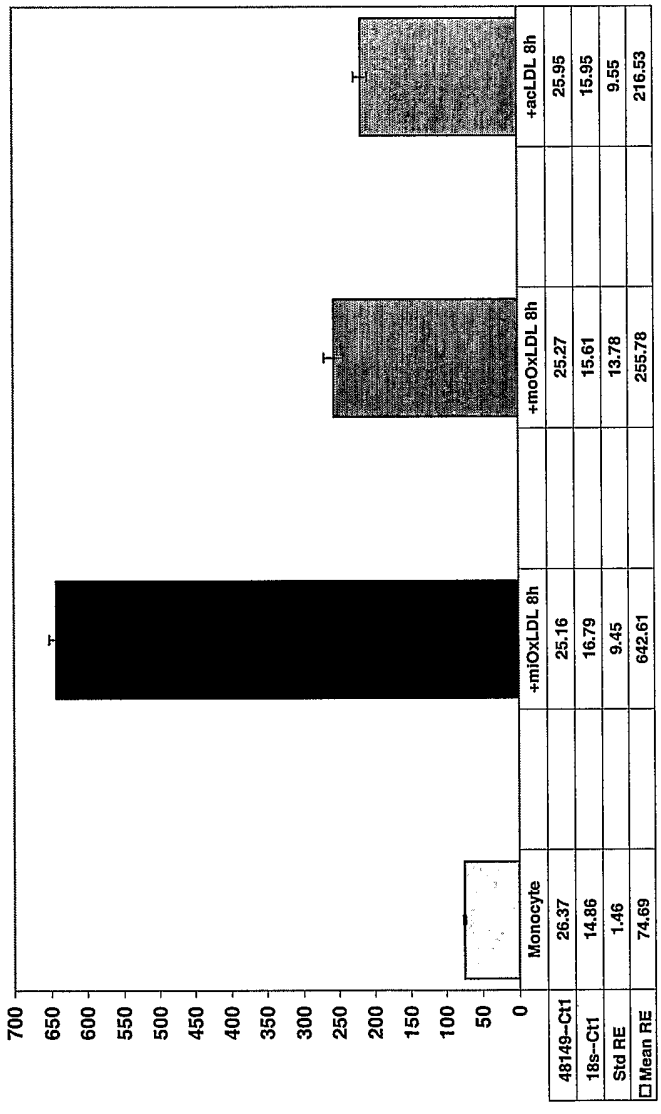

48149 Mac Lipid Loading #3
700
650
600
550
500
450
400
350
300
250
200
150
100
50
0
Monocyte
+miOxLDL 8h
+moOxLDL 8h
+acLDL 8h
48149--Ct1 26.37 25.16 25.27 25.95
18s--Ct1 14.86 16.79 15.61 15.95
Std RE 1.46 9.45 13.78 9.55
Mean RE 74.69 642.61 255.78 216.53

FIGURE 4

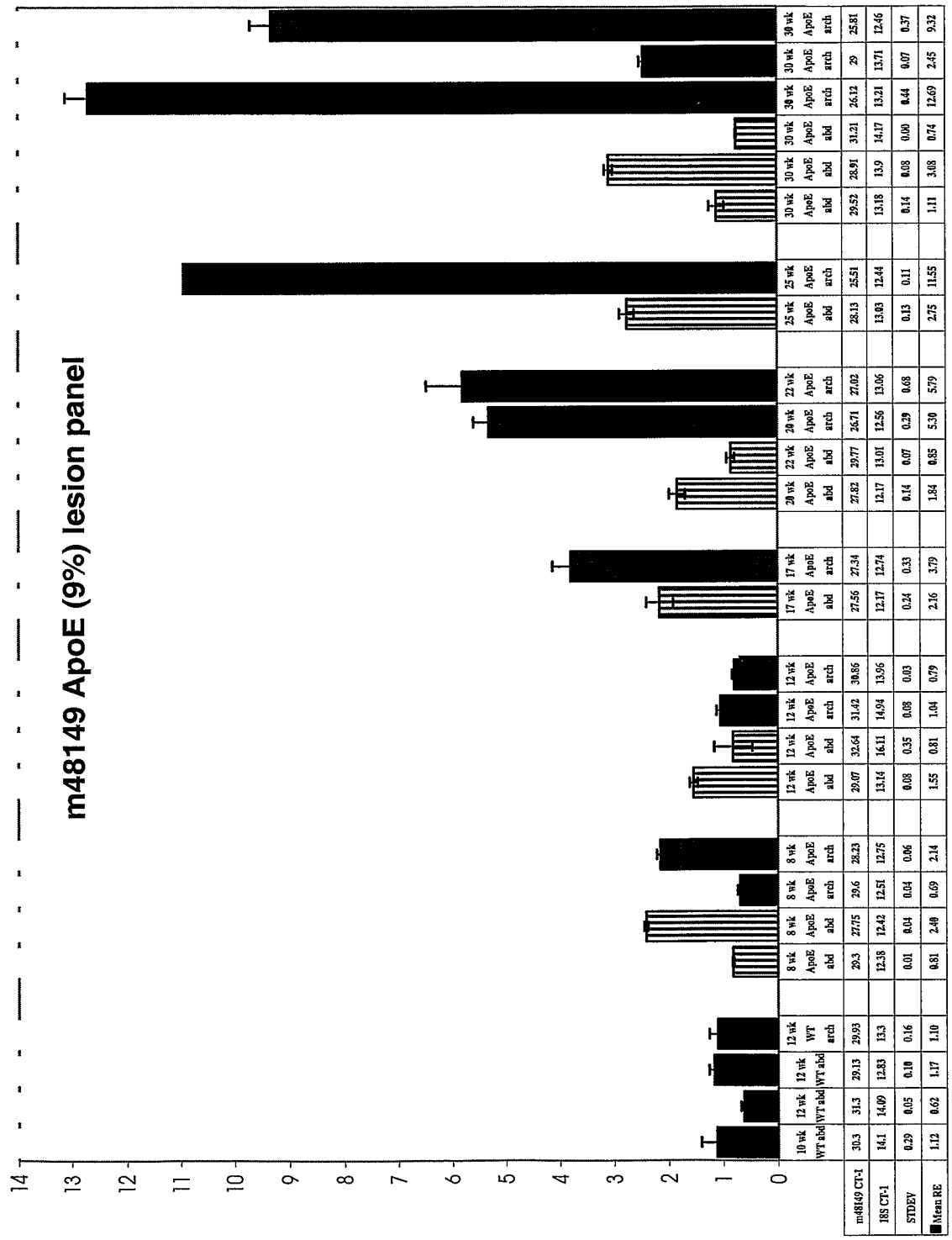

m48149 ApoE (9%) lesion panel
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
10 wk WT abd
12 wk WT abd
12 wk WT abd
12 wk WT arch
8 wk ApoE abd
8 wk ApoE abd
8 wk ApoE arch
8 wk ApoE arch
12 wk ApoE abd
12 wk ApoE abd
12 wk ApoE arch
12 wk ApoE arch
17 wk ApoE abd
17 wk ApoE arch
20 wk ApoE abd
20 wk ApoE arch
22 wk ApoE abd
22 wk ApoE arch
25 wk ApoE abd
25 wk ApoE arch
30 wk ApoE abd
30 wk ApoE abd
30 wk ApoE arch
30 wk ApoE arch
m48149 CT-1 30.3 31.3 29.13 29.93 29.3 27.75 29.6 28.23 29.07 32.64 31.42 30.86 27.56 27.34 27.82 29.77 26.71 27.02 28.13 25.51 29.52 28.91 31.21 26.12 29 25.81
18S CT-1 14.1 14.09 12.83 13.3 12.38 12.42 12.51 12.75 13.14 16.11 14.94 13.95 12.17 12.74 12.17 13.01 12.56 13.06 13.03 12.44 13.18 13.9 14.17 13.21 13.71 12.46
STDEV 0.29 0.05 0.10 0.16 0.01 0.04 0.04 0.06 0.08 0.35 0.08 0.03 0.24 0.33 0.14 0.07 0.29 0.68 0.13 0.11 0.14 0.08 0.00 0.44 0.07 0.37
Mean RE 1.12 0.62 1.17 1.10 0.81 2.40 0.69 2.14 1.55 0.81 1.04 0.79 2.16 3.79 1.84 0.85 5.30 5.79 2.75 11.55 1.11 3.08 0.74 12.69 2.45 9.32

FIGURE 5A

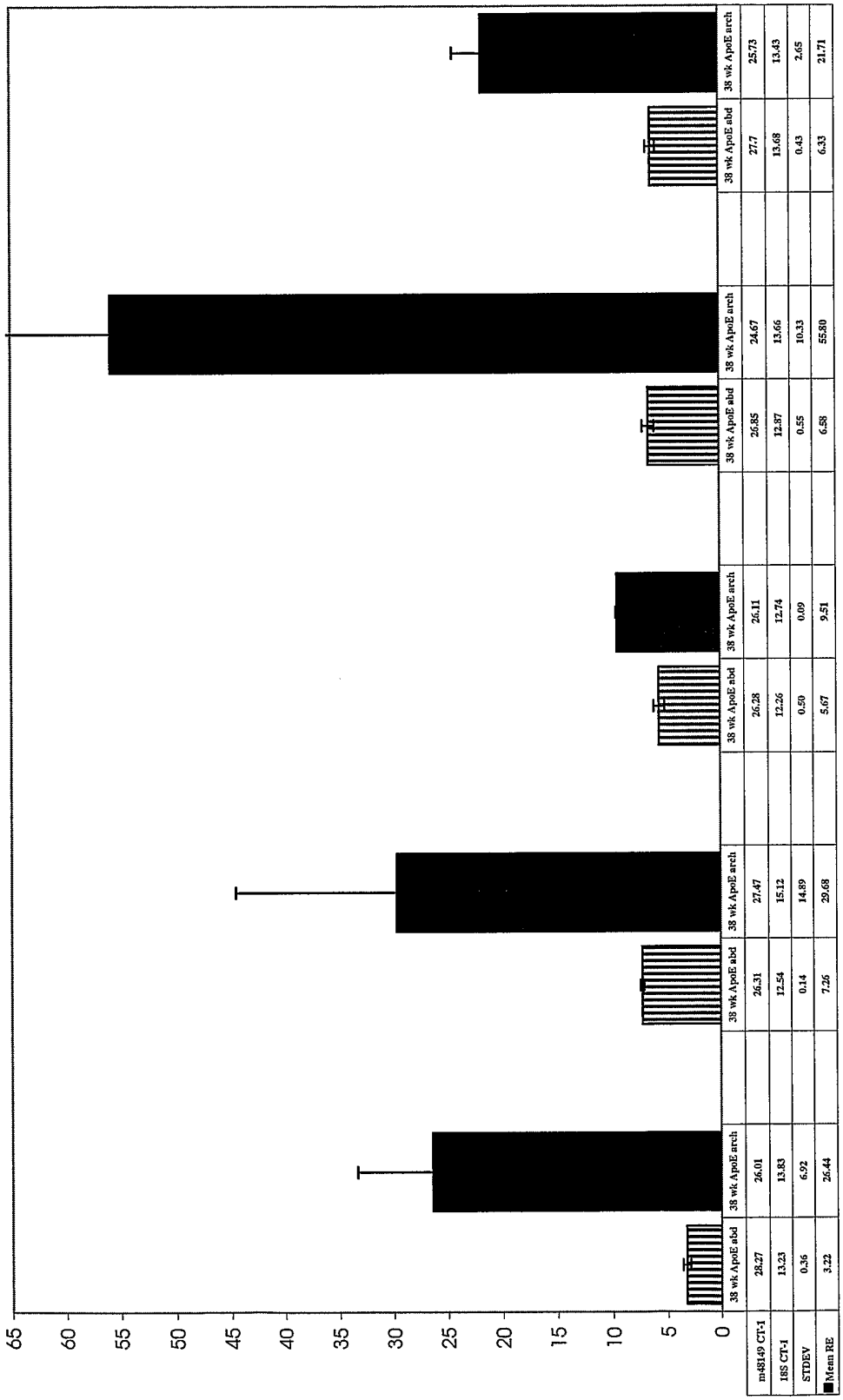
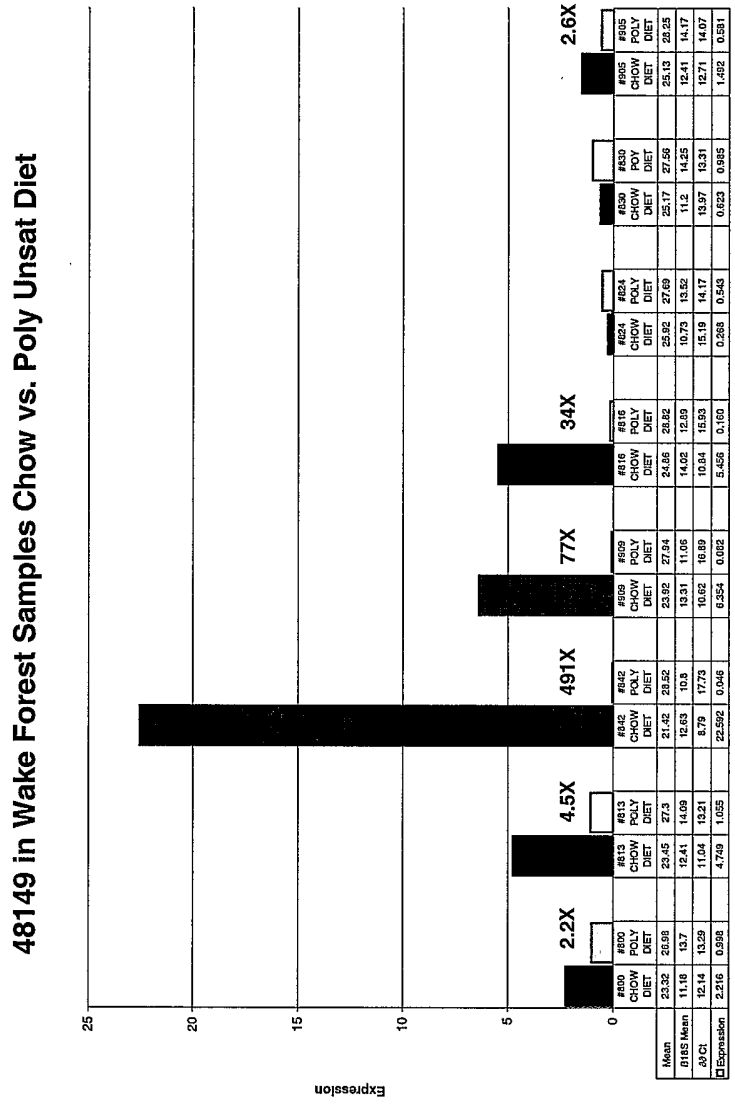

65
60
55
50
45
40
35
30
25
20
15
10
5
0
38 wk ApoE abd
38 wk ApoE arch
38 wk ApoE abd
38 wk ApoE arch
38 wk ApoE abd
38 wk ApoE arch
38 wk ApoE abd
38 wk ApoE arch
m49149 CT-1 28.27 26.01 26.31 27.47 26.28 26.11 26.85 24.67 27.7 25.73
18S CT-1 13.23 13.83 12.54 15.12 12.26 12.74 12.87 13.66 13.68 13.43
STDEV 0.36 6.92 0.14 14.89 0.50 0.09 0.55 10.33 0.43 2.65
Mean RE 3.22 26.44 7.26 29.68 5.67 9.51 6.58 55.80 6.33 21.71

FIGURE 5B



	#300 CHOW DIET	#300 POLY DIET	#813 CHOW DIET	#813 POLY DIET	#842 CHOW DIET	#842 POLY DIET	#609 CHOW DIET	#609 POLY DIET	#816 CHOW DIET	#816 POLY DIET	#824 CHOW DIET	#824 POLY DIET	#530 CHOW DIET	#530 POY DIET	#905 CHOW DIET	#905 POLY DIET
Mean	23.32	26.98	23.45	27.3	21.42	28.52	23.92	27.94	24.86	28.82	25.92	27.69	25.17	27.56	25.13	28.25
18S Mean	11.18	13.7	12.41	14.09	12.63	10.8	13.31	11.06	14.02	12.69	10.73	13.52	11.2	14.25	12.41	14.17
ΔΔCt	12.14	13.29	11.04	13.21	8.79	17.73	10.62	16.89	10.84	15.93	15.19	14.17	13.97	13.31	12.71	14.07
Expression	2.216	0.598	4.749	1.055	22.592	0.046	6.354	0.022	5.456	0.160	0.268	0.543	0.623	0.985	1.492	0.581

FIGURE 6

SEQUENCE LISTING

<110> Chun, Miyoung

<120> METHODS OF USING 48149, A HUMAN AMINOPEPTIDASE FAMILY MEMBER

<130> MPI01-174

<150> 60/335,084
<151> 2001-10-31

<160> 4

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 3494
<212> DNA
<213> Homo sapien

<400> 1
taatttttgc ccagtctgcc tgttgtgggg ctcctcccct ttggggatat aagcccggcc 60
tggggctgct ccgttctctg cctggcctga ggctccctga gccgcctccc caccatcacc 120
atggccaagg gcttctatat ttccaagtcc ctgggcatcc tggggatcct cctgggcgtg 180
gcagccgtgt gcacaatcat cgcactgtca gtggtgtact cccaggagaa gaacaagaac 240
gccaacagct ccccccgtggc ctccaccacc ccgtccgcct cagccaccac caaccccgcc 300
tcggccacca ccttggacca aagtaaagcg tggaatcgtt accgcctccc caacacgctg 360
aaacccgatt cctaccaggt gacgctgaga ccgtacctca cccccaatga caggggcctg 420
tacgttttta agggctccag caccgtccgt ttcacctgca aggaggccac tgacgtcatc 480
atcatccaca gcaagaagct caactacacc ctcagccagg ggcacagggt ggtcctgcgt 540
ggtgtgggag gctcccagcc ccccgacatt gacaagactg agctggtgga gcccaccgag 600
tacctggtgg tgcacctcaa gggctccctg gtgaaggaca gccagtatga gatggacagc 660
gagttcgagg gggagttggc agatgacctg gcgggcttct accgcagcga gtacatggag 720
ggcaatgtca gaaaggtggt ggccactaca cagatgcagg ctgcagatgc ccggaagtcc 780
ttcccatgct tcgatgagcc ggccatgaag gccgagttca acatcacgct tatccacccc 840
aaggacctga cagccctgtc caacatgctt cccaaaggtc ccagcacccc acttccagaa 900
gaccccaact ggaatgtcac tgagttccac accacgccca agatgtccac gtacttgctg 960
gccttcattg tcagtgagtt cgactacgtg gagaagcagg catccaatgg tgtcttgatc 1020
cggatctggg cccggcccag tgccattgcg gcgggccacg gcgattatgc cctgaacgtg 1080
acgggcccca tccttaactt ctttgctggt cattatgaca caccctaccc actcccaaaa 1140
tcagaccaga ttggcctgcc agacttcaac gccggcgcca tggagaactg gggactggtg 1200
acctaccggg agaactccct gctgttcgac cccctgtcct cctccagcag caacaaggag 1260
cgggtggtca ctgtgattgc tcatgagctg gcccaccagt ggttcgggaa cctggtgacc 1320
atagagtggt ggaatgacct gtggctgaac gagggcttcg cctcctacgt ggagtacctg 1380
ggtgctgact atgcggagcc cacctggaac ttgaaagacc tcatggtgct gaatgatgtg 1440
taccgcgtga tggcagtgga tgcactggcc tcctcccacc cgctgtccac acccgcctcg 1500
gagatcaaca cgccggccca gatcagtgag ctgtttgacg ccatctccta cagcaagggc 1560
gcctcagtcc tcaggatgct ctccagcttc ctgtccgagg acgtattcaa gcagggcctg 1620
gcgtcctacc tccacacctt tgcctaccag aacaccatct acctgaacct gtgggaccac 1680
ctgcaggagg ctgtgaacaa ccggtccatc caactcccca ccaccgtgcg ggacatcatg 1740
aaccgctgga ccctgcagat gggcttcccg gtcatcacgg tggataccag cacggggacc 1800
ctttcccagg agcacttcct ccttgacccc gattccaatg ttacccgccc ctcagaattc 1860
aactacgtgt ggattgtgcc catcacatcc atcagagatg gcagacagca gcaggactac 1920
tggctgatag atgtaagagc ccagaacgat ctcttcagca catcaggcaa tgagtgggtc 1980
ctgctgaacc tcaatgtgac gggctattac cgggtgaact acgacgaaga gaactggagg 2040
aagattcaga ctcagctgca gagagaccac tcggccatcc ctgtcatcaa tcgggcacag 2100
atcattaatg acgccttcaa cctggccagt gcccataagg tccctgtcac tctggcgctg 2160
aacaacaccc tcttcctgat tgaagagaga cagtacatgc cctgggaggc cgccctgagc 2220
agcctgagct acttcaagct catgtttgac cgctccgagg tctatggccc catgaagaac 2280

tacctgaaga agcaggtcac acccctcttc attcacttca gaaataatac caacaactgg 2340
agggagatcc cagaaaacct gatggaccag tacagcgagg ttaatgccat cagcaccgcc 2400
tgctccaacg gagttccaga gtgtgaggag atggtctctg gccttttcaa gcagtggatg 2460
gagaacccca ataataaccc gatccacccc aacctgcggt ccaccgtcta ctgcaacgct 2520
atcgcccagg gcggggagga ggagtgggac ttcgcctggg agcagttccg aaatgccaca 2580
ctggtcaatg aggctgacaa gctccgggca gccctggcct gcagcaaaga gttgtggatc 2640
ctgaacaggt acctgagcta caccctgaac ccggacttaa tccggaagca ggacgccacc 2700
tctaccatca tcagcattac caacaacgtc attgggcaag gtctggtctg ggactttgtc 2760
cagagcaact ggaagaagct ttttaacgat tatggtggtg gctcgttctc cttctccaac 2820
ctcatccagg cagtgacacg acgattctcc accgagtatg agctgcagca gctggagcag 2880
ttcaagaagg acaacgagga aacaggcttc ggctcaggca cccgggccct ggagcaagcc 2940
ctggagaaga cgaaagccaa catcaagtgg gtgaaggaga acaaggaggt ggtgctccag 3000
tggttcacag aaaacagcaa atagtcccca gcccttgaag tcacccggcc ccgatgcaag 3060
gtgcccacat gtgtccatcc cagcggctgg tgcagggcct ccattcctgg agcccgaggc 3120
accagtgtcc tcccctcaag gacaaagtct ccagcccacg ttctctctgc ctgtgagcca 3180
gtctagttcc tgatgaccca ggctgcctga gcacctccca gcccctgccc ctcatgccaa 3240
ccccgcccta ggcctggcat ggcacctgtc gcccagtgcc ctggggctga tctcagggaa 3300
gcccagctcc agggccagat gagcagaagc tctcgatgga caatgaacgg ccttgctggg 3360
ggccgccctg taccctcttt caccctttccc taaagaccct aaatctgagg aatcaacagg 3420
gcagcagatc tgtatatttt tttctaagag aaaatgtaaa taaaggattt ctagatgaaa 3480
aaaaaaaaaa aaaa 3494

<210> 2
<211> 967
<212> PRT
<213> Homo sapien

<400> 2
Met Ala Lys Gly Phe Tyr Ile Ser Lys Ser Leu Gly Ile Leu Gly Ile
1 5 10 15
Leu Leu Gly Val Ala Ala Val Cys Thr Ile Ile Ala Leu Ser Val Val
20 25 30
Tyr Ser Gln Glu Lys Asn Lys Asn Ala Asn Ser Ser Pro Val Ala Ser
35 40 45
Thr Thr Pro Ser Ala Ser Ala Thr Thr Asn Pro Ala Ser Ala Thr Thr
50 55 60
Leu Asp Gln Ser Lys Ala Trp Asn Arg Tyr Arg Leu Pro Asn Thr Leu
65 70 75 80
Lys Pro Asp Ser Tyr Gln Val Thr Leu Arg Pro Tyr Leu Thr Pro Asn
85 90 95
Asp Arg Gly Leu Tyr Val Phe Lys Gly Ser Ser Thr Val Arg Phe Thr
100 105 110
Cys Lys Glu Ala Thr Asp Val Ile Ile Ile His Ser Lys Lys Leu Asn
115 120 125
Tyr Thr Leu Ser Gln Gly His Arg Val Val Leu Arg Gly Val Gly Gly
130 135 140
Ser Gln Pro Pro Asp Ile Asp Lys Thr Glu Leu Val Glu Pro Thr Glu
145 150 155 160
Tyr Leu Val Val His Leu Lys Gly Ser Leu Val Lys Asp Ser Gln Tyr
165 170 175
Glu Met Asp Ser Glu Phe Glu Gly Glu Leu Ala Asp Asp Leu Ala Gly
180 185 190
Phe Tyr Arg Ser Glu Tyr Met Glu Gly Asn Val Arg Lys Val Val Ala
195 200 205
Thr Thr Gln Met Gln Ala Ala Asp Ala Arg Lys Ser Phe Pro Cys Phe
210 215 220
Asp Glu Pro Ala Met Lys Ala Glu Phe Asn Ile Thr Leu Ile His Pro
225 230 235 240
Lys Asp Leu Thr Ala Leu Ser Asn Met Leu Pro Lys Gly Pro Ser Thr
245 250 255
Pro Leu Pro Glu Asp Pro Asn Trp Asn Val Thr Glu Phe His Thr Thr

```
            260                 265                 270
Pro Lys Met Ser Thr Tyr Leu Leu Ala Phe Ile Val Ser Glu Phe Asp
        275                 280                 285
Tyr Val Glu Lys Gln Ala Ser Asn Gly Val Leu Ile Arg Ile Trp Ala
    290                 295                 300
Arg Pro Ser Ala Ile Ala Ala Gly His Gly Asp Tyr Ala Leu Asn Val
305                 310                 315                 320
Thr Gly Pro Ile Leu Asn Phe Phe Ala Gly His Tyr Asp Thr Pro Tyr
                325                 330                 335
Pro Leu Pro Lys Ser Asp Gln Ile Gly Leu Pro Asp Phe Asn Ala Gly
            340                 345                 350
Ala Met Glu Asn Trp Gly Leu Val Thr Tyr Arg Glu Asn Ser Leu Leu
        355                 360                 365
Phe Asp Pro Leu Ser Ser Ser Ser Ser Asn Lys Glu Arg Val Val Thr
    370                 375                 380
Val Ile Ala His Glu Leu Ala His Gln Trp Phe Gly Asn Leu Val Thr
385                 390                 395                 400
Ile Glu Trp Trp Asn Asp Leu Trp Leu Asn Glu Gly Phe Ala Ser Tyr
                405                 410                 415
Val Glu Tyr Leu Gly Ala Asp Tyr Ala Glu Pro Thr Trp Asn Leu Lys
            420                 425                 430
Asp Leu Met Val Leu Asn Asp Val Tyr Arg Val Met Ala Val Asp Ala
        435                 440                 445
Leu Ala Ser Ser His Pro Leu Ser Thr Pro Ala Ser Glu Ile Asn Thr
    450                 455                 460
Pro Ala Gln Ile Ser Glu Leu Phe Asp Ala Ile Ser Tyr Ser Lys Gly
465                 470                 475                 480
Ala Ser Val Leu Arg Met Leu Ser Ser Phe Leu Ser Glu Asp Val Phe
                485                 490                 495
Lys Gln Gly Leu Ala Ser Tyr Leu His Thr Phe Ala Tyr Gln Asn Thr
            500                 505                 510
Ile Tyr Leu Asn Leu Trp Asp His Leu Gln Glu Ala Val Asn Asn Arg
        515                 520                 525
Ser Ile Gln Leu Pro Thr Thr Val Arg Asp Ile Met Asn Arg Trp Thr
    530                 535                 540
Leu Gln Met Gly Phe Pro Val Ile Thr Val Asp Thr Ser Thr Gly Thr
545                 550                 555                 560
Leu Ser Gln Glu His Phe Leu Leu Asp Pro Asp Ser Asn Val Thr Arg
                565                 570                 575
Pro Ser Glu Phe Asn Tyr Val Trp Ile Val Pro Ile Thr Ser Ile Arg
            580                 585                 590
Asp Gly Arg Gln Gln Gln Asp Tyr Trp Leu Ile Asp Val Arg Ala Gln
        595                 600                 605
Asn Asp Leu Phe Ser Thr Ser Gly Asn Glu Trp Val Leu Leu Asn Leu
    610                 615                 620
Asn Val Thr Gly Tyr Tyr Arg Val Asn Tyr Asp Glu Glu Asn Trp Arg
625                 630                 635                 640
Lys Ile Gln Thr Gln Leu Gln Arg Asp His Ser Ala Ile Pro Val Ile
                645                 650                 655
Asn Arg Ala Gln Ile Ile Asn Asp Ala Phe Asn Leu Ala Ser Ala His
            660                 665                 670
Lys Val Pro Val Thr Leu Ala Leu Asn Asn Thr Leu Phe Leu Ile Glu
        675                 680                 685
Glu Arg Gln Tyr Met Pro Trp Glu Ala Ala Leu Ser Ser Leu Ser Tyr
    690                 695                 700
Phe Lys Leu Met Phe Asp Arg Ser Glu Val Tyr Gly Pro Met Lys Asn
705                 710                 715                 720
Tyr Leu Lys Lys Gln Val Thr Pro Leu Phe Ile His Phe Arg Asn Asn
                725                 730                 735
Thr Asn Asn Trp Arg Glu Ile Pro Glu Asn Leu Met Asp Gln Tyr Ser
            740                 745                 750
```

```
Glu Val Asn Ala Ile Ser Thr Ala Cys Ser Asn Gly Val Pro Glu Cys
        755                 760                 765
Glu Glu Met Val Ser Gly Leu Phe Lys Gln Trp Met Glu Asn Pro Asn
    770                 775                 780
Asn Asn Pro Ile His Pro Asn Leu Arg Ser Thr Val Tyr Cys Asn Ala
785                 790                 795                 800
Ile Ala Gln Gly Gly Glu Glu Glu Trp Asp Phe Ala Trp Glu Gln Phe
                805                 810                 815
Arg Asn Ala Thr Leu Val Asn Glu Ala Asp Lys Leu Arg Ala Ala Leu
            820                 825                 830
Ala Cys Ser Lys Glu Leu Trp Ile Leu Asn Arg Tyr Leu Ser Tyr Thr
        835                 840                 845
Leu Asn Pro Asp Leu Ile Arg Lys Gln Asp Ala Thr Ser Thr Ile Ile
    850                 855                 860
Ser Ile Thr Asn Asn Val Ile Gly Gln Gly Leu Val Trp Asp Phe Val
865                 870                 875                 880
Gln Ser Asn Trp Lys Lys Leu Phe Asn Asp Tyr Gly Gly Gly Ser Phe
                885                 890                 895
Ser Phe Ser Asn Leu Ile Gln Ala Val Thr Arg Arg Phe Ser Thr Glu
            900                 905                 910
Tyr Glu Leu Gln Gln Leu Glu Gln Phe Lys Lys Asp Asn Glu Glu Thr
        915                 920                 925
Gly Phe Gly Ser Gly Thr Arg Ala Leu Glu Gln Ala Leu Glu Lys Thr
    930                 935                 940
Lys Ala Asn Ile Lys Trp Val Lys Glu Asn Lys Glu Val Val Leu Gln
945                 950                 955                 960
Trp Phe Thr Glu Asn Ser Lys
                965
```

<210> 3
<211> 2904
<212> DNA
<213> Homo sapien

<400> 3

```
atggccaagg gcttctatat ttccaagtcc ctgggcatcc tggggatcct cctgggcgtg 60
gcagccgtgt gcacaatcat cgcactgtca gtggtgtact cccaggagaa gaacaagaac 120
gccaacagct ccccgtggc ctccaccacc ccgtccgcct cagccaccac caaccccgcc 180
tcggccacca ccttggacca aagtaaagcg tggaatcgtt accgcctccc caacacgctg 240
aaacccgatt cctaccaggt gacgctgaga ccgtacctca cccccaatga caggggcctg 300
tacgttttta agggctccag caccgtccgt ttcacctgca aggaggccac tgacgtcatc 360
atcatccaca gcaagaagct caactacacc ctcagccagg ggcacagggt ggtcctgcgt 420
ggtgtgggag gctcccagcc ccccgacatt gacaagactg agctggtgga gcccaccgag 480
tacctggtgg tgcacctcaa gggctccctg gtgaaggaca gccagtatga gatggacagc 540
gagttcgagg gggagttggc agatgacctg gcgggcttct accgcagcga gtacatggag 600
ggcaatgtca gaaaggtggt ggccactaca cagatgcagg ctgcagatgc ccggaagtcc 660
ttcccatgct tcgatgagcc ggccatgaag gccgagttca acatcacgct tatccacccc 720
aaggacctga cagccctgtc caacatgctt cccaaaggtc ccagcacccc acttccagaa 780
gaccccaact ggaatgtcac tgagttccac accacgccca agatgtccac gtacttgctg 840
gccttcattg tcagtgagtt cgactacgtg gagaagcagg catccaatgg tgtcttgatc 900
cggatctggg cccggcccag tgccattgcg gcgggccacg gcgattatgc cctgaacgtg 960
acgggcccca tccttaactt ctttgctggt cattatgaca caccctaccc actcccaaaa 1020
tcagaccaga ttggcctgcc agacttcaac gccggcgcca tggagaactg gggactggtg 1080
acctaccggg agaactccct gctgttcgac cccctgtcct cctccagcag caacaaggag 1140
cgggtggtca ctgtgattgc tcatgagctg gcccaccagt ggttcgggaa cctggtgacc 1200
atagagtggt ggaatgacct gtggctgaac gagggcttcg cctcctacgt ggagtacctg 1260
ggtgctgact atgcggagcc cacctggaac ttgaaagacc tcatggtgct gaatgatgtg 1320
taccgcgtga tggcagtgga tgcactggcc tcctcccacc cgctgtccac acccgcctcg 1380
gagatcaaca cgccggccca gatcagtgag ctgtttgacg ccatctccta cagcaagggc 1440
gcctcagtcc tcaggatgct ctccagcttc ctgtccgagg acgtattcaa gcagggcctg 1500
```

gcgtcctacc tccacacctt tgcctaccag aacaccatct acctgaacct gtgggaccac 1560
ctgcaggagg ctgtgaacaa ccggtccatc caactcccca ccaccgtgcg ggacatcatg 1620
aaccgctgga ccctgcagat gggcttcccg gtcatcacgg tggataccag cacggggacc 1680
ctttcccagg agcacttcct ccttgacccc gattccaatg ttacccgccc ctcagaattc 1740
aactacgtgt ggattgtgcc catcacatcc atcagagatg gcagacagca gcaggactac 1800
tggctgatag atgtaagagc ccagaacgat ctcttcagca catcaggcaa tgagtgggtc 1860
ctgctgaacc tcaatgtgac gggctattac cgggtgaact acgacgaaga gaactggagg 1920
aagattcaga ctcagctgca gagagaccac tcggccatcc ctgtcatcaa tcgggcacag 1980
atcattaatg acgccttcaa cctggccagt gcccataagg tccctgtcac tctggcgctg 2040
aacaacaccc tcttcctgat tgaagagaga cagtacatgc cctgggaggc cgccctgagc 2100
agcctgagct acttcaagct catgtttgac cgctccgagg tctatggccc catgaagaac 2160
tacctgaaga agcaggtcac acccctcttc attcacttca gaaataatac caacaactgg 2220
agggagatcc cagaaaacct gatggaccag tacagcgagg ttaatgccat cagcaccgcc 2280
tgctccaacg gagttccaga gtgtgaggag atggtctctg gccttttcaa gcagtggatg 2340
gagaacccca ataataaccc gatccacccc aacctgcggt ccaccgtcta ctgcaacgct 2400
atcgcccagg gcggggagga ggagtgggac ttcgcctggg agcagttccg aaatgccaca 2460
ctggtcaatg aggctgacaa gctccgggca gccctggcct gcagcaaaga gttgtggatc 2520
ctgaacaggt acctgagcta caccctgaac ccggacttaa tccggaagca ggacgccacc 2580
tctaccatca tcagcattac caacaacgtc attgggcaag gtctggtctg ggactttgtc 2640
cagagcaact ggaagaagct ttttaacgat tatggtggtg gctcgttctc cttctccaac 2700
ctcatccagg cagtgacacg acgattctcc accgagtatg agctgcagca gctggagcag 2760
ttcaagaagg acaacgagga aacaggcttc ggctcaggca cccgggccct ggagcaagcc 2820
ctggagaaga cgaaagccaa catcaagtgg gtgaaggaga acaaggaggt ggtgctccag 2880
tggttcacag aaaacagcaa atag 2904

<210> 4
<211> 438
<212> PRT
<213> Artificial Sequence

<220>
<223> Consensus Sequence

<400> 4
Leu Pro Thr Thr Val Lys Pro Leu His Tyr Asp Leu Thr Leu Lys Pro
1 5 10 15
Lys Phe Gly Phe Leu Pro Glu Lys Pro Asn Tyr Ala Asp Glu Lys Asn
20 25 30
Phe Thr Phe Ser Gly Ser Val Thr Ile Thr Leu Thr Asn Gln Thr Lys
35 40 45
Ala Ala Thr Asp Glu Ile Val Leu His Ala Lys Asp Leu Thr Ile Ser
50 55 60
Ser Thr Gly Glu Gly Val Arg Val Thr Leu Val Leu Val Asn Gly Ser
65 70 75 80
Gln Lys Leu Pro Glu Ser Val Glu Phe Ser Leu Gln Asp Glu Thr Asp
85 90 95
Phe Leu Ala Val Asp Asp Asn Lys Glu Lys Leu Thr Ile Asn Leu Pro
100 105 110
Glu Ala Leu Ser Ala Gly Gln Gly Gly Ser Pro Tyr Thr Leu Glu Ile
115 120 125
Glu Tyr Glu Gly Lys Leu Asn Asp Ile Ser Met Leu Gly Phe Tyr Arg
130 135 140
Ser Glu Tyr Thr Asp Gly Asp Gly Glu Thr Lys Tyr Met Ala Thr Thr
145 150 155 160
Gln Phe Glu Glu Pro Thr Asp Ala Arg Arg Ala Phe Pro Cys Phe Asp
165 170 175
Glu Pro Ser Phe Lys Ala Thr Phe Thr Ile Thr Ile Ile His Pro Lys
180 185 190
Gly Thr Thr Ala Leu Ser Asn Met Pro Glu Ile Ser Lys Asp Asp Asp
195 200 205
Gly Pro Thr Arg Val Ile Thr Thr Phe Glu Thr Thr Pro Lys Met Ser

```
    210                 215                 220
Thr Tyr Leu Leu Ala Phe Ile Val Gly Glu Leu Glu Tyr Ile Glu Thr
225                 230                 235                 240
Glu Thr Lys Asp Gly Tyr Ser Ala Arg Glu Val Pro Val Arg Val Tyr
                245                 250                 255
Ala Arg Pro Gly Ala Lys Asn Ala Gly Gln Gly Gln Tyr Ala Leu Glu
            260                 265                 270
Val Thr Lys Lys Leu Leu Glu Phe Tyr Glu Glu Tyr Phe Gly Ile Pro
        275                 280                 285
Tyr Pro Leu Pro Lys Leu Asp Gln Val Ala Val Pro Asp Phe Ser Ala
    290                 295                 300
Gly Ala Met Glu Asn Trp Gly Leu Ile Thr Tyr Arg Glu Pro Ala Leu
305                 310                 315                 320
Leu Tyr Asp Pro Arg Ser Ser Thr Asn Ser Asp Lys Gln Arg Val Ala
                325                 330                 335
Glu Val Ile Ala His Glu Leu Ala His Gln Trp Phe Gly Asn Leu Val
            340                 345                 350
Thr Met Lys Trp Trp Asp Asp Leu Trp Leu Asn Glu Gly Phe Ala Thr
        355                 360                 365
Tyr Met Glu Tyr Leu Gly Thr Asp Glu Leu Gly Gly Glu Pro Glu Trp
    370                 375                 380
Asn Ile Glu Ala Gln Phe Leu Leu Arg Asp Asp Val Ala Gln Leu Ala
385                 390                 395                 400
Leu Ala Ser Asp Ser Leu Gly Ser Ser His Pro Ile Thr Asn Lys Leu
                405                 410                 415
Val Glu Val Asn Thr Pro Ala Glu Ile Ser Glu Ile Phe Asp Ser Ala
            420                 425                 430
Ile Thr Tyr Ala Lys Gly
        435
```